RELIABILITY ASSESSMENT OF SLOPES CONSIDERING SAMPLING INFLUENCE AND SPATIAL VARIABILITY BY SOBOL' SENSITIVITY INDEX

M.K. Lo[1], and Y.F. Leung[2]

ABSTRACT

This paper presents an extended formulation of the Sobol' sensitivity index for geotechnical reliability assessments involving spatially variable soil properties. It incorporates the subsurface spatial correlation structure with the response surface method, which is then assimilated into the context of the Sobol' index approach. A Sobol' index map can be generated for the entire subsurface domain, identifying the sensitive zones which also represent the optimal sampling locations. In addition, the approach allows the derivation of the mean and variance of system response conditional to any sample value, without the needs to conduct separate conditional random field simulations. This is adopted for the assessment of reliability of slopes, where design charts are established for cases where a single sample is obtained within slopes of c_u or $c - \phi$ soils, with various conditions of geometries and spatial variability. The approach can also be applied to multiple sampling points, thereby facilitating a feedback mechanism where the planning of geotechnical investigation and evaluation of performance uncertainty can be considered in a holistic manner.

Keywords: Slope stability, probabilistic analyses, conditional random field, sampling location, Sobol' sensitivity index

[1]Ph.D. Candidate, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.

[2]Assistant Professor, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong. E-mail: andy.yf.leung@polyu.edu.hk

INTRODUCTION

Reliability of geotechnical systems depends heavily on the uncertainty of soil and rock properties. Recently, reliability approaches such as first-order second-moment (FOSM) method (Duncan 2000) and the Hasofer-Lind approach (first-order reliability method) (Hasofer and Lind 1974) have gained popularity in the field of geotechnical engineering. However, many of these methods do not explicitly consider the spatial variations in material properties, which is an important source of uncertainty in geotechnical engineering. For example, the significance of such has been discussed by Christian and Baecher (2011), who also suggested that traditional assumptions of perfect (or no) spatial correlations in geotechnical properties do not always yield conservative estimates. The characterization of spatial variability in geotechnical properties has been discussed in Phoon and Kulhawy (1999), Baecher and Christian (2003), Liu et al. (2017b), Liu and Leung (2017), etc. A spatially variable soil profile can be modeled by the random field theory (Vanmarcke 1984), which considers the geotechnical property as a set of spatially correlated random variables, and its applications include shallow foundations (Fenton and Griffiths 2003; Al-Bittar and Soubra 2014), slope stability (Cho 2010; Jiang et al. 2015), soil liquefaction (Popescu et al. 2005), etc. Meanwhile, the geotechnical profession has long recognized the importance of obtaining samples or conducting *in situ* tests at 'representative' locations, and the spatial uncertainty can be reduced considerably when such information is incorporated into the probabilistic analyses (Lloret-Cabot et al. 2012; Li et al. 2016a). Accordingly, conditional random field modelling is a numerical approach that can assess the influence of known sample values at designated locations. In particular, some recent studies quantified the reductions in performance uncertainty considering the available sample values and their locations. These include the probabilistic assessments of footing settlements by Lo and Leung (2017a), and those for slope stability by Li et al. (2016b) and Liu et al. (2017a).

Despite being a powerful approach to utilize existing field data, conditional random field modelling suffers from some key limitations from the risk management perspective.

Essentially, the approach requires sample values and locations to be specified prior to the analyses, and each probabilistic assessment involves hundreds to thousands of realizations, based on one set of such samples. This exercise becomes impractical if the objectives are to explore multiple options for future sampling locations, and to evaluate how the additional, currently unknown, sample values may affect the performance uncertainty.

A complete assessment of the performance uncertainty should be based on all possible random field realizations. Also, it is desirable to develop an efficient strategy to obtain the optimal sampling pattern tailored to the specific application, allowing assessments of the influence of future sample values to the overall system reliability. To this end, this paper extends the formulation of Sobol' sensitivity index for geotechnical reliability assessments involving spatial variability in soil properties. It facilitates a feedback mechanism so that the planning of geotechnical investigation and evaluation of performance uncertainty can be considered in a holistic manner. While the approach is potentially applicable to various geotechnical problems, this study focuses on slope stability and presents design charts to illustrate the relationships between sample values and associated uncertainty in factor of safety.

RELIABILITY ASSESSMENT THROUGH SENSITIVITY ANALYSIS BY SOBOL' INDEX

Sobol' sensitivity index is a probabilistic tool developed by Sobol' (2001) to assess the influence of each input parameter in a physical model. In general, each model parameter can be regarded as a random variable. A Sobol' index can then be associated with each parameter, and used to quantify its contribution to the variance of the model response. This concept has been applied to a number of geotechnical applications, including investigations on the parameters affecting the consolidation process (Houmadi et al. 2012) and ground settlements induced by tunneling (Miro et al. 2014). Others have combined the Sobol' index approach with the response surface (or surrogate modeling) method to enhance the computational efficiency, and studied the parameters associated with footing displacements and stability of pressurized tunnels (Mollon et al. 2011).

These previous studies mainly focused on the influence of parameters in 'deterministic' models, where spatial variability of the concerned parameters was not considered. In fact, the Sobol' index was originally developed for independent input variables, and cannot be directly applied to spatially correlated fields of soil properties. Therefore, this study extends the original theory for applications to correlated random variables, which represent the correlated properties at various locations. The Sobol' index can be evaluated for each location, resulting in a 'Sobol' index map' that reveals the relative importance of soil properties at all location to the system response. The maximum index value corresponds to the most influential location, which is equivalent to the optimal sampling location.

The proposed Sobol' index approach enables site-specific geotechnical reliability assessments to be performed with considerations of spatial variability and information from soil samples. Details of the assessment strategy are shown in a flowchart in Fig. S1, while the key elements are briefly described herein. Based on existing soil samples or experience at the site, the mean, variance and spatial correlation features of the parameters can be estimated. A set of unconditional (or conditional if prior samples are available) random field analyses can be performed for system response g. The Sobol' indices throughout the subsurface domain are then evaluated to (1) identify optimal locations for additional samples; and (2) derive the relationships between system reliability (mean and variance of g) and potential values of the future samples (referred to as 'sensitivity functions' herein). These can provide guidance on the adequacy of the design, and whether more samples should be planned to further reduce the uncertainty or risk levels. In this process, only one set of random field simulations is required unless major revisions of the original design are involved. In the following sections, the formulation of Sobol' index approach for spatially variable soils are presented, together with its application to single and multiple sample points in slope stability assessments.

FORMULATION FOR GEOTECHNICAL RELIABILITY ASSESSMENT

Single sample point

In general, the properties of a spatially variable soil can be represented by a trend component and the residuals, or deviations from the trend:

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{z} = \{z_1, z_2, \ldots, z_d\}^{\mathrm{T}}$ is a vector of soil properties at d different locations. $\boldsymbol{\varepsilon}$ represents the random deviation of the property at each location, and has a constant variance σ_z^2. $\boldsymbol{\mu}$ represents the trend, or expected value, at each location. In this study, $\boldsymbol{\mu}$ is assumed to be a constant vector, i.e., the random field is statistically stationary. $\boldsymbol{z}$ represents lognormal random fields of soil properties in this study, i.e. $z_i = \ln(z_i')$, where z_i' is the original value of the concerned property. The z_i components are correlated with each other, and the spatial correlation structure is described herein by a squared exponential function:

$$R_{ij} = \exp\left[-\left(\frac{x_i - x_j}{\theta_{\ln,x}}\right)^2 - \left(\frac{y_i - y_j}{\theta_{\ln,y}}\right)^2\right] \tag{2}$$

where R_{ij} is the spatial correlation (autocorrelation) between z_i and z_j, which also represents the ij-th element of the correlation matrix $\mathbf{R}$. The spatial coordinates of z_i and z_j are (x_i, y_i) and (x_j, y_j), respectively, while $\theta_{\ln,x}$ and $\theta_{\ln,y}$ are the autocorrelation distances in x and y directions. The system response of a geotechnical model can be represented as $g(\boldsymbol{z})$, which is also a random variable since the model input $\boldsymbol{z}$ is random. Alternatively, to facilitate the formulation of the proposed approach, g can be treated as a function of the standardized residual instead, i.e. $g(\boldsymbol{e})$, where $\boldsymbol{e} = \boldsymbol{\varepsilon}/\sigma_z$.

