Géotechnique **00**, 1–16 [http://dx.doi.org/10.1680/geot.XX.XXXXX]
This is the Author Manuscript of the work. The final published article is available at https://doi.org/10.1680/jgeot.16.P.336.

Characterising three-dimensional anisotropic spatial correlation of soil properties through *in situ* test results

W. F. LIU* and Y. F. LEUNG*

Proper characterisations of soil properties and their variations are pivotal to the geotechnical design process. While multiple *in situ* soil tests are routinely specified and performed in geotechnical investigation programmes, the information they provide regarding the spatial correlations of soil properties are often not fully utilised. This paper presents a holistic framework to characterise the three-dimensional, anisotropic spatial variability of soil properties, using results of *in situ* tests such as standard penetration tests or vane shear tests. The restricted maximum likelihood method is implemented with an anisotropic covariance model, leading to improved predictive capabilities compared to conventional approaches, and allows quantification of the uncertainties on soil properties at unsampled locations, represented as distributions of prediction variance across the entire subsurface three-dimensional domain. The magnitudes of prediction variance at different locations can be used to provide guidance on the necessities and locations of additional soil sampling. They can also provide key input parameters for random field models in site-specific probabilistic analyses of geotechnical projects. The proposed approach is applied to the study of two project sites in Hong Kong, where it is shown that the three-dimensional spatial correlation features may be interpreted together with the geological settings at the site.

KEYWORDS: Site investigation, *in situ* soil testing, anisotropic spatial correlation, statistical analysis

INTRODUCTION

Spatial uncertainty involving geotechnical properties is often the cause of unexpected engineering performance. To quantify such uncertainty, the spatial variability or correlation structure of *in situ* soil and rock properties are usually analysed through the autocorrelation function (Vanmarcke, 1977) or geostatistical methods (Matheron, 1971). Recent applications of these techniques in geotechnical engineering include characterisations of spatial variability in geological profiles (e.g. Dasaka & Zhang, 2012; Li *et al.*, 2016; Liu *et al.*, 2017) and soil properties (e.g. DeGroot & Baecher, 1993; Chiasson *et al.*, 1995; Wang & Chiasson, 2006; Liu & Chen, 2010; Jaksa & Fenton, 2000; Zhang & Chen, 2012; Firouzianbandpey *et al.*, 2014; Wang & Zhao, 2016). Meanwhile, the influence of spatial variability on the performance of geo-infrastructure have been investigated by Griffiths & Fenton (2009); Griffiths *et al.* (2009); Cho (2010); Cho & Park (2010); Kasama & Whittle (2011); Al-Bittar & Soubra (2014); Jiang *et al.* (2015); Li *et al.* (2015). Many of these studies involved probabilistic analyses using random field simulations, where assumptions of stationarity were usually adopted. In an actual site setting, however, these assumptions are not always satisfied without appropriate data transformation such as log transformation or Box-Cox transformation (Box & Cox, 1964) and residual analysis approaches. Indeed, practical applications of reliability methods at a particular project hinge on proper characterisation of the site-specific spatial correlation features of the concerned geotechnical properties.

In general, spatial correlations of geotechnical properties may demonstrate features of three-dimensional anisotropy, manifested by significantly different scales of fluctuation in various directions. Strong anisotropy effects often occur in rock masses with certain predominant joint sets, or natural soils with discontinuities or weak seams. Anisotropy may also result from the depositional or post-depositional processes including fluvial or colluvial processes, spatially-varying diagenesis, fault movements, deformation of shear zones and discontinuities (Zhu & Zhang, 2013). For example, previous researchers (e.g. Phoon, 1995; DeGroot, 1996; Phoon & Kulhawy, 1999a,b; El-Ramly *et al.*, 2003; Huber, 2013) have reported horizontal correlations of soil properties spanning longer distances than the vertical correlations. Recently, Zhu & Zhang (2013) discussed various patterns of two-dimensional anisotropic correlations, assuming the same correlation structure in the two lateral directions. They derived the corresponding functions of scales of fluctuation, and illustrated these through generating a series of random field models. However, there has been limited study to characterise the general three-dimensional correlation structures of *in situ* soil data using rigorous statistical methods, or discussions on how the spatial variability of geotechnical properties are related to the geological setting of a site.

This study extends the integrated framework by Liu *et al.* (2017) to evaluate the three-dimensional anisotropic correlation structures of soil properties at two sites in Hong Kong, where

Manuscript received...
*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University

in situ tests including standard penetration tests (SPT) and vane shear tests (VST) were conducted. Based on the restricted maximum likelihood (REML) method, the deterministic trend structures and correlation parameters of the residuals in three principal directions are obtained by maximising the log-likelihood function with an evolutionary algorithm. The framework also ensures that fundamental assumptions of REML (e.g. stationarity) are satisfied, thereby enhancing the reliability of statistical analyses and making them more representative of the site conditions in random field models. The correlation features, in terms of spatial dependence and range parameters, are elaborated in the context of the respective geological settings to illustrate the possible inter-relationships. Finally, the predictive capabilities of the proposed approach are demonstrated through comparisons with traditional curve fitting methods and other geostatistical approaches, showing the improved accuracies in estimating geotechnical properties at unsampled locations. More importantly, the associated uncertainties at unsampled locations can be quantified through estimates of the prediction variance, which provide useful information for reliability analyses or decisions regarding additional sampling points.

ANISOTROPIC COVARIANCE MODEL

Spatial random variables are often expressed as a combination of the deterministic trend structure and the residual effects. With $\boldsymbol{x}$ representing the spatial coordinates of sampled points, a general linear mixed regression model for spatial data, $\boldsymbol{z}(\boldsymbol{x})$, can be formulated by:

$$\boldsymbol{z}(\boldsymbol{x}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{X}\boldsymbol{\beta}$ represents the large scale trend, with $\mathbf{X}$ being the deterministic component matrix that contains information on spatial coordinates, and $\boldsymbol{\beta}$ being the vector of regression coefficients corresponding to the adopted trend structure (linear, quadratic, cubic, etc.). The residual, $\boldsymbol{\varepsilon}$, is a combination of the correlation structure (with smooth scale variation of variance σ_e^2) and a white noise process (with variance σ_n^2), since white noise effects are assumed not to correlate with distance. The covariance matrix of $\boldsymbol{\varepsilon}$ is related to the correlation structure, $\mathbf{R}$, by:

$$\mathbf{V} = \mathbf{Var}(\boldsymbol{\varepsilon}) = \sigma_e^2\mathbf{R} + \sigma_n^2\mathbf{I} = (\sigma_e^2 + \sigma_n^2)\left[s\mathbf{R} + (1-s)\mathbf{I}\right]$$
$$\text{where } 0 \le s = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_n^2} \le 1 \tag{2}$$

In equation (2), $\mathbf{I}$ is the identity matrix, and s represents the spatial dependence, which incorporates the white noise effect into the covariance model. In the current study, the Gaussian (squared exponential) model is adopted to describe the autocorrelation at different separation distances, i.e. components of $\mathbf{R}$ matrix. For example, in the one-dimensional case, the isotropic model is represented by:

$$R(h_{ij}) = \exp\left(-\frac{h_{ij}^2}{\theta^2}\right) \tag{3}$$

where h_{ij} is the separation distance between points i and j, and θ is the range parameter. The θ parameter is sometimes known as the autocovariance distance, and represents the separation distance at which the correlation R decays to $1/e$ (e.g., DeGroot & Baecher, 1993). An alternative parameter to describe the influence of a sample is the scale of fluctuation, δ (e.g., Phoon & Kulhawy, 1999a), which is taken as the separation distance where $R = 0.05$ (Elkateb *et al.*, 2003; Rue & Held, 2005). For the Gaussian function, $\delta \approx \sqrt{\pi}\theta$ is an approximate relationship between θ and δ. As will be shown later, the current study adopts the REML method to determine the trend coefficients ($\boldsymbol{\beta}$) and correlation parameters (s and θ) of the residuals.

An anisotropic correlation structure involves direction-dependent scales of fluctuation, characterised by the different θ (or δ) in various principal directions, rotation angles of the principal axes, and deformation angles between them. There are two types of spatial anisotropy, namely geometric anisotropy and zonal anisotropy. Geometric anisotropy occurs when the range (θ), but not the spatial dependence (s), varies in different directions. It is often adopted to describe directional correlation effects, which is conceptually similar to the differential stretching and rotation of the coordinates axes of an isotropic field (Rue & Held, 2005). On the other hand, zonal anisotropy also involves a directional variable s, and is more relevant when strong layering effects exist. This study mainly focuses on the geometric anisotropy within a particular soil layer. The following sections will introduce the corresponding formulations of two-dimensional and three-dimensional anisotropic spatial correlation structures, and determination of the associated parameters using the REML method.

Two-dimensional anisotropy

This section outlines the derivation of two-dimensional anisotropic correlation structures, which will then be generalised into three-dimensional forms in the next section. Vanmarcke (1977) proposed the elliptical expression for two-dimensional variations in scales of fluctuation, which was later adopted by Zhu & Zhang (2013). In their study, the spatial correlations in the lateral (x and y) directions are assumed to be the same, but can differ from that in the vertical (z) direction. The current study adopts a similar structure for two-dimensional anisotropy, but the mathematical representations are revised for subsequent generalisation to three dimensions. Some examples of the elliptical representations of two- and three-dimensional anisotropic models are shown in Fig. 1.

For a transversely anisotropic field, the range parameters in the principal directions are denoted as θ_1 and θ_2, respectively.

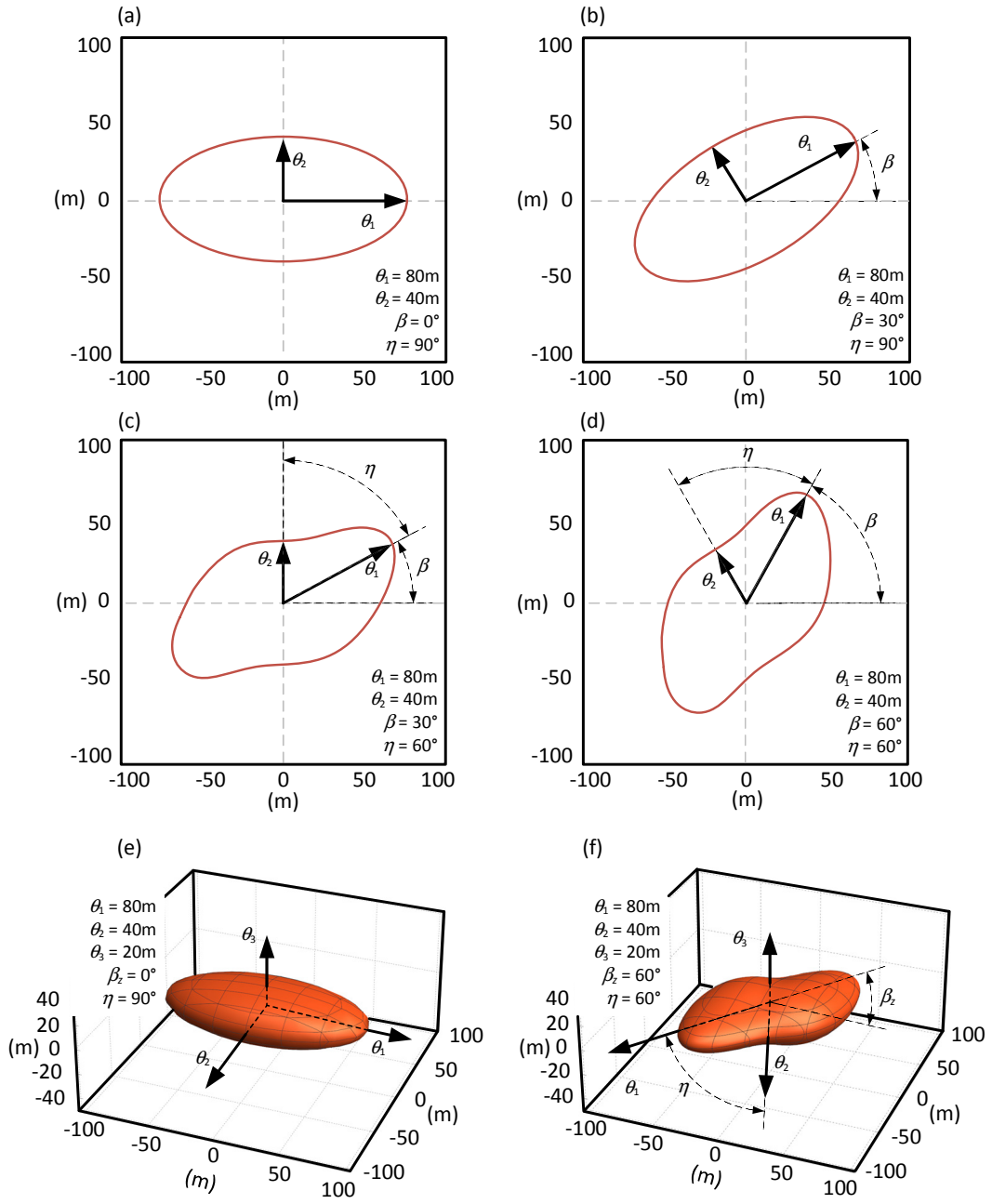

Fig. 1. Examples of anisotropic 2-D and 3-D spatial correlation models with different range parameters: (a) and (b) 2-D correlation without deformation; (c) and (d) 2-D correlation with deformation; (e) and (f) 3-D correlation

The principal axes may be rotated at an angle β relative to the x and z-axes. In general, θ_1 is not always larger than θ_2 under the current formulation, although the horizontal range is usually longer than the vertical range in the spatial correlation of many soils. If the principal directions remain orthogonal to each other, the Gaussian correlation function and the range in other orientations on the $x-z$ plane ($\phi \in [0, 2\pi]$) can be represented by:

$$R(h_{ij}) = \exp\left(-\frac{h_1^2}{\theta_1^2} - \frac{h_2^2}{\theta_2^2}\right)$$

$$\text{where} \quad \begin{Bmatrix} h_1 \\ h_2 \end{Bmatrix} = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{Bmatrix} h_x \\ h_z \end{Bmatrix} \tag{4a}$$

$$\theta(\phi) = \sqrt{\frac{\theta_1^2\theta_2^2(1+\tan^2\phi_r)}{\theta_2^2 + \theta_1^2\tan^2\phi_r}} \tag{4b}$$

where $\phi_r = \phi - \beta$, and h_x and h_z are the horizontal and vertical separation distances, respectively.

In cases where the two principal axes are not orthogonal to each other, the ellipse representing $\theta(\phi)$ becomes 'deformed', with the deformation dependent on the angle, η, between the two principal axes (Fig. 1(c)-(d)). This pattern of anisotropy can be found in rock masses that have undergone significant folding or tectonic movements, or those with two or more non-orthogonal joint sets. In this study, the transformation is represented by:

$$\begin{Bmatrix} h_1' \\ h_2' \end{Bmatrix} = \begin{bmatrix} \sin\eta & \cos\eta \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} h_1 \\ h_2 \end{Bmatrix} \tag{5}$$

where h_1' and h_2' represent the separation distances along the two principal axes in the new coordinate system. It should be noted, however, that in this new coordinate space, the total separation distance ($h_{ij}' = \sqrt{{h_1'}^2 + {h_2'}^2}$) becomes quantitatively different from the actual separation (h_{ij}) in the original space. Therefore, a scaling factor K is introduced in this study, to ensure that the separation distance remains the same as described by different coordinate spaces corresponding to η. This can be achieved by:

$$K^2({h_1'}^2 + {h_2'}^2) = h_1^2 + h_2^2 \tag{6}$$

Combining equations (5) and (6), and considering the parametric representations of h_1 and h_2 ($h_1 = h\cos\phi_r$ and $h_2 = h\sin\phi_r$):

$$K = \sqrt{\frac{1+\tan^2\phi_r}{(\sin\eta - \tan\phi_r\cos\eta)^2 + \tan^2\phi_r}} \tag{7}$$

Incorporating the rotation of coordinate system (β) and non-orthogonal principal axes (η), the Gaussian correlation function and range parameters become:

$$R_\eta(h_{ij}) = \exp\left[-\frac{(h_1\sin\eta + h_2\cos\eta)^2}{\theta_1^2} - \frac{(h_2)^2}{\theta_2^2}\right] \tag{8a}$$

$$\theta_\eta(\phi) = \frac{1}{K}\sqrt{\frac{\theta_1^2\theta_2^2(1+\tan^2\phi_r)}{\theta_2^2(\sin\eta - \tan\phi_r\cos\eta)^2 + \theta_1^2\tan^2\phi_r}} \tag{8b}$$

Three-dimensional anisotropy

The two-dimensional spatial correlation model can be versatile for transversely anisotropic geo-materials, such as sedimentary rocks or soils with a dominant set of bedding and stratification. It may not be adequate, however, for soils that demonstrate features of three-dimensional spatial variability, such as residual soils with properties heavily influenced by directional variations in the weathering process. Therefore, this study considers the additional dimension and proposes a more general spatial correlation structure, and its applications will be illustrated in later sections.

Extending the elliptical expression for a three-dimensional model, the corresponding range parameters would fall on the surface of an ellipsoid (Fig. 1(e)-(f)), and those on the principal axes are denoted as θ_1, θ_2 and θ_3. If the three principal axes remain orthogonal to each other, the Gaussian correlation function and the range in other orientations in the three-dimensional space ($\phi, \omega \in [0, 2\pi]$) can be represented by:

$$R(h_{ij}) = \exp\left(-\frac{h_1^2}{\theta_1^2} - \frac{h_2^2}{\theta_2^2} - \frac{h_3^2}{\theta_3^2}\right)$$

$$\text{where } \begin{Bmatrix} h_1 \\ h_2 \\ h_3 \end{Bmatrix} = \boldsymbol{\Omega}_3\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1 \begin{Bmatrix} h_x \\ h_y \\ h_z \end{Bmatrix}$$

$$\boldsymbol{\Omega}_1 = \begin{bmatrix} \cos\beta_z & \sin\beta_z & 0 \\ -\sin\beta_z & \cos\beta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\Omega}_2 = \begin{bmatrix} \cos\beta_y & 0 & -\sin\beta_y \\ 0 & 1 & 0 \\ \sin\beta_y & 0 & \cos\beta_y \end{bmatrix}$$

$$\boldsymbol{\Omega}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta_x & \sin\beta_x \\ 0 & -\sin\beta_x & \cos\beta_x \end{bmatrix} \tag{9a}$$

$$\theta(\phi,\omega) = \sqrt{\frac{\theta_1^2\theta_2^2\theta_3^2(1+\tan^2(\frac{\pi}{2}-\omega_r))}{\theta_2^2\theta_3^2\cos^2\phi_r + \theta_1^2\theta_3^2\sin^2\phi_r + \theta_1^2\theta_2^2\tan^2(\frac{\pi}{2}-\omega_r)}} \tag{9b}$$

where β_x, β_y and β_z represent the rotation angles about x, y and z-axes, respectively. The relationships between (ϕ, ω) and (ϕ_r, ω_r) in general three-dimensional conditions are more complex than those shown in equation (4b), and the details are presented in the Appendix.