The Sobol' index can be adopted to assess the influence of property at each soil location, e_i, to the variance of g. By the law of total variance, the first order Sobol' index can be defined as:

$$S(e_i) = \frac{\mathrm{Var}_{e_i}\left[\mathrm{E}_{\boldsymbol{e}_{-i}}(g|e_i)\right]}{\mathrm{Var}(g)} = 1 - \frac{\mathrm{E}_{e_i}\left[\mathrm{Var}_{\boldsymbol{e}_{-i}}(g|e_i)\right]}{\mathrm{Var}(g)} \tag{3}$$

In the context of spatially variable subsurface domain, $\mathrm{E}_{e_i}[\mathrm{Var}_{\boldsymbol{e}_{-i}}(g|e_i)]$ is the expected value of system variance if sample value (e_i) is known for location i. Such variance arises from the fact that properties are unknown at other locations $(-i)$. $\mathrm{Var}_{e_i}[\mathrm{E}_{\boldsymbol{e}_{-i}}(g|e_i)]$ represents the variance in expected system response, due to all possible values of e_i. These two terms are denoted as $\mathrm{E}[\mathrm{Var}(g|e_i)]$ and $\mathrm{Var}[\mathrm{E}(g|e_i)]$ hereafter. In other words, the Sobol' index can be interpreted as the variance reduction in model response, when a soil sample is obtained at the i-th location. To facilitate the calculation of the Sobol' index, the response surface method may be used to reduce the number of simulations of the geotechnical model. In this study, the second order Polynomial Chaos Expansion (PCE) is adopted (Ghanem and Spanos 1991), which is given by:

$$g = a_0 + \sum_{j=1}^{M} a_j \xi_j + \sum_{j_1=1}^{M} \sum_{j_2=j_1}^{M} a_{j_1 j_2} (\xi_{j_1} \xi_{j_2} - \delta_{j_1 j_2}) \tag{4}$$

where $\xi_1, \ldots, \xi_M$ are independent standard normal variables, which may be grouped as a vector $\boldsymbol{\xi}$ with M principal components, representing a realization of soil profile. $\delta_{j_1 j_2}$ is the Kronecker delta. The unknown PCE coefficients $(a_0, a_j, a_{j_1 j_2})$ are determined by a non-intrusive approach, which involves generating and simulating N realizations of the geotechnical model for g, followed by a linear regression analysis. The prediction accuracy of the coefficients is measured by an indicator known as Q^2. Meanwhile, an adaptive algorithm proposed by Blatman and Sudret (2010) is adopted, whereby only the PCE coefficients that can increase the Q^2 are kept in the expansion (Eq. (4)). Once the PCE coefficients are determined, the expectation and variance of g can be evaluated as below:

$$\mathrm{E}(g) = a_0 \tag{5a}$$

$$\mathrm{Var}(g) = \sum_{j=1}^{M} a_j^2 + \sum_{j_1=1}^{M} \sum_{j_2=j_1}^{M} a_{j_1 j_2}^2 (1 + \delta_{j_1 j_2}) \tag{5b}$$

To utilize the PCE for evaluation of $S(e_i)$, it should be noted that although the vectors $\boldsymbol{\xi}$ and $\boldsymbol{e}$ are not equivalent, they are related by the following transformation:

$$\boldsymbol{e} = \mathbf{H}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\xi} = \mathbf{C}\boldsymbol{\xi} \tag{6}$$

where $\mathbf{H}$ and $\boldsymbol{\Lambda}$ are obtained from spectral decomposition of the correlation matrix ($\mathbf{R} = \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}^{\mathrm{T}}$), with $\mathbf{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_d]$ being a matrix containing d orthonormal eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix with d descending eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_d)$. The residual at the i-th location is therefore $e_i = \mathbf{C}_i\boldsymbol{\xi}$. $\mathbf{C}_i$ is the i-th row of $\mathbf{C}$, and is abbreviated as $\boldsymbol{c}$ in the subsequent formulation. Taking a conditional expectation on both sides of Eq. (4), the mean response of g conditioned on sample value e_i can be represented by:

$$\mathrm{E}(g|e_i) = a_0 + a_j \sum_{j=1}^{M} \mathrm{E}(\xi_j|e_i) + a_{j_1 j_2} \sum_{j_1=1}^{M} \sum_{j_2=j_1}^{M} [\mathrm{E}(\xi_{j_1}\xi_{j_2}|e_i) - \delta_{j_1 j_2}] \tag{7}$$

Since $\boldsymbol{\xi}$ is multivariate normal and e is standard normal, the conditional distribution of $\boldsymbol{\xi}$ given e_i is multivariate normal with mean of $\boldsymbol{c}^{\mathrm{T}}e_i$ and covariance of $\mathbf{I} - \boldsymbol{c}^{\mathrm{T}}\boldsymbol{c}$ ($\mathbf{I}$ is the identity matrix). Therefore,

$$\mathrm{E}(\xi_j|e_i) = c_j e_i \tag{8a}$$

$$\mathrm{E}(\xi_{j_1}\xi_{j_2}|e_i) = \delta_{j_1 j_2} - c_{j_1}c_{j_2} + c_{j_1}c_{j_2}e_i^2 \tag{8b}$$

Substituting (8) into (7), $\mathrm{E}(g|e_i)$ can be rewritten as:

$$
\begin{aligned}
& \mathrm{E}(g|e_i) = r_0 + r_1 e_i + r_2 e_i^2 \\
\text{where} \quad & r_1 = \sum_{j=1}^{M} a_j c_j \\
& r_2 = \sum_{j_1=1}^{M} \sum_{j_2=j_1}^{M} a_{j_1 j_2} c_{j_1} c_{j_2} \\
& r_0 = a_0 - r_2
\end{aligned}
\tag{9}
$$

Using Eq. (9), the mean response can be evaluated efficiently for any sample value, without resorting to additional conditional random field simulations. To evaluate $\mathrm{Var}[\mathrm{E}(g|e_i)]$, it should be noted that e_i is a standard normal variable, and $\mathrm{Var}(e_i) = 1$; $\mathrm{Var}(e_i^2) = 2$; $\mathrm{Cov}(e_i, e_i^2) = 0$. Therefore,

$$\mathrm{Var}\left[\mathrm{E}(g|e_i)\right] = r_1^2 + 2r_2^2 \tag{10}$$

With Eqs. (9) and (10), the Sobol' index $S(e_i)$ can be computed by Eq. (3). A Sobol' index map can be generated once $S(e_i)$ are calculated for all locations, and can be used to identify the most influential location, where the $S(e_i)$ value is maximum.

While Eq. (9) represents the conditional mean of g for normal random variable e, the soil property in this study is modeled as lognomal random variable. In other words, e in Eq. (9) represents the residual in log-transformed space: $e = (\ln z' - \mu_{\ln z'})/\sigma_{\ln z'}$, but the residual in original space is $e' = (z' - \mu_{z'})/\sigma_{z'}$. Conversion between the original residual e' and transformed residual e is given by:

$$e = \frac{\ln(e'\mathrm{CV}(z') + 1)}{\sqrt{\ln(1 + \mathrm{CV}(z')^2)}} + \frac{1}{2}\sqrt{\ln(1 + \mathrm{CV}(z')^2)} \tag{11}$$

where $\mathrm{CV}(z') = \sigma_{z'}/\mu_{z'}$. Substituting (11) into (9), the conditional mean in terms of e' is:

$$
\begin{aligned}
&\mathrm{E}(g|e_i') = s_0 + s_1 \ln\left[e_i'\mathrm{CV}(z') + 1\right] + s_2(\ln\left[e_i'\mathrm{CV}(z') + 1\right])^2 \\
\text{where} \quad &s_0 = r_0 + \frac{r_1}{2}\sqrt{\ln(1+\mathrm{CV}(z')^2)} + \frac{r_2}{4}\ln(1+\mathrm{CV}(z')^2) \\
&s_1 = \frac{r_1}{\sqrt{\ln(1+\mathrm{CV}(z')^2)}} + r_2 \\
&s_2 = \frac{r_2}{\ln(1+\mathrm{CV}(z')^2)}
\end{aligned}
\tag{12}
$$