The deformation angle between the principal axes can also be incorporated. As a simplified description of the concept, only the deformation on $\theta_1 - \theta_2$ plane ($\eta = \eta_{12}$) is illustrated herein, in which case the transformation of coordinate system can be represented as follows:

$$\begin{Bmatrix} h_1' \\ h_2' \\ h_3' \end{Bmatrix} = \begin{bmatrix} \sin\eta & \cos\eta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} h_1 \\ h_2 \\ h_3 \end{Bmatrix} \tag{10}$$

where h_1', h_2' and h_3' represent the separation distances along three principal axes in the new coordinate system. Similar to the two-dimensional case, a scaling factor K can be defined. Incorporating this into the three-dimensional rotations of coordinate system, the Gaussian correlation function and range parameters become:

$$K = \sqrt{\frac{1+\tan^2(\frac{\pi}{2}-\omega_r)}{(\cos\phi_r\sin\eta-\sin\phi_r\cos\eta)^2+\sin^2\phi_r+\tan^2(\frac{\pi}{2}-\omega_r)}} \tag{11a}$$

$$R_\eta(h_{ij}) = \exp\left[-\frac{(h_1\sin\eta+h_2\cos\eta)^2}{\theta_1^2}-\frac{(h_2)^2}{\theta_2^2}-\frac{(h_3)^2}{\theta_3^2}\right] \tag{11b}$$

$$\theta_\eta(\phi,\omega) = \frac{1}{K}\sqrt{\frac{\theta_1^2\theta_2^2\theta_3^2\left[1+\tan^2(\frac{\pi}{2}-\omega_r)\right]}{\theta_2^2\theta_3^2(\sin\eta\cos\phi_r-\cos\eta\sin\phi_r)^2+\theta_1^2\theta_3^2\sin^2\phi_r+\theta_1^2\theta_2^2\tan^2(\frac{\pi}{2}-\omega_r)}} \tag{11c}$$

DETERMINATION OF CORRELATION PARAMETERS

While the previous section presents the mathematical models for anisotropic correlation structures, the key to its practical application lies on the evaluation of the associated parameters that characterise site-specific spatial correlations in an actual site setting. Under the current framework, this refers to the determination of trend coefficients $\boldsymbol{\beta}$, and correlation parameters s, β, η and θ in different directions. In some previous studies (e.g. Cai *et al.*, 2016; Firouzianbandpey *et al.*, 2014), the sample data were analysed using the method of moments (MoM), which requires subjective decisions on the trend assumptions, bin size in residual analyses and curve fitting of semivariograms. These decisions and assumptions may not be robust, especially when spatial variability in three dimensions are involved. Also, the correlations in different directions were often decoupled in previous studies, and rotation of the principal directions cannot be considered rigorously since their directions are an *a priori* assumption in the decoupling process. In contrast, the current study adopts the REML method, implemented through an optimisation algorithm to determine all concerned parameters in the coupled three-dimensional formulation.

Liu *et al.* (2017) proposed an integrated framework to characterise spatial variability of geological profiles, combining REML with residual diagnostics and an optimal detrending process to enhance the robustness of residual analyses. The approach is extended in the current study to investigate the three-dimensional correlation structures of geotechnical properties, and the key elements are outlined herein.

In equation (1), the $\mathbf{X}$ matrix depends only on the spatial coordinates of the sampled points and not the correlation parameters. Meanwhile, the spatial correlation structure of $\boldsymbol{\varepsilon}$ can be determined by grouping the unknown parameters into a vector $\boldsymbol{\Theta}$. For example, for the three-dimensional anisotropic conditions described earlier, $\boldsymbol{\Theta} = \{s, \theta_1, \theta_2, \theta_3, \beta_x, \beta_y, \beta_z, \eta\}$. The values of these parameters can be obtained by maximising the following log-likelihood function with respect to $\boldsymbol{\Theta}$:

$$\begin{aligned} L(\boldsymbol{\Theta}|\boldsymbol{y}) = &-\frac{n-p}{2}\log(2\pi)-\frac{1}{2}\log|\mathbf{V}|-\frac{1}{2}\log|\mathbf{W}| \\ &-\frac{1}{2}\boldsymbol{y}^T\mathbf{V}^{-1}\mathbf{Q}\boldsymbol{y} \\ \text{where } \mathbf{W} &= \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} \\ \mathbf{Q} &= \mathbf{I}-\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^T\mathbf{V}^{-1} \end{aligned} \tag{12}$$

In equation (12), n is the number of data points, and p is the number of coefficients in the trend structure, and $\boldsymbol{y} = (\mathbf{I}\text{-}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\boldsymbol{z}$ is the vector of filtered dataset with the trend components filtered out. The $\mathbf{V}$ matrix depends on the correlation structure through equations (2), (8a) and (9a), which are in turn dependent on the parameters in $\boldsymbol{\Theta}$. The determination of $\boldsymbol{\Theta}$ can be treated as an optimisation problem, aiming to obtain the values of $\boldsymbol{\Theta}$ that maximise $L(\boldsymbol{\Theta}|\boldsymbol{y})$. This is achieved in the current study using an evolutionary algorithm known as Differential Evolution (Storn & Price, 1997), which is a robust and efficient technique that has been applied in a number of engineering problems (e.g. Leung *et al.*, 2011).

Once the covariance structure is determined, the trend coefficients ($\hat{\boldsymbol{\beta}}$) and predicted residuals ($\hat{\boldsymbol{\varepsilon}}$) can be estimated through generalised least squares (GLS). Predictions at unsampled locations, $\hat{\boldsymbol{z}}(\boldsymbol{x}_0)$, and the corresponding prediction variance, $\boldsymbol{\sigma}_{\boldsymbol{z}}^2(\boldsymbol{x}_0)$, can be estimated based on the Best Linear Unbiased Prediction (BLUP) technique (Atkinson *et al.*, 2008; Santra *et al.*, 2012):

$$\hat{\boldsymbol{z}}(\boldsymbol{x}_0) = \mathbf{X}_0\hat{\boldsymbol{\beta}} + \mathbf{K}^T\mathbf{V}^{-1}\hat{\boldsymbol{\varepsilon}} \tag{13a}$$

$$\begin{aligned} &\boldsymbol{\sigma}_{\boldsymbol{z}}^2(\boldsymbol{x}_0) = \text{diag}(\mathbf{K}_0-\mathbf{K}^T\mathbf{V}^{-1}\mathbf{K}+\mathbf{M}^T(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{M}) \\ &\text{where } \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\boldsymbol{z} \\ &\qquad \hat{\boldsymbol{\varepsilon}} = \boldsymbol{z}-\mathbf{X}\hat{\boldsymbol{\beta}} \end{aligned} \tag{13b}$$

In equation (13a), $\mathbf{X}_0$ is the deterministic component matrix of prediction locations, $\mathbf{K}$ represents the covariance matrix between observations and predictions, i.e., $\mathbf{K} = \text{cov}\{\boldsymbol{z}(\boldsymbol{x}), \boldsymbol{z}(\boldsymbol{x}_0)\}$, $\mathbf{K}_0 = \text{cov}\{\boldsymbol{z}(\boldsymbol{x}_0), \boldsymbol{z}(\boldsymbol{x}_0)^T\}$ and $\mathbf{M} = \mathbf{X}_0{}^T - \mathbf{X}^T\mathbf{V}^{-1}\mathbf{K}$.

The REML method determines the covariance of residuals together with the trend coefficients, and does not require subjective decisions on the bin size or curve fitting of autocorrelation function, so it can be applied to irregularly-spaced sampling points. However, the analyses can be affected by the selection of trend structure, i.e. the order of polynomial (linear, quadratic, cubic, etc.) used to describe the deterministic trend. As discussed by Liu *et al.* (2017), a higher order trend will fit the data better and reduce the residuals and their variance. However, an ever-increasing trend flexibility may lead to overfitting of the data, which means random

noise is included in the statistical model and compromises its predictive power. Meanwhile, stationarity conditions (constant mean and variance for residuals across the domain) are not always satisfied in actual site data, although these assumptions are commonly adopted in random finite element analyses.

To enhance robustness of the analyses on correlation parameters, the REML method is incorporated into an integrated framework which consists of the following major components: (1) the trend structure is determined by evaluating the significance of higher order terms through the Wald statistics, and their influence on the predictive power of the model through leave-one-out cross validation (Haslett & Hayes, 1998); and (2) data transformation is performed through log transformation or Box-Cox transformation (Box & Cox, 1964), with residual diagnostics to ensure normality and stationarity conditions are satisfied. In particular, for a positive-valued raw dataset $\boldsymbol{z}^*$, the transformed dataset, $\boldsymbol{z}$, can be represented by the following equation:

$$z_i = \begin{cases} \dfrac{(z_i^*)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(z_i^*) & \text{if } \lambda = 0 \end{cases} \tag{14}$$

where λ is the transformation parameter, estimated by minimising the residual sum of squares of $\boldsymbol{z}$. It should be noted that equation (14) itself does not guarantee normality or stationarity conditions. As discussed by Box & Cox (1964), the approach assumes that the transformed data have the highest likelihood to be normally distributed. The current study therefore incorporates residual diagnostics including the Kolomogrov-Smirnov test for normality and Breusch-Pagan test for constant variance conditions. The polynomial order of the trend structure is raised with λ revised in a sequential manner to ensure that both conditions are satisfied. Details on the integration of these components and implementation of the framework can be found in Liu *et al.* (2017).

CASE STUDIES

The proposed formulations are applied to characterise the three-dimensional spatial variability of geotechnical properties at two project sites in Hong Kong, namely the Hong Kong Convention and Exhibition Centre (HKCEC) and the Chek Lap Kok (CLK) site, with their details shown in Table 1. Borehole information at the two sites and *in situ* test results were obtained from geotechnical investigation reports of previous government projects in the areas, archived in the Civil Engineering Library maintained by the Civil Engineering and Development Department of the Hong Kong Government. Information on SPT blow counts (SPT-N values) and undrained shear strengths (s_u) from field vane shear tests is available at the HKCEC site and CLK site, respectively. The following sections present details of the spatial correlation analyses, while benefits of the proposed model are demonstrated through comparisons with other conventional characterisation models.

Descriptions of study sites

HKCEC is located in the northern side of Hong Kong Island, and was constructed on reclaimed land. The subsurface soil profile consists of fill material underlain by soft marine deposits, firm alluvium containing medium dense silty sand/sandy silt, and completely decomposed coarse-grained granite (CDG). Data of SPT-N values have been collected from 50 boreholes over an area of 200 m $\times$ 210 m, and only test results in CDG are analysed using the proposed approach in this study. Although it is a typical local practice to perform SPT in fine-grained materials such as the marine deposits and alluvial layers, the SPT-N values may not be representative of their engineering properties (British Standard, 2007). Meanwhile, the variability of the fill layer is expected to be largely dominated by the construction activities during reclamation, instead of the natural material variability.