From the risk assessment perspective, a key concern is the remaining uncertainty in system response after a sample value has been obtained. This is represented by the conditional variance, $\mathrm{E}[\mathrm{Var}(g|e_i)]$, or the conditional standard deviation (SD). There are two methods to estimate the conditional SD. The first method is to estimate directly from the Sobol' index:

$$
\mathrm{E}\left[\mathrm{SD}(g|e_i)\right] \approx \sqrt{\mathrm{E}\left[\mathrm{Var}(g|e_i)\right]} = \sqrt{\left[1 - S(e_i)\right]\mathrm{Var}(g)} = \sqrt{1 - S(e_i)}\mathrm{SD}(g) \tag{13}
$$

This method has been employed by Lo and Leung (2017b) to investigate the uncertainty reduction in footing displacement, and is accurate if the conditional SD is insensitive to the sample value. If the sample value heavily influences the conditional SD, an explicit function for conditional SD is required. Therefore, a conditional variance function $\mathrm{Var}(g|e_i) = V(\boldsymbol{\theta}, e_i)$ is proposed, with $\boldsymbol{\theta}$ being a vector of parameters describing the variance function. During the construction of PCE, N combinations of model response and sample values (g_i, e_i) are obtained. $\boldsymbol{\theta}$ can then be obtained by maximizing the following log-likelihood function (Davidian and Carroll 1987):

$$
L(\boldsymbol{\theta}) = -\sum_{i=1}^{N} \ln\left[V(\boldsymbol{\theta}, e_i)\right] - \sum_{i=1}^{N} \frac{\left[g_i - \mathrm{E}(g|e_i)\right]^2}{V(\boldsymbol{\theta}, e_i)} \tag{14}
$$

where $\mathrm{E}(g|e_i)$ is evaluated from Eq. (9). The maximization is performed in this study by an evolutionary searching algorithm known as Differential Evolution (Storn and Price 1997), which is not prone to converging at local maxima. Once $\mathrm{Var}(g|e_i)$ is obtained, it can be converted to $\mathrm{Var}(g|e_i')$ through Eq. (11).

Multiple sample points

Geotechnical investigation usually involves multiple samples being retrieved from borehole(s). It is therefore beneficial to extend the sensitivity analysis framework, for determination of multiple sample locations and the corresponding reduction in system uncertainty. This can be achieved by evaluating the n-th order Sobol' sensitivity index $S(\boldsymbol{e})$. The definition of the n-th order Sobol' index is similar to the 1st order index in Eq. (3), with the standardized residual e_i being replaced by a residual vector, $\boldsymbol{e} = \{e_1, e_2, \ldots, e_n\}^{\mathrm{T}}$, representing samples from n locations. $S(\boldsymbol{e})$ can be interpreted as the averaged variance reduction of the system response, when soil samples are available from n locations.

From a second order PCE, $S(\boldsymbol{e})$ can be formulated by first considering the joint distribution of vectors $\boldsymbol{\xi}$ and $\boldsymbol{e}$:

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_M \\ e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 & C_{11} & C_{21} & \cdots & C_{n1} \\ 0 & 1 & \cdots & 0 & C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & C_{1M} & C_{2M} & \cdots & C_{nM} \\ C_{11} & C_{12} & \cdots & C_{1M} & 1 & R_{12} & \cdots & R_{1n} \\ C_{21} & C_{22} & \cdots & C_{2M} & R_{21} & 1 & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nM} & R_{n1} & R_{n2} & \cdots & 1 \end{pmatrix} \right] = N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{I} & \mathbf{C}_s^{\mathrm{T}} \\ \mathbf{C}_s & \mathbf{R}_s \end{pmatrix} \right] \tag{15}$$

with covariance between $\boldsymbol{\xi}$ and $\boldsymbol{e}$ being $\mathrm{Cov}(e_i, \xi_j) = \mathrm{Cov}(\mathbf{C}_i \boldsymbol{\xi}, \xi_j) = C_{ij}$. $\mathbf{C}_s$ and $\mathbf{R}_s$ in Eq. (15) is a subset of the full $\mathbf{C}$ (Eq. (6)) and $\mathbf{R}$ matrices, and they are not identical. By multivariate normal theory, the conditional distribution of $\boldsymbol{\xi}$ given $\boldsymbol{e}$ is also multivariate normal with mean of $\mathbf{C}_s^{\mathrm{T}} \mathbf{R}_s^{-1} \boldsymbol{e}$ and covariance of $\mathbf{I} - \mathbf{C}_s^{\mathrm{T}} \mathbf{R}_s^{-1} \mathbf{C}_s$. Hereafter, they are expressed as $\mathbf{K}\boldsymbol{e}$ and $\mathbf{I} - \mathbf{Q}$, respectively. Based on this conditional distribution, the following conditional expectations can be obtained:

$$\mathrm{E}(\xi_j | \boldsymbol{e}) = \mathbf{K}_j \boldsymbol{e} \tag{16a}$$

$$\mathrm{E}(\xi_{j_1}\xi_{j_2}|\boldsymbol{e}) = \delta_{j_1 j_2} - \mathbf{Q}_{j_1 j_2} + (\mathbf{K}_{j_1}\boldsymbol{e})(\mathbf{K}_{j_2}\boldsymbol{e}) \tag{16b}$$

Substituting Eq. (16) into the conditional mean of a 2nd order PCE, the conditional mean of g can be written as:

$$\begin{aligned}
&\mathrm{E}(g|\boldsymbol{e}) = r_0 + \sum_{i=1}^{n} r_i e_i + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n} r_{i_1 i_2} e_{i_1} e_{i_2} \\
\text{where} \quad & r_0 = a_0 - \sum_{j_1=1}^{M}\sum_{j_2=j_1}^{M} a_{j_1 j_2}\mathbf{Q}_{j_1 j_2} \\
& r_i = \left[\sum_{j=1}^{M} a_j \mathbf{K}_j\right]_i \\
& r_{i_1 i_2} = \begin{cases} \mathbf{P}_{ii} & \text{if } i_1 = i_2 = i \\ \mathbf{P}_{i_1 i_2} + \mathbf{P}_{i_2 i_1} & \text{if } i_1 \neq i_2 \end{cases} \\
& \mathbf{P} = \left[\sum_{j_1=1}^{M}\sum_{j_2=j_1}^{M} a_{j_1 j_2}\mathbf{K}_{j_1}^{\mathrm{T}}\mathbf{K}_{j_2}\right]
\end{aligned} \tag{17}$$

It is also possible to represent the conditional mean in terms of residuals in original space ($\boldsymbol{e}'$), and derive the s coefficients similar to Eq. (12). However, the corresponding formulation is complex, and it may be easier to directly apply Eq. (17) with log-transformed $\boldsymbol{e}$. Meanwhile, $\mathrm{Var}[\mathrm{E}(g|\boldsymbol{e})]$ is given by:

$$\mathrm{Var}\left[\mathrm{E}(g|\boldsymbol{e})\right] = \sum_{i=1}^{n}\sum_{k=1}^{n} r_i r_k \mathrm{Cov}(e_i, e_k) + \sum_{i_1=1}^{n}\sum_{i_2=i_1}^{n}\sum_{k_1=1}^{n}\sum_{k_2=k_1}^{n} r_{i_1 i_2} r_{k_1 k_2}\mathrm{Cov}(e_{i_1}e_{i_2}, e_{k_1}e_{k_2}) \tag{18}$$

To evaluate the components in Eq. (18), it should be noted that the components in $\boldsymbol{e}$ are spatially correlated. The following covariance can be obtained by the theory of characteristic functions (with details in the Appendix):

$$\mathrm{Cov}(e_i, e_k) = R_{ik} \tag{19a}$$

$$\mathrm{Cov}(e_i, e_{k_1} e_{k_2}) = 0 \tag{19b}$$

$$\mathrm{Cov}(e_{i_1} e_{i_2}, e_{k_1} e_{k_2}) = R_{i_1 k_1} R_{i_2 k_2} + R_{i_1 k_2} R_{i_2 k_1} \tag{19c}$$

and $\mathrm{Var}[\mathrm{E}(g|\boldsymbol{e})]$ can then be used to calculate the n-order Sobol' index.

To obtain the maximum index $S(\boldsymbol{e})$, the spatial configuration of the sample vector $\boldsymbol{e}$ has to be optimized. The n locations for vector $\boldsymbol{e}$ may be searched simultaneously using various optimization algorithms. Alternatively, the multiple sample locations can be determined sequentially. For example, starting with $i = 1$, the Sobol' index is evaluated across the domain to identify the location that corresponds to the maximum value of $S(e_1)$. Once this location is decided for the first sample, the Sobol' index is evaluated again with $i = 2$, based on the selected location of sample 1, to identify the second sample location leading to the maximum value of $S(e_1, e_2)$. This stepwise procedure is then repeated until the target number of samples is reached. It is also possible to impose various constraints when searching for the optimal sample locations. For example, when multiple samples are retrieved from a single borehole, they need to share the same horizontal coordinates.