The second study site, CLK, is located in the western waters of Hong Kong, north of the Lantau Island, where large-scale reclamation had been completed during construction of the Hong Kong International Airport. The offshore deposits in this area have been identified as the Hang Hau Formation (Strange & Shaw, 1986), which are one of the most geotechnically variable formations within the superficial deposits of Hong Kong. The study site has an area of 590 m $\times$ 280 m, and is underlain by soft clayey marine deposit, overlying alluvial deposits and completely to moderately decomposed granite. This study mainly focuses on the spatial variability of the marine deposit layer, which comprises soft to very soft silty clay, with thickness varying from 9 m to 18.5 m across the site. A total of 591 vane shear tests (VST) had been conducted at various depths below the seabed level, providing data on undrained shear strengths of the clayey soil. Only the peak shear strengths are used for characterisation of the three-dimensional spatial correlation structure.

Results of spatial correlation structures

The proposed approach is applied to investigate the spatial correlation features of CDG and marine clay deposits at the two sites, which include determination of range parameters and rotation angles in three directions. The deformation (η) model between principal axes is not incorporated in the subsequent analyses, mainly because this phenomenon is more influential in rock strata with distinct shear zones that may result from deformation, folding and/or tectonic movements. On the other hand, the CDG at HKCEC site results from the chemical weathering of granite which contains orthogonal joint sets in the region (GEO, 2007). The deformation processes associated with the η model are not dominant for sedimentary or residual soils in seismically inactive areas investigated in the current study.

Table 1. Site areas and details of *in situ* tests at the two case studies

Case study	*in situ* test data	Site area	Sample size (Number of tests)	Spacing of tests (m)	
				Lateral	Vertical
HKCEC	SPT-N values of saprolitic soil	210 m × 200 m	230	22.3 – 32.6	1.5 – 3.0
CLK	s_u of soft clay from VST	590 m × 280 m	591	1.7 – 111.1	1.0 – 3.0

Table 2. Spatial correlation parameters for two case studies

Case study	Trend structure	Transformation parameter	Spatial dependence	Range (m)			Rotation angle		
		λ	s	θ_1	θ_2	θ_3	β_z	β_y	β_x
HKCEC	Horizontal: quadratic Vertical: linear	-0.118	0.87	50.0	49.1	1.8	0°	241°	19°
CLK	Horizontal: quartic Vertical: quartic	0.339	0.76	1.3	12.0	11.6	116°	-	-

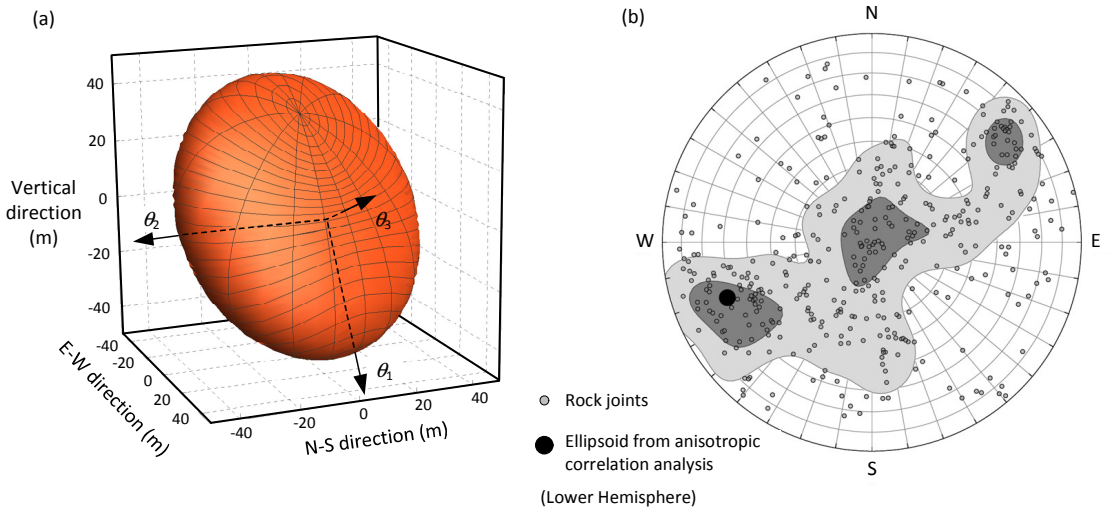

Fig. 2. Comparison between spatial variability characteristics of CDG and joint set orientations of underlying granite: (a) Orientation of ellipsoid from spatial correlation analyses; (b) Stereoplot of joint sets and ellipsoid at HKCEC

Following the procedures described earlier, the spatial coordinates of the *in situ* tests are utilised to evaluate the matrix $\mathbf{X}$, and the log-likelihood function in equation (12) is then maximised by Differential Evolution to obtain the correlation parameters $\boldsymbol{\Theta}$. Table 2 presents these parameters that characterise the spatial correlation features based on the SPT-N values and peak s_u from VST results at the two sites. For these two sites, the adopted trend structures do not include cross-terms between vertical and horizontal directions, as those terms are found to be insignificant, or would weaken the predictive power under the proposed approach. In other site settings, the cross-terms may be important and can be easily incorporated into the framework. Also, it should be noted that the subscripts 1, 2 and 3 for θ indicate the principal axes after rotations about x, y and z-axes (equation (9a)), but they do not imply the relative lengths of the different ranges. In other words, θ_1 is not necessary the longest (major) range, nor is θ_3 always the shortest (minor).

For the HKCEC case, the orientations of principal directions and the associated ranges are also shown graphically through the ellipsoid in Fig. 2(a). The ranges for two principal directions, θ_1 and θ_2 are significantly longer than that of θ_3, and the resulting ellipsoid resembles a planar surface. Since the CDG layer consists of residual soils that result from weathering of the underlying granite, the stereoplot of joint set directions for the granite, comprising 400 acoustic televiewer records from boreholes across the same site, is also shown in Fig. 2(b). The orientation of the ellipsoid obtained from the proposed approach, which represents the spatial correlation ranges, appears to correlate well with the major joint set of the parental rock underneath the CDG layer. This may be attributed to the fact that chemical weathering at the site is promoted by groundwater flow along pre-existing discontinuities within the rock mass and the subsequent wetting and drying processes due to changes in the groundwater table (GEO, 2007). The process is therefore concentrated along the discontinuities and proceeds inwards via microfractures. More variations in the weathering process are therefore expected in the direction normal to the joint plane, leading to a smaller range in terms of spatial correlation. As the original rock fabric is destroyed to become

CDG, the anisotropic spatial variations in soil properties follow similar patterns, manifested by the similar dip directions and angles of the ellipsoid in Fig. 2(a) and major joint set of parental rock material in Fig. 2(b).

For the marine clay deposit at CLK, only rotation about the vertical direction (β_z) is considered, as the layer thickness (9-18.5 m) is very small compared to the lateral extents of the domain (590 m $\times$ 280 m). Unlike the case of the weathering of granite, the soil layers above and below the marine clay in the CLK entail different origins. The rotations along lateral directions (β_x and β_y) of spatial correlation are therefore insignificant when considering the sedimentation process of the layer. As shown in Table 2, the vertical correlation range (θ_3) is very close to one of the horizontal ranges (θ_2), and significantly longer than that in the other horizontal direction (θ_1). Considering also the value of β_z, the orientation of the major principal direction appears to coincide with the dominate water flow direction in the region. Admittedly, however, it is difficult to confirm whether the erosion by the currents might have shaped or led to the directions of spatial correlation observed in the marine sediments.

Comparisons with conventional site characterisation methods

To illustrate the strengths of the proposed approach, three series of analyses are performed to compare its predictive capabilities with other site characterisation approaches. These include (1) the conventional curve fitting procedure where a simple polynomial function is adopted to represent variations of data values (SPT-N or s_u) with depth; (2) surface fitting where a three-dimensional polynomial function is adopted to include variations in the lateral directions, and (3) the proposed approach incorporating anisotropic spatial correlations and REML with residual diagnostics. In fact, the polynomial function in (2) is estimated through the GLS method, and is the same as the trend structure in approach (3), but the spatial distribution of the residuals are not considered in approach (2).

The predictive power of the proposed approach is assessed through the 'leave-one-out cross validation' method, which is performed by removing one observation at one time from the dataset, and then predicting its value using the remaining data while assuming the spatial correlation parameters remain unchanged (Haslett & Hayes, 1998). Conceptually, this same procedure also applies to the curve fitting and surface fitting analyses, but since they adopt one-to-one relationships between the spatial locations and the data value, the distribution of residuals are not considered explicitly in the cross validation.

Fig. 3 shows the predicted and measured values of SPT-N and peak s_u at the two sites. Compared to the curve fitting approach, surface fitting also considers variations in lateral directions and better represents the spatial distributions of test results. These two approaches, however, do not account for the aleatory uncertainty, or natural variations of *in situ* soil materials. In other words, the residual term (ε) is assumed to be zero in these approaches, and the predictive power is solely dependent on the polynomial trend function. On the other hand, the proposed approach appropriately defines the spatial correlation features using existing information (Table 2), which are then utilised for predictions at unsampled locations through equation (13a). Taking into consideration the natural variability of the residuals, this approach produces the most accurate estimates. Comparisons of the three approaches are also illustrated in Fig. 4, where the measured values are plotted against the predicted values. Predictions by the proposed method are closest to the 1:1 line with the highest value of coefficient of determination (R^2).

Comparisons with geostatistical method with two-dimensional spatial correlation model

To further illustrate the benefits of the proposed three-dimensional correlation analysis by REML, comparisons are also made with the two-dimensional correlation model that considers only vertical (z) and horizontal ($x-y$) spatial correlations, without rotation of the principal axes. The *in situ* test data at HKCEC are analysed by MoM, based on an approach similar to that by Firouzianbandpey *et al.* (2014): the correlation coefficients (sR) at different h_z are evaluated using data from each borehole separately, and the average values are obtained considering the results from all boreholes. Similar procedures are performed again to evaluate the correlation in the horizontal direction, by considering data from separate 1-m thick sublayers and then obtaining the average values. It should be noted that the same trend structures are adopted in REML and MoM analyses. The results are shown in Fig. 5, which also includes the curves fitted to the average moment estimates.