The proposed Sobol' index approach is essentially a post-processing technique for random field analyses, performed in this study through the PCE. The benefits of the proposed approach are multifold. Contrary to previous conditional random field modeling techniques, the Sobol' index is a global sensitivity index, which means it encompasses all possible values of an uncertain input parameter. In addition, by constructing a Sobol' index map, all potential sample locations are assessed simultaneously for the formulation of optimal sampling strategy. There are no needs for pre-determined sample values or sampling patterns, and it is not necessary to perform separate conditional random field simulations to investigate multiple scenarios, thereby reducing the computational demands substantially. In the following sections, the Sobol' index approach will be validated through comparisons with conditional random field simulations, and then applied to reliability assessments involving slope stability.

APPLICATIONS IN RELIABILITY ASSESSMENTS OF SLOPES

Comparisons with conditional random field simulations

The proposed Sobol' index approach can be validated by comparisons with conditional random field simulations, through which the advantages of the approach are also illustrated. Essentially, it allows efficient evaluation of the mean and standard deviation of factor of safety (FS) from any sample value e', which can then be used to obtain the reliability index or failure probability of the slope. This section considers a 35° slope with a height of 5 m, simulated in *FLAC* with a model boundary 15 m below the top of the slope. The soil is idealized as Tresca material, with mean undrained shear strength (c_u) of 36.3 kPa, which would correspond to a 'deterministic' FS of 2.0, if the soil profile was assumed to be uniform. Instead, the c_u profile is simulated as lognormally distributed in the model, with coefficient of variation of 0.4, horizontal autocorrelation distances of $\theta_{\ln,x} = 14.3$ m and $\theta_{\ln,y} = 2.5$ m. Throughout this study, the element size in the *FLAC* model is about 0.5 m (vertical) $\times$ 1.0 m (horizontal), which are always smaller than half of the autocorrelation distances in the corresponding directions. The influence of element size on FS values of slopes was studied by Dawson et al. (1999), who showed that for a 10-m slope with slope angles between 15°-45°, the difference in FS values obtained by a fine mesh (60×60) and a coarse mesh (20×20) is smaller than 4%. As part of the verification process in this study, a separate analysis was conducted with 4 times the number of elements (each 0.25 m $\times$ 0.5 m in size), and the subsequent changes to the Sobol' indices, conditional mean and SD curves are mostly under 5%.

Based on these spatial variability features, unconditional random field simulations are first performed without specifying any sample locations in the domain. 1,000 realizations of c_u profiles are generated by Latin Hypercube Sampling with Dependence (LHSD), which is an extension of LHS and aims to introduce stratification while maintaining the spatial correlation of random variables. Details of LHSD and its incorporation into Cholesky decomposition are described in Packham and Schmidt (2010) and Lo and Leung (2017a), who also showed that when LHSD is coupled with PCE, 1,000 realizations are sufficient to obtain robust estimates of the PCE coefficients and hence probability density of the FS for slopes with similar features of spatially variable soils.

Using this approach, the unconditional mean and standard deviation of FS are found to be E(FS)= 1.697 and SD(FS)= 0.308 in this case. Through decomposition of the $\mathbf{R}$ matrix and Eqs. (9) and (10), the Sobol' index map is then derived, and is shown in Fig. 1(a). The optimal sample location is found at the depth of 6.25 m from the slope top, with horizontal separation of 5.8 m from the slope toe, where the Sobol' index is at the maximum of 0.397. This means that on average (considering all possible sample values), an approximate SD reduction of 22.3% can be achieved if a sample is obtained at this location. The concentric shape of the Sobol' index variation is related to the ratio between $\theta_{\ln,x}$ and $\theta_{\ln,y}$. Since $\theta_{\ln,x}$ is more than 5.5 times larger than $\theta_{\ln,y}$ in this example, the corresponding contour is elongated in the x-direction. In case of a layered soil profile with $\theta_{\ln,x} = \infty$, the Sobol' index contour will also display a 'layered' feature as all locations at the same depth are equally important because of the highly correlated properties.

The conditional mean equation, considering standardized residuals in original space (E(FS$|e'$)), is then calculated from the PCE through Eq. (12) and plotted in Fig. 1(b), where the coefficients are found to be $(s_0, s_1, s_2) = (1.76, 0.47, -0.175)$ for this case. As mentioned earlier, the conditional variance equation can be obtained by maximizing the log-likelihood function (Eq. (3)) involving $V(\boldsymbol{\theta}, e)$. In this study, the generalized logistic function (Richards 1959) is adopted for $V(\boldsymbol{\theta}, e)$ since it has a flexible functional form with an asymmetric S-shape:

$$\mathrm{Var}(\mathrm{FS}|e) = V(\boldsymbol{\theta}, e) = l + \frac{u - l}{\{1 + t\exp[-b(e - m)]\}^{1/t}} \tag{20}$$

The shape of the function is controlled by the parameters $\boldsymbol{\theta} = \{l, u, t, b, m\}$, where l is the lower bound, u is the upper bound, t is the asymmetry parameter, b is the growth rate, and m is the point of inflection. The maximization is performed by Differential Evolution, with the following optimized parameters: $\{l, u, t, b, m\} = \{0.013, 0.12, 1.6, 1, 0.56\}$ in this case. Through this 'sensitivity function', the relationship between uncertainty in FS and the potential sample values can be developed. For example, the conditional variance $\mathrm{Var}(\mathrm{FS}|e)$ is converted to the conditional SD curve in terms of original residual, i.e. $\mathrm{SD}(\mathrm{FS}|e')$, and plotted in Fig. 1(c). The function allows rapid determination of the system reliability once the sample value becomes available.

These results by the Sobol' index approach can be validated through conducting separate conditional random field simulations, where sample values are assigned at the optimal sampling point mentioned earlier. Eight individual sets of analyses are performed, with different sample values assigned at that location: $e' = (-1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2)$, which corresponds to about 95% of the possible cases considering the lognormal distribution of c_u. For each set of analysis, 1,000 realizations of the conditional random field are simulated and assessed using the LHSD-PCE approach. The corresponding conditional mean and SD results are also plotted in Figs. 1(b) and (c), showing close agreements with the relationships developed by the proposed Sobol' index approach, and therefore validating its accuracy. It is important to note that an additional 8,000 *FLAC* analyses are required by conditional random field modeling, whereas the proposed Sobol' index approach only requires 1,000 unconditional random field simulations. The associated reductions in computational demands are substantial.

In fact, the influence of sampling at any (non-optimal) location can be evaluated by the proposed approach, by adopting another $\boldsymbol{c}$ vector corresponding to the intended sample location i. To illustrate such effects, a sensitivity analysis is performed with another sample location indicated in Fig. 1(a). The prediction intervals, taken as $\mathrm{E}(\mathrm{FS}|e') \pm \mathrm{SD}(\mathrm{FS}|e')$, are shown in Fig. 1(d) for both sample locations. The prediction interval arising from the optimal location is steeper and narrower than that of the non-optimal location, which again demonstrates that FS is more sensitive to the c_u value at the optimal location.

Design charts for slopes in c_u soils

Based on the proposed Sobol' index approach, a series of design charts can be developed for efficient assessments of slope reliability according to the available sample value of shear strength parameters. For example, a long embankment may be designed based on mean shear strength parameters across the site. Making use of the charts in this study, the failure probabilities of individual sections can be easily evaluated when a soil sample is retrieved at any of those locations.

Table 1 presents the various slope geometries and subsurface conditions considered in this study. For slopes in c_u soils, the $FLAC$ model geometries are similar to those described earlier, with slope height $H = 5$ m, and model boundary 15 m below the top of the slope. Four different slope angles are considered: $\beta = 20°$, $30°$, $35°$ and $40°$, and the width of the slope is therefore $W = H/\tan\beta$. c_u is modeled as a lognormal random field, with the mean values (μ_{cu}) chosen such that deterministic analyses (assuming uniform soil profiles) would lead to FS of 1.0, 1.5 or 2.0. Such 'deterministic FS' provides a useful indicator of the degree of strength mobilization, which can be easily applied by most practitioners in a typical design process. Various patterns of spatial variability for c_u are considered in developing the design charts. The coefficient of variation (i.e. CV_{cu}) is assigned to be 0.15 or 0.4, while the autocorrelation distance $\theta_{\ln,x} = 0.5W$, $2W$ or ∞, and $\theta_{\ln,y} = 0.25H$ or $0.5H$. Considering these variations, 36 series of probabilistic analyses are performed for each slope angle.

For each series of analyses, a PCE is constructed using 1,000 c_u profile realizations. It should be noted that for all analyses in this study, the value of Q^2, which measures accuracy of PCE coefficients, exceeds 0.92. With the coefficients determined, the optimal sampling locations are then obtained using the Sobol' index approach. It was found that the sampling locations are insensitive to the mean values of the property, and therefore cases with different μ_{cu} values are averaged. According to Fig. 2(a), the optimal sampling depths D are all below the slope, i.e., $D/H > 1$, which is expected since deep-seated failures are more common for slopes in c_u soils. In general, D becomes shallower as β increases or with higher variability in c_u, since the possibilities of shallow slope failures increase with potential weak spots at shallow depths. Fig. 2(a) also shows the horizontal distance between the optimal sampling location and the slope toe (L), normalized by the width of slope (W). In most cases, it would be reasonable to adopt $L = 0.5W$.

Based on these optimal sample locations, the sensitivity functions (conditional mean $\mathrm{E}(\mathrm{FS}|e')$ and conditional standard deviation $\mathrm{SD}(\mathrm{FS}|e')$ relationships) can be evaluated by the procedures described earlier. Fig. 3 shows that under the same deterministic FS (of 1.5) and spatial variability conditions (i.e. CV_{cu}, $\theta_{\ln,x}$, $\theta_{\ln,y}$), the sensitivity functions are relatively insensitive to the slope angle (from 20° to 40°), with a narrow range around the average value. Therefore, the obtained functions are averaged across various slope angles in the subsequent figures. Also, various sets of analyses are performed with the model scale doubled, as illustrated in Table 1. The results are also shown in Fig. 3 and indicated that the sensitivity functions are scale-invariant. In other words, regardless of the size of the slope, the functions will be identical for the same $\theta_{\ln,x}/W$ and $\theta_{\ln,y}/H$ ratios. Meanwhile, the intention of Fig. 3 is not to imply that autocorrelation distances will be scaled up or down with the slope geometries at a particular site. In fact, for site-specific applications, it is necessary to evaluate the $\theta_{\ln,x}/W$ and $\theta_{\ln,y}/H$ ratios based on the site conditions, and use the corresponding sensitivity functions pertaining to those conditions.

For example, Fig. 4 represents the design charts with sensitivity functions for different values of deterministic FS, and demonstrates the influence of spatial correlation features on slope reliability. In particular, the $\mathrm{E}(\mathrm{FS}|e')$ function becomes steeper with larger values of $\theta_{\ln,x}$, $\theta_{\ln,y}$ and/or CV_{cu}, which means the sample has a larger conditioning effect to FS in these cases. In Figs. 4(a) and (b), the $\mathrm{E}(\mathrm{FS}|e')$ functions for $\theta_{\ln,x} = 2W$ are omitted for brevity, but the curves lie approximately midway between those of $\theta_{\ln,x} = 0.5W$ and $\theta_{\ln,x} = \infty$ for positive e' values, and are close to the latter for negative e' values. The corresponding $\mathrm{SD}(\mathrm{FS}|e')$ curves are also shown in Figs. 4(c) to (h), and they are found to increase with CV_{cu}, $\theta_{\ln,x}$ and $\theta_{\ln,y}$. Using these relationships, the reliability of a slope can be quickly assessed for any sampled value of c_u represented as standardized residual e'. For other values of deterministic FS, it is reasonable to interpolate the results as a monotonic trend is observed among the various curves. As discussed earlier, these relationships are independent of the problem scale and may also be applied to other slope dimensions.

Design charts for slopes in $c - \phi$ soils

Similar procedures are applied to slopes in $c-\phi$ soils to establish the corresponding design charts. In this section, a slope height of $H = 10$ m is adopted, and the model boundary is 15 m below the top of the slope. The H to model boundary ratio is smaller than that for c_u soils, since slopes in 'sandy' soils often fail with relatively shallow slip surfaces, and this phenomenon will be discussed here in terms of sensitivity analyses. This section will focus on the variability of friction angle ϕ, while c is assumed to be a constant of 5 kPa. As shown in Table 1, ϕ is modeled as a lognormal random field, with mean values (μ_ϕ) designated to achieve deterministic FS of 1.0, 1.5 or 2.0. The coefficient of variation of ϕ (i.e. CV_ϕ) is assigned to be 0.05 or 0.1, while other variations are similar to those adopted for c_u slopes. 'Dry' slopes are first investigated and the influence of water table will be discussed later.

Again, the optimal sampling locations are found to be insensitive to the μ_ϕ values, and therefore cases with different μ_ϕ are averaged. The normalized sampling positions are shown in Figs. 2(c) and (d), where the optimal sampling depths are all within the slope, with $D/H < 1$. In other words, stability of $c - \phi$ slopes are mostly influenced by shear strength close to the slope face, which is another manifestation of the shallow slip surfaces usually observed in these slopes. Similar to the c_u slopes, a larger slope angle or CV_ϕ value would lead to a shallower optimal sampling depth, and D/H ranges from around 0.75 to 0.95 for all the studied cases. Meanwhile, L is approximately $0.4W$ for most cases.

Fig. 5 shows the conditional mean and SD curves for $c - \phi$ slopes with sample obtained from the optimal location. Similar to the case of c_u slopes, additional analyses are performed with a double model size, and the results are found to be invariant to the model scale (Fig. 3). In addition, the curves are found to be insensitive to the horizontal autocorrelation distance of ϕ once $\theta_{\ln,x}$ exceeds $2W$. This may be attributed to the relatively shallow slip surfaces, which tend to intercept several horizontal soil layers, making $\theta_{\ln,y}$ generally more influential when $\theta_{\ln,x}$ extends beyond the width of the slope. On the other hand, Figs. 5(a) and (b) shows that the E(FS|e') curves becomes steeper with increasing values of $\theta_{\ln,y}$, which means the sample significance increases as the soil becomes more uniform.

Interestingly, the SD(FS|e') curves in Figs. 5(c) to (h) are shifted downwards with increasing $\theta_{\ln,y}$, which is an opposite trend compared to c_u slopes. In fact, the conditional SD curves depend both on the value of $\theta_{\ln,y}$, and the conditioning effects of the sample. For a probabilistic assessment without any sample value, the 'unconditional' SD of slope performance will always increase with $\theta_{\ln,y}$, since there are less random effects that 'average out' the soil variability, leading to more uncertain performance. This trend is, however, counteracted by the knowledge of sample value at a designated location, in which case a larger $\theta_{\ln,y}$ means a greater conditioning effect of the sample, as a thicker layer becomes associated with that value. The conditioning effects appear to be less dominant for c_u slopes that involve deep-seated failure mass, but are more influential for $c - \phi$ slopes with shallow slip surfaces. Similar influence of conditioning to the trend reversal of SD is also observed in probabilistic assessments of footing performance presented by Lo and Leung (2017a).

The influence of water table to the reliability of $c - \phi$ slopes is also investigated in this study, where the water level is modeled such that 3/4 of the slope is submerged, as shown in the inset of Fig. 6(a). The scenario with deterministic FS=1.5 and slope angle of 30° is selected for comparisons of E(FS|e') and SD(FS|e'), as shown in Fig. 6. It should be noted that the mean friction angles are different for dry and submerged slopes in order to achieve the same deterministic FS, with $\mu_\phi = 32.5°$ for the former case and $\mu_\phi = 42.5°$ for the latter. Fig. 6 shows that the E(FS|e') relationships are steeper for submerged slope, resulting in higher E(FS|e') for positive e' and lower E(FS|e') for negative e'. Meanwhile, the SD(FS|e') relationships are shifted upwards when water table is present. However, the differences in both mean and standard deviation of FS are not substantial comparing the dry and submerged cases. In general, the optimal sampling depths for the submerged slope is slightly deeper (by about $0.05H$) compared to the dry slope case.