The estimates by the proposed REML approach are also shown in Figs. 5(a) and (b). It should be noted that two correlation functions, in x (east-west) and y (north-south) directions, are extracted by the proposed approach for comparison with a single function in $x-y$ (lateral) direction for the two-dimensional MoM analyses. Meanwhile, the measured SPT-N values are also plotted against the predicted values by MoM in Fig. 5, which shows more scattering than the three-dimensional analyses in Fig. 4. The prediction errors of the two approaches are presented in Fig. 5(d), which again shows better accuracy associated with the proposed three-dimensional approach.

The prediction capabilities of various methods are also compared in Table 3, through examining the corresponding root-mean-square-error (RMSE) and the mean absolute percentage error (MAPE), which are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\widetilde{\varepsilon_i}^2} \tag{15a}$$

$$\text{MAPE}\ (\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{|\widetilde{\varepsilon_i}|}{z_i} \tag{15b}$$

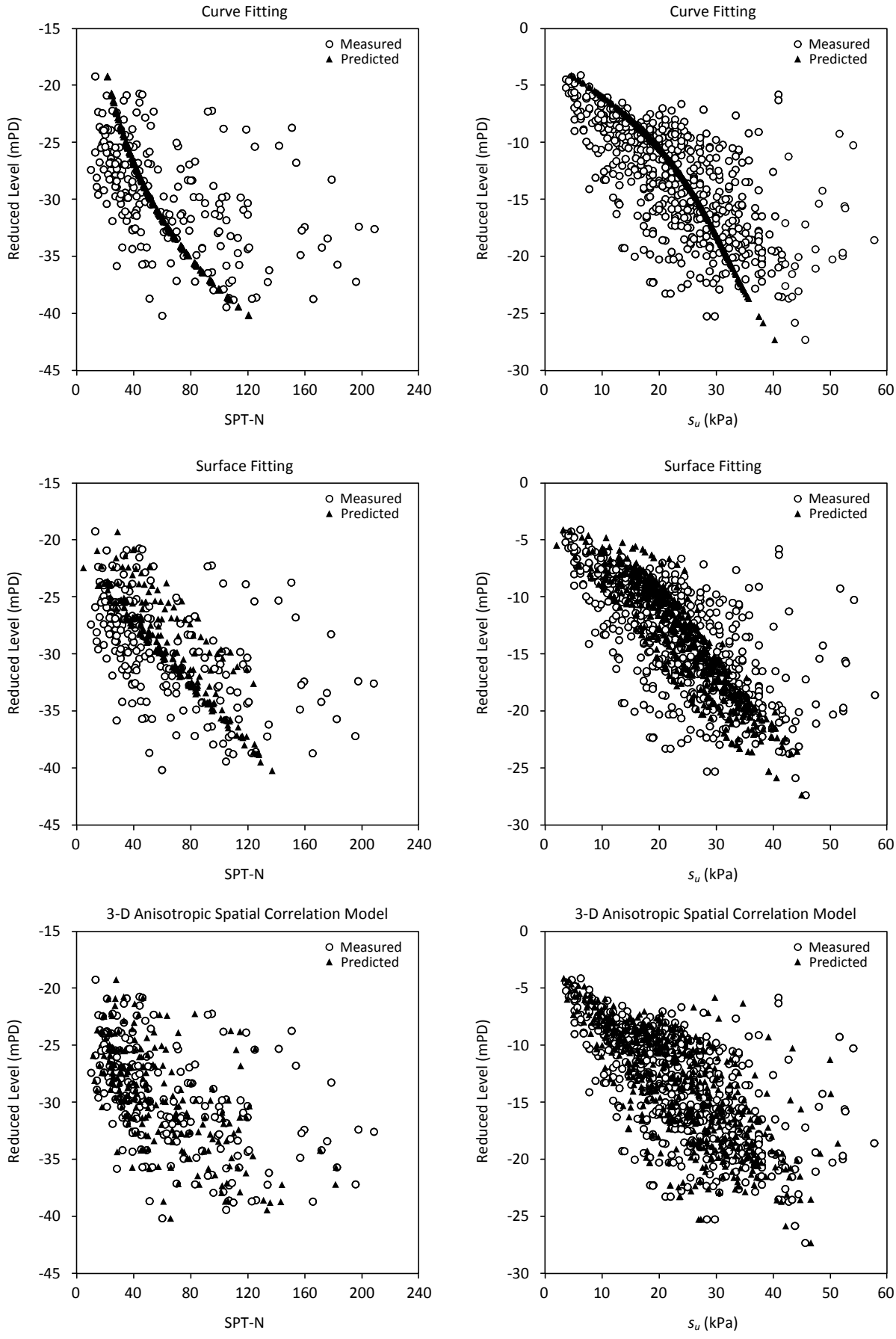

Fig. 3. Measured and predicted values of SPT-N at HKCEC (left) and s_u at CLK (right) with depth

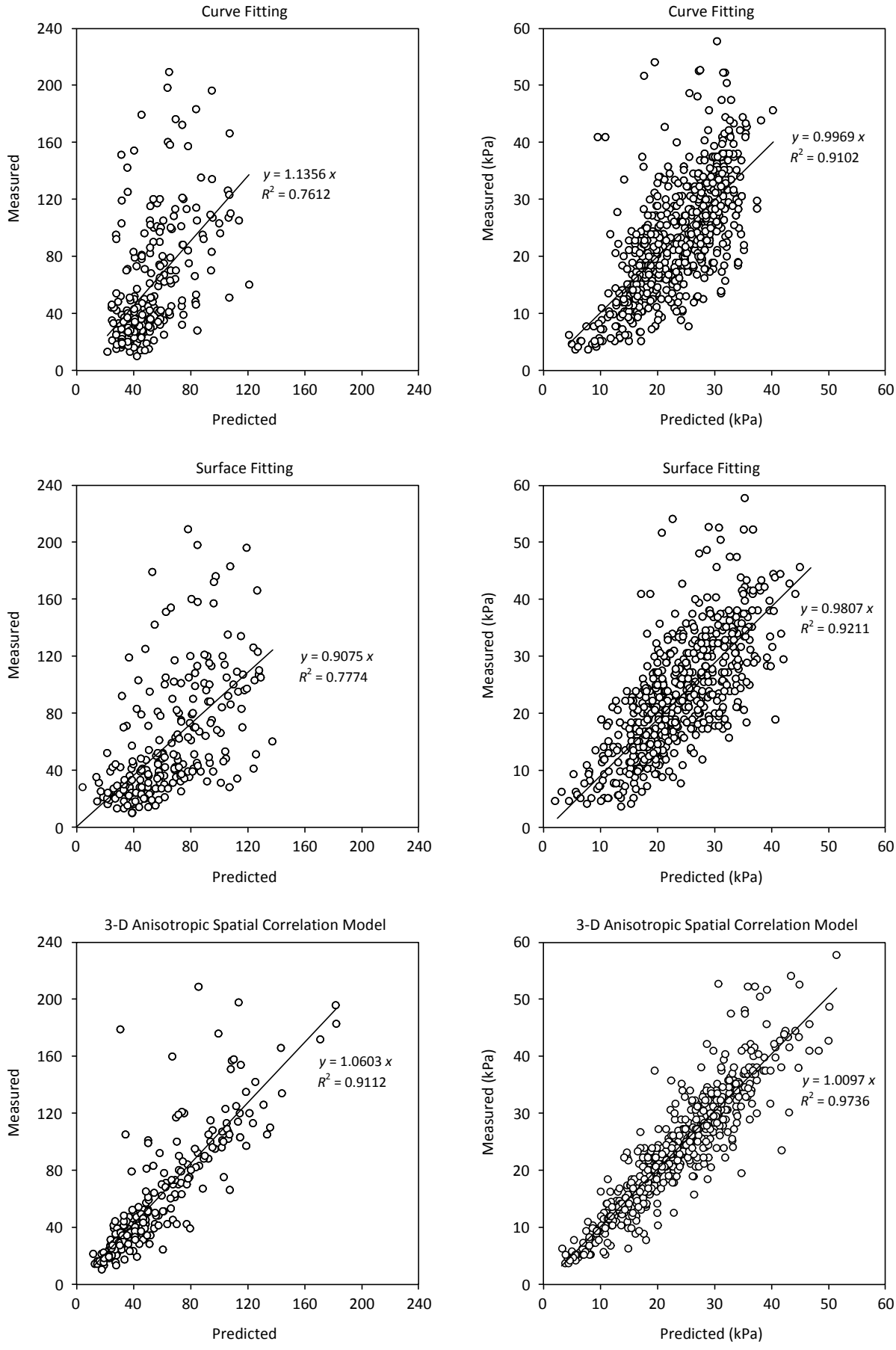

Fig. 4. Comparisons of measured and predicted values of SPT-N (left) and s_u (right) using three approaches

Table 3. Comparisons of prediction performance among four approaches

Case study	RMSE				MAPE (%)			
	Curve fitting	Surface fitting	2-D correlation by MoM	3-D correlation by REML	Curve fitting	Surface fitting	2-D correlation by MoM	3-D correlation by REML
HKCEC	37.07	35.62	28.22	22.48	48.58	59.43	39.28	21.12
CLK	7.62	7.16	7.12	4.14	31.23	30.83	25.68	13.66