Example applications of the design charts

To illustrate the application of the presented charts, this section considers a scenario where a 5 m high cut slope of 30° is to be constructed. Based on previous geotechnical investigation and/or knowledge of the site, suppose the mean c_u of the clayey deposits is estimated to be 27 kPa, with $\mathrm{CV}_{cu} = 0.4$. The standard deviation of c_u (i.e., σ_z) is therefore 10.8 kPa, while $\theta_{\ln,x} = 17.3$ m $(= 2W)$ and $\theta_{\ln,y} = 2.5$ m. With this mean c_u value, the deterministic FS is 1.5. To refine the estimates of failure probability for a particular section of the slope, a soil sample is to be obtained. Making use of Fig. 2, the optimal location of the sample would be about 7.5 m deep from the top of slope, and 4.3 m away from the toe of the future slope profile. Suppose a clay sample is then retrieved from this location with c_u determined to be 22.7 kPa. This would correspond to $e' = -0.4$ (the sampled value is $-0.4\sigma_z$ away from the mean). From the design charts in Fig. 4, taking deterministic FS= 1.5, $\theta_{\ln,y} = 0.5H$ and with $e' = -0.4$, the conditional mean and SD of FS become 1.27 and 0.18, respectively. Assuming the FS has a normal distribution, the slope has a failure probability of 0.067.

Consider another possible scenario at the same site, where prior samples or knowledge of the soil properties are limited, and the mean c_u at the site cannot be determined with confidence. Nonetheless, the optimal sample location is still the same since it does not depend on the mean property value. Suppose a sample is retrieved and the c_u of that sample is determined to be 22.7 kPa. Without other information, this is taken as the mean value and the deterministic FS is found to be 1.26. Making use of the design charts in Fig. 4, the conditional mean and SD of FS may be further approximated, assuming $e' = 0$ and with estimates of CV_{cu} and $\theta_{\ln,y}$ according to published literature or the engineers' judgement. It is also possible to test how the failure probability changes with different CV or θ assumptions, and these estimates may provide useful information to support the engineering decisions.

Applications for multiple sample points

This section illustrates the application of Sobol' index approach considering multiple sample points, making use of the sequential algorithm described earlier. The slope is in c_u soils with a slope height of 5 m and slope angle of 35°. The *FLAC* model boundary is 15 m below the top of the slope. c_u is simulated as a lognormal random field with a mean value 27.3 kPa, corresponding to a deterministic FS of 1.5. $\text{CV}_{cu} = 0.4$, while $\theta_{\ln,x} = 2W = 14.3$ m and $\theta_{\ln,y} = 0.25\ H = 1.25$ m. Based on these parameters, the unconditional mean and standard deviation of FS are E(FS)=1.28 and SD(FS)=0.18.

Considering the scenario where six soil samples are to be obtained from two boreholes in the slope, the optimal sample locations and resulting conditional FS distribution can be evaluated by the proposed approach. The first sample is determined by searching throughout the entire slope profile, and the corresponding lateral coordinates become those of the first borehole. If the optimal location of subsequent samples falls within a close horizontal distance (taken as 3 m in this case) from the first borehole, then a constrained search is performed such that the sample would share the same lateral coordinates as this borehole. Otherwise this sample indicates the location of the second borehole.

Fig. 7(a) shows the location of the six samples thus determined, while Table 2 summarizes their coordinates and the corresponding Sobol' index values. The cumulative reductions in SD(FS) are also shown in percentages as the number of samples increases, approximated by $100[1 - \sqrt{1 - S(\boldsymbol{e})}](\%)$ as indicated in Eq. (13). The vertical spacing between the samples is 2 to 2.5 m, which is about 2 times $\theta_{\ln,y}$. The horizontal spacing between the two boreholes is 7.3 m, which is about 0.5 times $\theta_{\ln,x}$, which is consistent with the recommendation by Li et al. (2016b). Once the sample locations are determined, sensitivity analysis is conducted to obtain the conditional mean equation $\text{E}(\text{FS}|\boldsymbol{e})$, according to Eq. (17). The coefficients of the conditional mean equation are given in Table 3.

The improvements in response prediction through the proposed sampling strategy are further illustrated through 1,000 realizations of c_u profiles, simulated by the LHSD technique described by Lo and Leung (2017a). The profiles are then analyzed by *FLAC* to obtain the FS values, denoted by $\mathrm{FS}_1, \mathrm{FS}_2, \ldots, \mathrm{FS}_{1000}$. If an unconditional random field assessment was performed, the prediction errors can be represented by $\mathrm{FS}_i - \mathrm{E(FS)}$ (i from 1 to 1,000). On the other hand, the conditional predictions can be performed using the sampling strategy presented in Table 2, and 1,000 sample combinations can be realized as $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_{1000}$. The associated prediction errors are therefore $\mathrm{FS}_i - \mathrm{E}(\mathrm{FS}|\boldsymbol{e}_i)$, where the coefficients of $\mathrm{E}(\mathrm{FS}|\boldsymbol{e}_i)$ are given in Table 3 and no separate conditional random field simulations are required. Fig. 7(b) shows the density plots of the unconditional errors and the conditional errors, with their standard deviations being 0.184 and 0.102, respectively, and hence the percentage reduction in SD is about 44.3%. This SD reduction is identical to the value shown in Table 2, which is calculated using a different approach, i.e., directly through the Sobol' index. In other words, this numerical testing also verifies the accuracy of the proposed approach for multiple sample points.

EXAMPLE APPLICATION OF DESIGN CHARTS TO CASE STUDY

The case study of the James Bay hydroelectric project in Quebec was described in detail by Christian et al. (1994), and is revisited in this study to illustrate the application of the proposed design charts and their differences with the FOSM method. The project involved construction of embankments on soft sensitive clays, with the first stage consisting of two berms with a total height (H) of 12 m. The two berms are separated from each other in the horizontal direction and the total width (W) is approximately 90 m (Fig. 8a). Field vane shear tests had been conducted in 35 boreholes across the site to characterize c_u of the marine clay and lacustrine clay below the embankment, and the details are also shown in Fig. 8b. The spatial variability features were investigated by DeGroot and Baecher (1993) and Christian et al. (1994) (Fig. 8c), with the autocorrelation distances in horizontal and vertical directions estimated to be 37.3 m and 1.1 m from their results, assuming similar spatial correlation features for marine and lacustrine clays.

Some approximations are inevitable when applying the design charts for the case study.

Based on the slope geometries and spatial correlation features, the charts for $\theta_{\ln,y} = 0.25H$ and $\theta_{\ln,x} = 0.5W$ are adopted. The deterministic FS was estimated to be 1.453 ($\approx$ 1.5) by Christian et al. (1994), with the critical failure plane intersecting the lower lacustrine clay that has $\mathrm{CV}_{cu} = 0.272$. Therefore, when applying the design charts, the results from $\mathrm{CV}_{cu} = 0.15$ and $\mathrm{CV}_{cu} = 0.4$ are averaged. Considering the total H and W of the two berms, the average slope angle would be 8°, and the corresponding optimal (single) sample location is at $L = 0.5W$ and $D = 2.4H$ by extrapolating Fig. 2. The sampled value at that depth is very close to the mean c_u, and therefore $e' = 0$.

Based on these approximations and by averaging the results from Figs. 4(a) and 4(b), the conditional mean FS is estimated to be 1.382, and the conditional SD of FS is 0.083 from Figs. 4(g) and 4(h). This SD value is lower than the value estimated with FOSM by Christian et al. (1994), which was 0.205 considering only the contributions from spatial variations of c_u (total SD was 0.257, which included factors such as variations in the fill and crust layers). The discrepancy can be attributed to two main reasons. Firstly, the FOSM method assumes the soil to be homogeneous ($\theta_{\ln,y} = \theta_{\ln,x} = \infty$) when the partial derivative $\partial\mathrm{FS}/\partial c_u$ was evaluated, which leads to an increase in SD estimates. For example, if the design charts are applied to the scenario of $\theta_{\ln,x} = \infty$ and $\theta_{\ln,y} = 0.5H$, with other conditions being the same (i.e., Figs. 4(c) and 4(d)), the estimated SD will become 0.154, and will increase further as $\theta_{\ln,y}$ approaches the assumptions of FOSM. In addition, the proposed approach and design charts explicitly considered the influence of the soil sample, which also reduces the SD in FS estimates. It should be noted, however, that the James Bay project involves multiple samples which cannot be accounted for only by using the design charts. Yet, the estimates through the charts can be useful indicators for practitioners. If all samples need to be considered in a more rigorous manner, the approach presented in Eqs. (15) to (19) can be adopted.