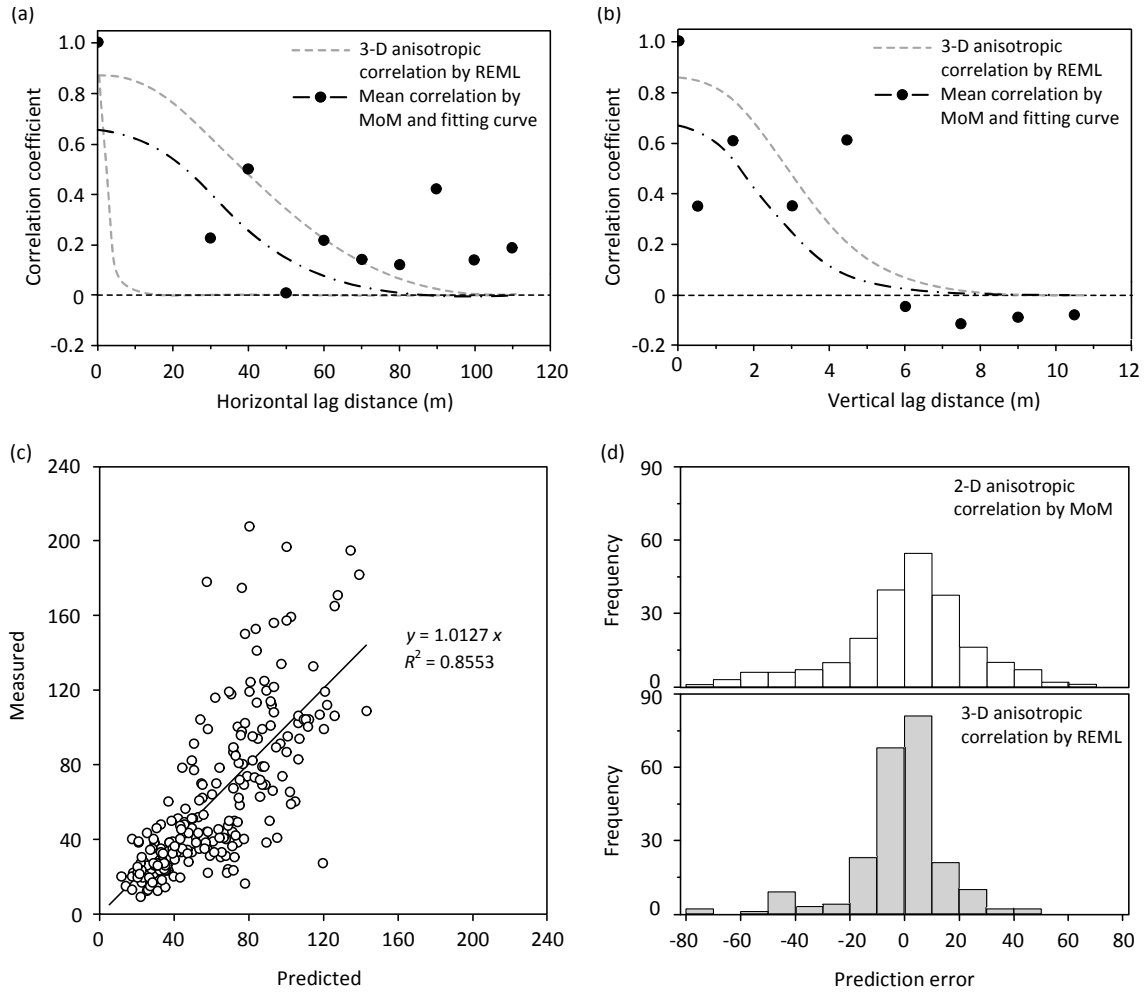

Fig. 5. Correlation structures in (a) horizontal direction and (b) vertical direction; (c) Measured and predicted SPT-N values using 2-D correlation model by MoM; (d) Prediction errors using 2-D and 3-D anisotropic spatial correlation models

where $\widetilde{\varepsilon_i}$ refers to the prediction error at sample location i. Both RMSE and MAPE indicate best performance by the proposed approach considering three-dimensional anisotropic spatial correlations. For example, the MAPE values are approximately 14-21% under the proposed approach, which are less than half of those by curve fitting and surface fitting approaches, and also significantly lower than the MoM analysis with two-dimensional correlation structure. The prediction errors are generally larger for SPT-N values in CDG (HKCEC) than s_u by VST in marine clay deposits (CLK). This may be partly attributed to occasional presence of corestones in saprolitic soils, which can introduce bias to the measured N values. More discussions on the influence of test data quality will be presented in later sections.

Prediction variance by proposed approach

The prediction variance (σ_z^2), estimated through equation (13b), represents the uncertainty associated with the predictions at unsampled locations, and may be interpreted as the confidence level in these estimates. Such uncertainty varies spatially across the three-dimensional domain, depending on the autocorrelation structure and locations of existing sampled points, and contains contributions from both uncertainties in deterministic trend structure and the residual effects.

Fig. 6 shows examples of the prediction variance contours for SPT-N values in the CDG layer at HKCEC. In general, the prediction variance is low near sampled locations and increases with distance away from data points, depending on the spatial correlation parameters. It is not straightforward to validate the estimated prediction variance contour, since the values of σ_z^2

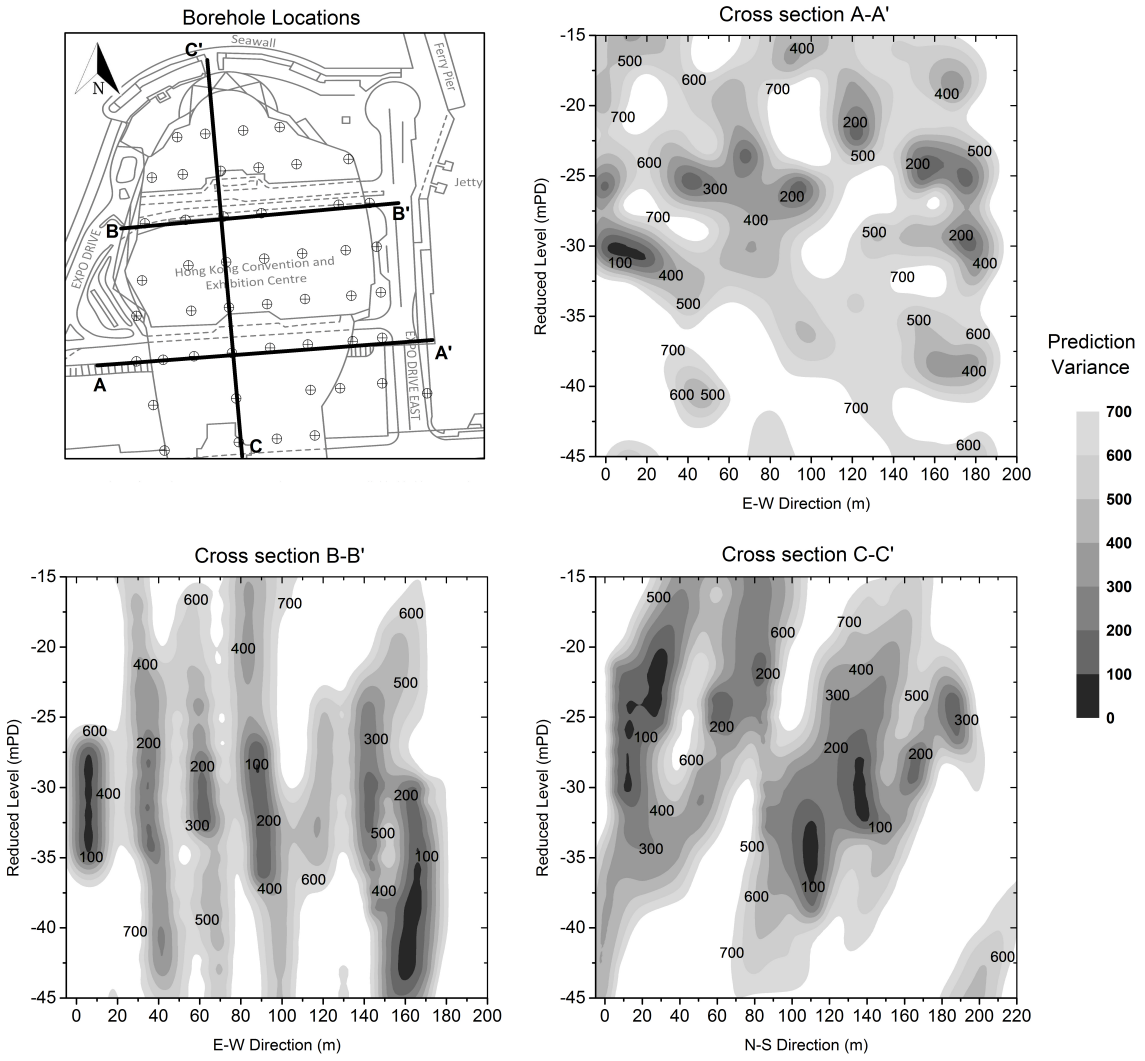

Fig. 6. Prediction variance in transformed space for SPT-N values at HKCEC

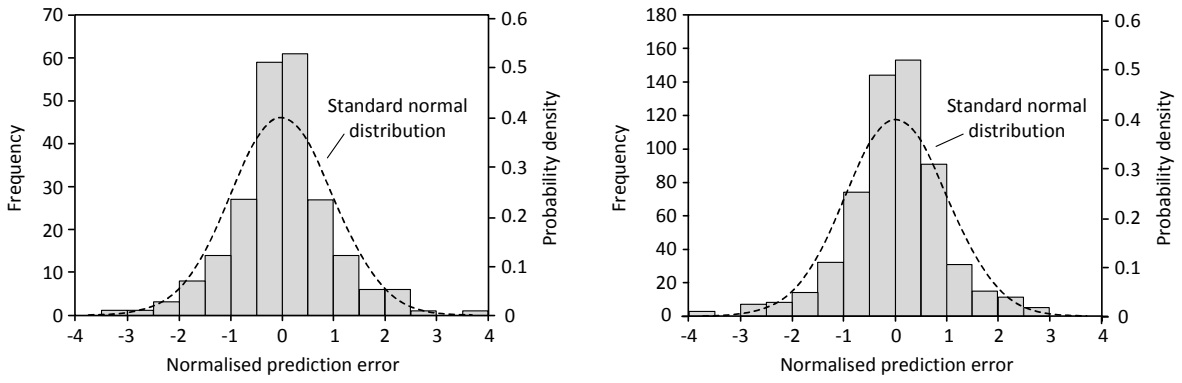

Fig. 7. Normalised prediction errors for SPT-N at HKCEC (left) and s_u at CLK (right)

are different at every location within the domain, but there can be only one measured z value and hence one prediction error at each sampled point. An indirect approach is to 'normalise' the prediction error obtained in leave-one-out cross validation ($\tilde{\varepsilon}$), by the value of σ_z estimated at every sample location. These normalised errors for the two sites are plotted as histograms in Fig. 7, and their distributions broadly follow the standard normal distribution. This means that on average, the errors

associated with $\hat{z}$ estimates at unsampled locations can be expected to follow a normal distribution, with variance denoted by the corresponding σ_z^2 at the location. This justifies the representation of uncertainty in the domain using the prediction variance contour.