CONCLUSION

This paper extends the Sobol' index approach for the simulation of spatially variable soil properties, and explores its applications to the reliability assessment of slopes. The approach can be used to identify the optimal locations of samples which will bring maximum reduction to the uncertainty in system performance. It also allows the derivation of sensitivity functions for site-specific risk assessment of system performance, which is computationally efficient since the influence of different sample values can be evaluated without performing additional conditional random field simulations.

Design charts are presented for the case of a single sample in c_u slopes and $c - \phi$ slopes. These charts enable efficient assessment of the failure probability of slopes, represented as the conditional mean and standard deviations under various conditions of slope geometries and spatial variability. It should be noted, however, that probabilistic methods should not be taken as replacement of the understanding of local geology, which can be helpful in identifying potential anomalously weak layers that could be one of the main causes of slope failures. Also, the charts are developed based on the assumption of statistical stationarity in the shear strength parameters of the soil, with constant mean values of c_u or ϕ. They should be applied with proper engineering judgement, especially when the geotechnical conditions deviate significantly from these assumptions. In those cases, it is also possible to conduct detailed risk assessments through the proposed Sobol' index approach.

The proposed approach is extended to consider multiple sampling points through evaluation of multi-ordered Sobol' index. This is illustrated through an example where the optimal locations of six samples along two boreholes are determined. A significant reduction in the performance uncertainty can be achieved, and is verified by the much smaller prediction errors in the FS of slope, compared to the case when no samples are available. The proposed Sobol' index approach is shown to be a useful post-processing tool on the existing random field simulation results, and is capable of revealing high risk zones related to the specific geotechnical applications. This can be integrated into a risk assessment framework to assist the decision-making process associated with the uncertainty of system performance arising from spatial variability of geotechnical properties.

SUPPLEMENTAL DATA

Descriptions on the details of risk-based design process utilizing the proposed Sobol' index approach, together with Fig. S1, are available online in the ASCE Library (www.ascelibrary.org).

ACKNOWLEDGEMENT

The work presented in this paper is financially supported by the Research Grants Council of the Hong Kong Special Administrative Region (Project No. 25201214). Also, the authors would like to acknowledge the valuable advice by Dr Zhen Pang of the Department of Applied Mathematics, The Hong Kong Polytechnic University.

APPENDIX

According to Anderson (1984), if $\boldsymbol{X} = \{X_1, \ldots, X_p\}^T$ is a multivariate normal distribution, i.e. $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the characteristic function ϕ is given by:

$$\phi(\boldsymbol{t}) = \mathrm{E}\left[\exp(i\boldsymbol{t}^T\boldsymbol{X})\right] = \exp\left(i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right) \tag{21}$$

The n-th non-central moment of $\boldsymbol{X}$ can be obtained by differentiating ϕ:

$$\mathrm{E}\left[X_1 \ldots X_n\right] = \frac{1}{i^n}\frac{\partial \phi}{\partial t_1 \ldots \partial t_n}\bigg|_{\boldsymbol{t}=\boldsymbol{0}} \tag{22}$$

The first four non-central moments are hence derived:

$$\begin{aligned}
\mathrm{E}[X_1] &= \mu_1 \\
\mathrm{E}[X_1X_2] &= \frac{1}{i^2}\frac{\partial \phi}{\partial t_1 \partial t_2}\bigg|_{\boldsymbol{t}=\boldsymbol{0}} = \sigma_{12} + \mu_1\mu_2 \\
\mathrm{E}[X_1X_2X_3] &= \frac{1}{i^3}\frac{\partial \phi}{\partial t_1 \partial t_2 \partial t_3}\bigg|_{\boldsymbol{t}=\boldsymbol{0}} = \sigma_{12}\mu_3 + \sigma_{13}\mu_2 + \sigma_{23}\mu_1 + \mu_1\mu_2\mu_3 \\
\mathrm{E}[X_1X_2X_3X_4] &= \frac{1}{i^4}\frac{\partial \phi}{\partial t_1 \partial t_2 \partial t_3 \partial t_4}\bigg|_{\boldsymbol{t}=\boldsymbol{0}} = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} \\
&\quad + \sigma_{12}\mu_3\mu_4 + \sigma_{13}\mu_2\mu_4 + \sigma_{14}\mu_2\mu_3 + \sigma_{23}\mu_1\mu_4 + \sigma_{24}\mu_1\mu_3 + \sigma_{34}\mu_1\mu_2
\end{aligned} \tag{23}$$

which lead to these covariances:

$$\mathrm{Cov}[X_1, X_2X_3] = \mathrm{E}[X_1X_2X_3] - \mathrm{E}[X_1]\mathrm{E}[X_2X_3] = \sigma_{12}\mu_3 + \sigma_{13}\mu_2$$

$$\begin{aligned}\mathrm{Cov}[X_1X_2, X_3X_4] &= \mathrm{E}[X_1X_2X_3X_4] - \mathrm{E}[X_1X_2]\mathrm{E}[X_3X_4] \\ &= \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23} + \sigma_{13}\mu_2\mu_4 + \sigma_{14}\mu_2\mu_3 + \sigma_{23}\mu_1\mu_4 + \sigma_{24}\mu_1\mu_3\end{aligned} \tag{24}$$

Since $\boldsymbol{e} \sim N(\mathbf{0}, \mathbf{R})$,

$$\mathrm{Cov}(e_i, e_{k_1}e_{k_2}) = 0$$

$$\mathrm{Cov}(e_{i_1}e_{i_2}, e_{k_1}e_{k_2}) = R_{i_1k_1}R_{i_2k_2} + R_{i_1k_2}R_{i_2k_1} \tag{25}$$

REFERENCES

Al-Bittar, T. and Soubra, A.-H. (2014). "Probabilistic analysis of strip footings resting on spatially varying soils and subjected to vertical or inclined loads." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0001046, 04013043.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. John Wiley & Sons, Inc.

Baecher, G. B. and Christian, J. T. (2003). *Reliability and Statistics in Geotechnical Engineering*. Wiley.

Blatman, G. and Sudret, B. (2010). "An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis." *Probabilistic Engineering Mechanics*, 25(2), 183 – 197.

Cho, S. E. (2010). "Probabilistic assessment of slope stability that considers the spatial variability of soil properties." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0000309, 975–984.

Christian, J. T. and Baecher, G. B. (2011). "Unresolved problems in geotechnical risk and reliability." *Georisk 2011: Geotech. Risk Assess. Manage.*, 224, 50–63.

Christian, J. T., Ladd, C. C., and Baecher, G. B. (1994). "Reliability applied to slope stability analysis." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)0733-9410(1994)120:12(2180), 2180–2207.

Davidian, M. and Carroll, R. J. (1987). "Variance function estimation." *Journal of the American Statistical Association*, 82(400), 1079–1091.

Dawson, E. M., Roth, W. H., and Drescher, A. (1999). "Slope stability analysis by strength reduction." *Géotechnique*, 49(6), 835–840.

DeGroot, D. J. and Baecher, G. B. (1993). "Estimating autocovariance of in-situ soil properties." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1993)119:1(147), 147–166.

Duncan, J. M. (2000). "Factors of safety and reliability in geotechnical engineering." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2000)126:4(307), 307–316.

Fenton, G. A. and Griffiths, D. V. (2003). "Bearing-capacity prediction of spatially random c - ϕ soils." *Can. Geotech. J.*, 40(1), 54–65.

Ghanem, R. G. and Spanos, P. D. (1991). *Stochastic Finite Element: A Spectral Approach.* Springer, New York.

Hasofer, A. M. and Lind, N. C. (1974). "An exact and invariant first order reliability format." *J. Eng. Mech. Div.*, 100(EM1), 111121.

Houmadi, Y., Ahmed, A., and Soubra, A. H. (2012). "Probabilistic analysis of a one-dimensional soil consolidation problem." *Georisk*, 6(1), 36–49.

Jiang, S.-H., Li, D.-Q., Cao, Z.-J., Zhou, C.-B., and Phoon, K.-K. (2015). "Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0001227, 04014096.

Li, X. Y., Zhang, L. M., and Li, J. H. (2016a). "Using conditioned random field to characterize the variability of geologic profiles." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0001428, 04015096.