The distribution of prediction variance over the three-dimensional domain may be incorporated into conditioned random field or geostatistical models (e.g. Leung & Lo, 2015; Lo & Leung, 2016) for probabilistic analyses of geotechnical performance, utilising the sample values, locations of existing information, as well as spatial correlation structure across the domain. The contours can also provide guidance regarding the necessity and locations of additional *in situ* tests or soil samples to achieve a specific level of confidence in the predictions of system performance.

DISCUSSIONS

The proposed approach utilises *in situ* test data to reveal the anisotropic correlation features of material properties in the three-dimensional domain. Yet, the importance of a good understanding of the site geological conditions cannot be overstated. In fact, as shown in earlier sections, the results from spatial correlation analysis should be interpreted together with the geological settings of the particular site. Also, the approach is not intended to be a replacement to any physical soil or rock testing methods. Instead, it is proposed to supplement the typical geotechnical investigation procedures, by providing added-value to the physical test results through evaluation of the uncertainty using rigorous geostatistical techniques.

In situ tests such as SPT may suffer from errors arising from non-standardised practice. Although SPT tests in Hong Kong are performed by mechanised hammer systems for improved efficiency and consistency, it is still possible that different operators may introduce inconsistencies into the testing results. These will be reflected, to a certain degree, through the 'noise' effects represented by the spatial dependence term (s). However, as for all statistical analyses, this approach will be inevitably affected by measurement errors.

In fact, the approach can also be applied to laboratory testing results, such as triaxial tests or oedometer tests, on samples obtained from a site. However, the number of samples retrieved for laboratory testing is usually limited, and one may argue that the accuracies of laboratory test results can also suffer from soil disturbance during the retrieval, transportation and handling processes of the specimens.

The current study adopts the Gaussian (squared exponential) model in the formulation of three-dimensional covariance structure. Under the same framework, the autocovariance structure can also be formulated readily using other models, such as the exponential, spherical, or the more flexible Matérn model (Matérn, 1960). If these models are implemented in REML for analyses of the above case studies, the differing function shapes may lead to different range parameters compared to those presented in Table 2. However, the subsequent estimates will be similar and the general findings from these case studies will not be affected. Also, potential modification of the approach may be achieved by replacing the elliptical function (e.g. equations (4b) and (9b)) by more irregular representations of correlation ranges in various orientations. This would, however, increase the complexity in application of the approach.

CONCLUSION

This paper presents the three-dimensional autocovariance model and its application to characterise the spatial variability of SPT-N values and undrained shear strengths using *in situ* soil test data. The Gaussian autocorrelation model is adopted and the model is implemented using the REML method under the framework by Liu *et al.* (2017), which ensures the stationarity of the residuals, and the determination of an optimal trend structure through rigorous geostatistical techniques.

The approach is applied to two sites in Hong Kong, and the spatial correlation features may be interpreted together with the local geological settings. These results supplement the physical tests in revealing the characteristics of variations within the subsurface domain. The approach is also shown to produce more accurate predictions of soil properties than the traditional site characterisation practice, through the additional consideration of spatial variations of residuals.

Under the proposed approach, uncertainty in the soil properties may be quantified and represented by the predication variance estimates, which can be useful input for probabilistic analyses of geotechnical performance. These estimates may also be considered in the derivation of cost-effective soil sampling strategies at project sites where pre-existing information is available.

ACKNOWLEDGEMENTS

The work presented in this paper is financially supported by the Research Grants Council of the Hong Kong Special Administrative Region (Project No. 25201214). Mr. M.K. Lo has provided valuable suggestions on the formulation and implementation of the anisotropic covariance model. Also, the authors would like to acknowledge the permission of the Civil Engineering and Development Department, the Government of Hong Kong Special Administrative Region, to present analyses of data obtained from the Civil Engineering Library.

APPENDIX

Rotations of the three-dimensional ellipsoid of range parameters can be illustrated by considering the vector of separation distance in spherical coordinates (h, ω, ϕ) and Cartesian coordinates (x, y, z), as shown in Fig. 8. Without rotation, $h_x = h \sin\omega \cos\phi$, $h_y = h \sin\omega \sin\phi$ and $h_z = h \cos\omega$. In this study, rotation about z-axis is first considered, followed by rotations

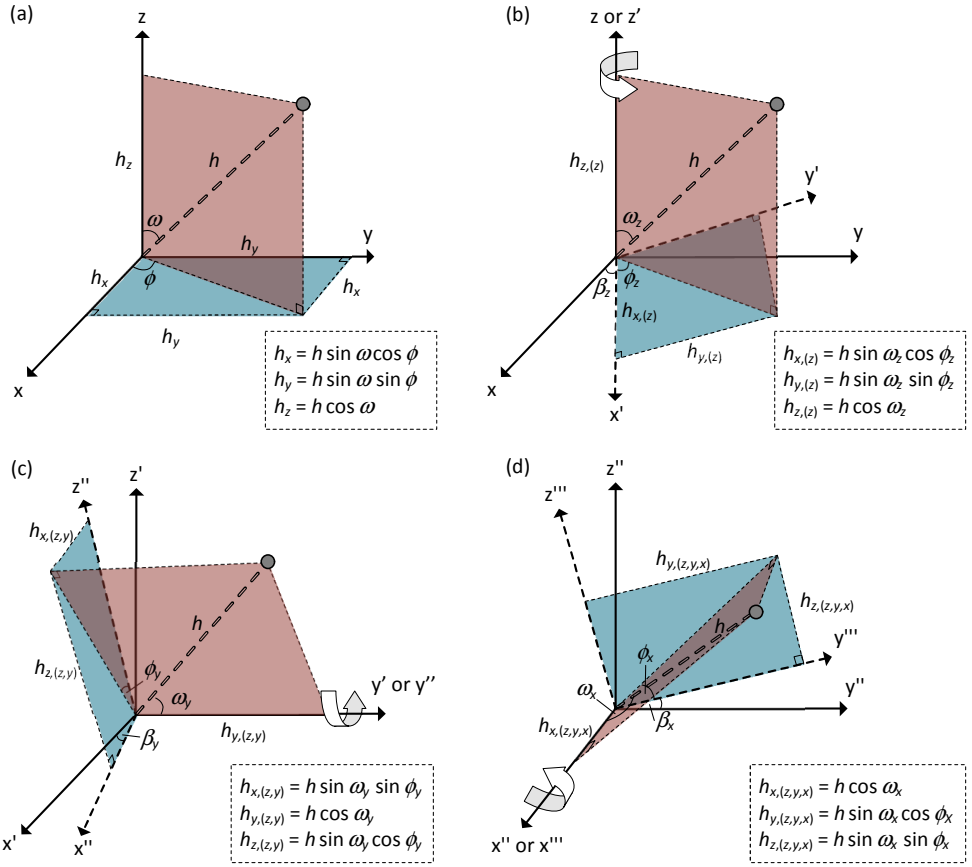

Fig. 8. (a) Original coordinate system; (b) Rotation about z-axis; (c) Rotation about y'-axis; (d) Rotation about x''-axis

about y-axis and then x-axis. Upon rotation about z-axis, the three components, $h_{x,(z)}$, $h_{y,(z)}$ and $h_{z,(z)}$, are represented by:

$$\begin{aligned} h_{x,(z)} &= h \sin\omega_z \cos\phi_z \\ h_{y,(z)} &= h \sin\omega_z \sin\phi_z \\ h_{z,(z)} &= h \cos\omega_z \\ \text{where } \omega_z &= \cos^{-1}\left[\frac{h_z}{h}\right] = \omega \\ \phi_z &= \tan^{-1}\left[\frac{h_y}{h_x}\right] - \beta_z = \phi - \beta_z \end{aligned} \tag{16}$$

It should be noted that after the first rotation, the remaining two axes should be labelled as x' and y', as they now differ from those in the original system (Fig. 8(b)). Subsequent rotations can be incorporated sequentially, whereby the associated ω_y, ω_x, ϕ_y, ϕ_x (subscripts labelled as x and y, instead of x', y' or x'', y'', for convenience) in the new system can be derived as:

$$\begin{aligned} \omega_y &= \cos^{-1}\left[\frac{h_{y,(z)}}{h}\right] \\ \phi_y &= \tan^{-1}\left[\frac{h_{x,(z)}}{h_{z,(z)}}\right] - \beta_y \\ \omega_x &= \cos^{-1}\left[\frac{h_{x,(z,y)}}{h}\right] \\ \phi_x &= \tan^{-1}\left[\frac{h_{z,(z,y)}}{h_{y,(z,y)}}\right] - \beta_x \end{aligned} \tag{17}$$

where $h_{x,(z,y)}$, $h_{y,(z,y)}$ and $h_{z,(z,y)}$ represent the three components under the coordinate system rotated about z and then y'-axes. They can be derived in similar manner as shown in equation (16). Finally, ω_r and ϕ_r, which incorporate rotations about the three axes, can be obtained by combining equations (16), (17) and the definitions of $h_{x,(z,y,x)}(= h_1)$, $h_{y,(z,y,x)}(= h_2)$ and $h_{z,(z,y,x)}(= h_3)$ (Fig. 8) in the new coordinate system:

$$\begin{aligned} \omega_r &= \cos^{-1}\left[\frac{h_{z,(z,y,x)}}{h}\right] = \cos^{-1}\left\{\sin\left[\cos^{-1}(\sin A \sin B)\right] \sin\left[\tan^{-1}(\tan A \cos B) - \beta_x\right]\right\} \\ \phi_r &= \tan^{-1}\left[\frac{h_{y,(z,y,x)}}{h_{x,(z,y,x)}}\right] = \tan^{-1}\left\{\tan\left[\cos^{-1}(\sin A \sin B)\right] \cos\left[\tan^{-1}(\tan A \cos B) - \beta_x\right]\right\} \\ \text{where } A &= \cos^{-1}\left[\sin\omega \sin(\phi - \beta_z)\right] \\ B &= \tan^{-1}\left[\tan\omega \cos(\phi - \beta_z)\right] - \beta_y \end{aligned} \tag{18}$$

NOTATION

h_1, h_2, h_3	separation distance along orthogonal principal axes
h_1', h_2', h_3'	separation distance along non-orthogonal principal axes
h_{ij}	separation distance between points i and j
h_x, h_y, h_z	separation distances along x, y and z-axes
$\mathbf{K}$	covariance matrix between observations and predictions
$\mathbf{K}_0$	covariance matrix between predictions
K	scaling factor
n	number of data points
$\mathbf{R}$	autocorrelation matrix
$R(\cdot)$	Gaussian autocorrelation function
s	spatial dependence
$\mathbf{V}$	covariance matrix
$\mathbf{X}$	design matrix of observations
$\mathbf{X}_0$	design matrix of predictions
$\boldsymbol{x}$	vector of sample locations
$\boldsymbol{x}_0$	vector of prediction locations
$\boldsymbol{y}$	vector of filtered dataset
$\boldsymbol{z}$	transformed dataset
$\boldsymbol{z}^*$	raw dataset
$\hat{\boldsymbol{z}}$	predictions at unsampled locations
$\boldsymbol{\beta}$	vector of trend coefficients
$\hat{\boldsymbol{\beta}}$	vector of estimated trend coefficients
β, β_x, β_y, β_z	rotation angles about coordinate axes
δ	scale of fluctuation
$\boldsymbol{\varepsilon}$	residuals
$\hat{\boldsymbol{\varepsilon}}$	predicted residuals at unsampled locations
$\widetilde{\varepsilon_i}$	prediction error at sample location i
η	deformation angle between principal axes
$\boldsymbol{\Theta}$	vector of correlation parameters
θ	range parameter
λ	Box-Cox transformation parameter
σ_e^2	error variance in the regression model
σ_n^2	white noise process
$\boldsymbol{\sigma}_{\boldsymbol{z}}^2$	prediction variance at unsampled locations
ϕ, ω, ϕ_r,ω_r	spatial orientations
$\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{\Omega}_3$	rotation matrix
RMSE	root-mean-square-error
MAPE	mean absolute percentage error

REFERENCES

Al-Bittar, T. & Soubra, A.-H. (2014). Probabilistic analysis of strip footings resting on spatially varying soils and subjected to vertical or inclined loads. *Journal of Geotechnical and Geoenvironmental Engineering* **140**, No. 4, 04013043, doi:10.1061/(ASCE)GT.1943-5606.0001046.

Atkinson, P. M., Pardo-Iguzquiza, E. & Chica-Olmo, M. (2008). Downscaling cokriging for super-resolution mapping of continua in remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing* **46**, No. 2, 573–580.

Box, G. E. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* , 211–252.

British Standard (2007). *Eurocode 7 – Geotechnical Design – part 2: Ground investigation and testing*. British Standard Institution, doi: BS EN 1997-2:2007.

Cai, G., Lin, J., Liu, S. & Puppala, A. J. (2016). Characterization of spatial variability of CPTU data in a liquefaction site improved by vibro-compaction method. *KSCE Journal of Civil Engineering* , 1–11.

Chiasson, P., Lafleur, J., Soulié, M. & Law, K. T. (1995). Characterizing spatial variability of a clay by geostatistics. *Canadian Geotechnical Journal* **32**, No. 1, 1–10.

Cho, S. E. (2010). Probabilistic assessment of slope stability that considers the spatial variability of soil properties. *Journal of Geotechnical and Geoenvironmental Engineering* **136**, No. 7, 975–984, doi:10.1061/(ASCE)GT.1943-5606.0000309.

Cho, S. E. & Park, H. C. (2010). Effect of spatial variability of cross-correlated soil properties on bearing capacity of strip footing. *International journal for numerical and analytical methods in geomechanics* **34**, No. 1, 1–26.

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

DeGroot, D. J. (1996). Analyzing spatial variability of in situ soil properties. In *Uncertainty in the Geologic Environment: From Theory to Practice*, vol. 1, pp. 210–238.

DeGroot, D. J. & Baecher, G. B. (1993). Estimating autocovariance of in-situ soil properties. *Journal of Geotechnical Engineering* **119**, No. 1, 147–166.

El-Ramly, H., Morgenstern, N. & Cruden, D. (2003). Probabilistic stability analysis of a tailings dyke on presheared clay shale. *Canadian Geotechnical Journal* **40**, No. 1, 192–208.

Elkateb, T., Chalaturnyk, R. & Robertson, P. K. (2003). An overview of soil heterogeneity: quantification and implications on geotechnical field problems. *Canadian Geotechnical Journal* **40**, No. 1, 1–15.

Firouzianbandpey, S., Griffiths, D. V., Ibsen, L. B. & Andersen, L. V. (2014). Spatial correlation length of normalized cone data in sand: case study in the north of Denmark. *Canadian Geotechnical Journal* **51**, No. 8, 844–857.

GEO (2007). *Engineering geological practice in Hong Kong*. Geotechnical Engineering Office, Civil Engineering and Development Department, The Government of Hong Kong Special Administrative Region.

Griffiths, D. V. & Fenton, G. A. (2009). Probabilistic settlement analysis by stochastic and random finite-element methods. *Journal of Geotechnical and Geoenvironmental Engineering* **135**, No. 11, 1629–1637, doi:10.1061/(ASCE)GT.1943-5606.0000126.

Griffiths, D. V., Huang, J. & Fenton, G. A. (2009). Influence of spatial variability on slope reliability using 2-D random fields. *Journal of Geotechnical and Geoenvironmental Engineering* **135**, No. 10, 1367–1378, doi:10.1061/(ASCE)GT.1943-5606.0000099.

Haslett, J. & Hayes, K. (1998). Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* , 201–215.

Huber, M. (2013). *Soil variability and its consequences in geotechnical engineering*. Ph.D. thesis, University of Stuttgart.

Jaksa, M. B. & Fenton, G. A. (2000). Random field modeling of CPT data. *Journal of Geotechnical and Geoenvironmental Engineering* **126**, No. 12, 1212–1216.

Jiang, S.-H., Li, D.-Q., Cao, Z.-J., Zhou, C.-B. & Phoon, K.-K. (2015). Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. *Journal of Geotechnical and Geoenvironmental Engineering* **141**, No. 2, 04014096, doi:10.1061/(ASCE)GT.1943-5606.0001227.

Kasama, K. & Whittle, A. J. (2011). Bearing capacity of spatially random cohesive soil using numerical limit analyses. *Journal of Geotechnical and Geoenvironmental Engineering* **137**, No. 11, 989–996, doi:10.1061/(ASCE)GT.1943-5606.0000531.

Leung, Y. F. & Lo, M. K. (2015). Anisotropic spatial correlation of compression indices of marine clay deposits in Hong Kong. In *International Conference on Soft Ground Engineering, Singapore*, pp. 581–588.

Leung, Y. F., Soga, K. & Klar, A. (2011). Multi-objective foundation optimization and its application to pile reuse. In *Geo-Frontiers 2011*, pp. 75–84, doi:10.1061/41165(397)9.

Li, J., Cassidy, M. J., Huang, J., Zhang, L. & Kelly, R. (2016). Probabilistic identification of soil stratification. *Géotechnique* **66**, No. 1, 16–26, doi:10.1680/jgeot.14.P.242.

Li, J., Tian, Y. & Cassidy, M. J. (2015). Failure mechanism and bearing capacity of footings buried at various depths in spatially random soil. *Journal of Geotechnical and Geoenvironmental Engineering* **141**, No. 2, 04014099, doi:10.1061/(ASCE)GT.1943-5606.0001219.

Liu, C.-N. & Chen, C.-H. (2010). Estimating spatial correlation structures based on CPT data. *Georisk* **4**, No. 2, 99–108.

Liu, W. F., Leung, Y. F. & Lo, M. K. (2017). Integrated framework for characterization of spatial variability of geological profiles. *Canadian Geotechnical Journal* **54**, No. 1, 47–58.

Lo, M. K. & Leung, Y. F. (2016). Probabilistic analyses of slopes and footings with spatially variable soils considering cross-correlation and conditioned random field. *Journal of Geotechnical and Geoenvironmental Engineering* , Accepted.

Matérn, B. (1960). Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden från Statens Skogsforskningsinstitut* **49**, No. 5, 1–144.

Matheron, G. (1971). *The theory of regionalized variables and its application*. France: Fontainebleau.

Phoon, K.-K. (1995). *Reliability-based design of foundations for transmission line structures*. Ph.D. thesis, Cornell University.

Phoon, K.-K. & Kulhawy, F. H. (1999a). Characterization of geotechnical variability. *Canadian Geotechnical Journal* **36**, No. 4, 612–624, doi:10.1139/t99-038.

Phoon, K.-K. & Kulhawy, F. H. (1999b). Evaluation of geotechnical property variability. *Canadian Geotechnical Journal* **36**, No. 4, 625–639, doi:10.1139/t99-039.

Rue, H. & Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Santra, P., Das, B. S. & Chakravarty, D. (2012). Spatial prediction of soil properties in a watershed scale through maximum likelihood approach. *Environmental Earth Sciences* **65**, No. 7, 2051–2061.

Storn, R. & Price, K. (1997). Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11**, No. 4, 341–359.

Strange, P. & Shaw, R. (1986). *Geology of Hong Kong Island and Kowloon: 1: 20,000 sheets 11 & 15*. Geotechnical Control Office, Civil Engineering Services Dept.

Vanmarcke, E. H. (1977). Probabilistic modeling of soil profiles. *ASCE Journal of Geotechnical Engineering Division* **103**, No. GT11, 1227–1246.

Wang, Y. & Zhao, T. (2016). Interpretation of soil property profile from limited measurement data: a compressive sampling perspective. *Canadian Geotechnical Journal* **53**, No. 9, 1547–1559.

Wang, Y.-J. & Chiasson, P. (2006). Stochastic stability analysis of a test excavation involving spatially variable subsoil. *Canadian Geotechnical Journal* **43**, No. 10, 1074–1087.

Zhang, L. & Chen, J.-J. (2012). Effect of spatial correlation of standard penetration test (SPT) data on bearing capacity of driven piles in sand. *Canadian Geotechnical Journal* **49**, No. 4, 394–402.

Zhu, H. & Zhang, L. M. (2013). Characterizing geotechnical anisotropic spatial variations using random field theory. *Canadian Geotechnical Journal* **50**, No. 7, 723–734.