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

Liu, L.-L., Cheng, Y.-M., and Zhang, S. H. (2017a). "Conditional random field reliability analysis of a cohesion-frictional slope." *Comput. Geotech.*, 82, 173–186.

Liu, W. F. and Leung, Y. F. (2017). "Characterising three-dimensional anisotropic spatial correlation of soil properties through *in situ* test results." *Under review.*

Liu, W. F., Leung, Y. F., and Lo, M. K. (2017b). "Integrated framework for characterization of spatial variability of geological profiles." *Can. Geotech. J.*, 54(1), 47–58.

Lloret-Cabot, M., Hicks, M. A., and van den Eijnden, A. P. (2012). "Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using kriging." *Géotechnique Letters*, 2(3), 123–127.

Lo, M. K. and Leung, Y. F. (2017a). "Probabilistic analyses of slopes and footings with spatially variable soils considering cross-correlation and conditioned random fields." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0001720, 04017044.

Lo, M. K. and Leung, Y. F. (2017b). "Risk assessment of geotechnical performance on spatially varying soil by the use of sensitivity index." *Geo-Risk 2017*, Geotechnical special publication 284, J. Huang et al., eds., ASCE, Denver, CO, 467–476.

Miro, S., Hartmann, D., and Schanz, T. (2014). "Global sensitivity analysis for subsoil parameter estimation in mechanized tunneling." *Comput. Geotech.*, 56, 80–88.

Mollon, G., Dias, D., and Soubra, A. (2011). "Probabilistic analysis of pressurized tunnels against face stability using collocation-based stochastic response surface method." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0000443, 385–397.

Packham, N. and Schmidt, W. M. (2010). "Latin hypercube sampling with dependence and applications in finance." *Journal of Computational Finance*, 13(3), 81–111.

Phoon, K. K. and Kulhawy, F. H. (1999). "Characterization of geotechnical variability." *Can. Geotech. J.*, 36(4), 612–624.

Popescu, R., Prevost, J. H., and Deodatis, G. (2005). "3D effects in seismic liquefaction of stochastically variable soil deposits." *Géotechnique*, 55(1), 21–31.

Richards, F. J. (1959). "A flexible growth function for empirical use." *Journal of experimental Botany*, 10(2), 290–301.

Sobol', I. M. (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Math. Comput. Simulation*, 55, 271–280.

Storn, R. and Price, K. (1997). "Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces." *Journal of Global Optimization*, 11(4), 341–359.

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

List of Tables

TABLE 1. Slope geometries and shear strength variability parameters adopted in Sobol' index analyses

Parameters		c_u soils	c_u soils (double scale)	$c-\phi$ soils	$c-\phi$ soils (double scale)
Slope angle, β (for all cases)		20°, 30°, 35°, 40°			
Soil unit weight (for all cases)		20 kN/m^3			
Slope Height, H		5 m	10 m	10 m	20 m
Deterministic FS		1, 1.5, 2	1.5	1, 1.5, 2	1.5
Variability of shear	CV$_{cu}$	0.15, 0.4	0.15, 0.4	-	-
strength parameters	CV$_{\phi}$	-	-	0.05, 0.1	0.05, 0.1
	$\theta_{\ln,x}$ (for all cases)	$0.5W$, $2W$, ∞			
	$\theta_{\ln,y}$ (for all cases)	$0.25H$, $0.5H$			

TABLE 2. Six sample locations from two boreholes for a slope in c_u soils

n (No. of samples)	Borehole number	Depth from slope top (m)	$S(e_1, ..., e_n)$	Cumulative % reduction in SD
1	1	6.25	0.24	12.8
2	1	8.75	0.37	20.6
3	2	4.2	0.47	27.2
4	1	10.75	0.56	33.7
5	1	13.25	0.63	39.2
6	2	2.2	0.69	44.3

TABLE 3. Conditional mean coefficients for multiple samples in c_u slope

Order of coefficient	r	Value
0th order	r_0	1.318
1st order	r_1	0.080
	r_2	0.057
	r_3	0.050
	r_4	0.047
	r_5	0.038
	r_6	0.039
2nd order	r_{11}	-0.0078
	r_{12}	0.0012
	r_{13}	-0.0097
	r_{14}	0.0105
	r_{15}	0.0098
	r_{16}	-0.0095
	r_{22}	-0.0103
	r_{23}	0.0046
	r_{24}	-0.0019
	r_{25}	0.0118
	r_{26}	-0.0028
	r_{33}	-0.0026
	r_{34}	0.0078
	r_{35}	0.0137
	r_{36}	-0.0073
	r_{44}	-0.0096
	r_{45}	0.0016
	r_{46}	0.0056
	r_{55}	-0.0090
	r_{56}	0.0085
	r_{66}	0.0023

List of Figures

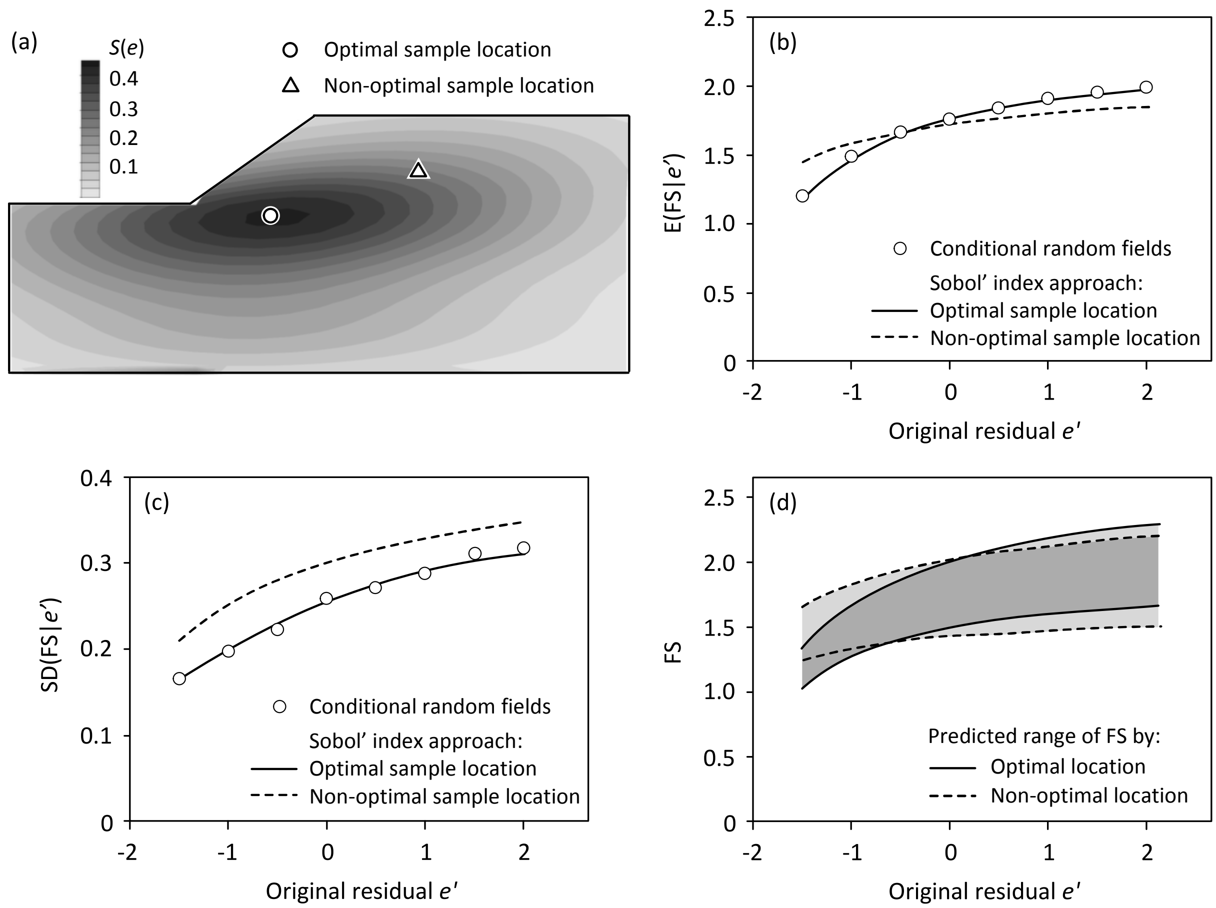

FIG. 1. Comparison between Sobol' index approach and individual conditional random field simulations

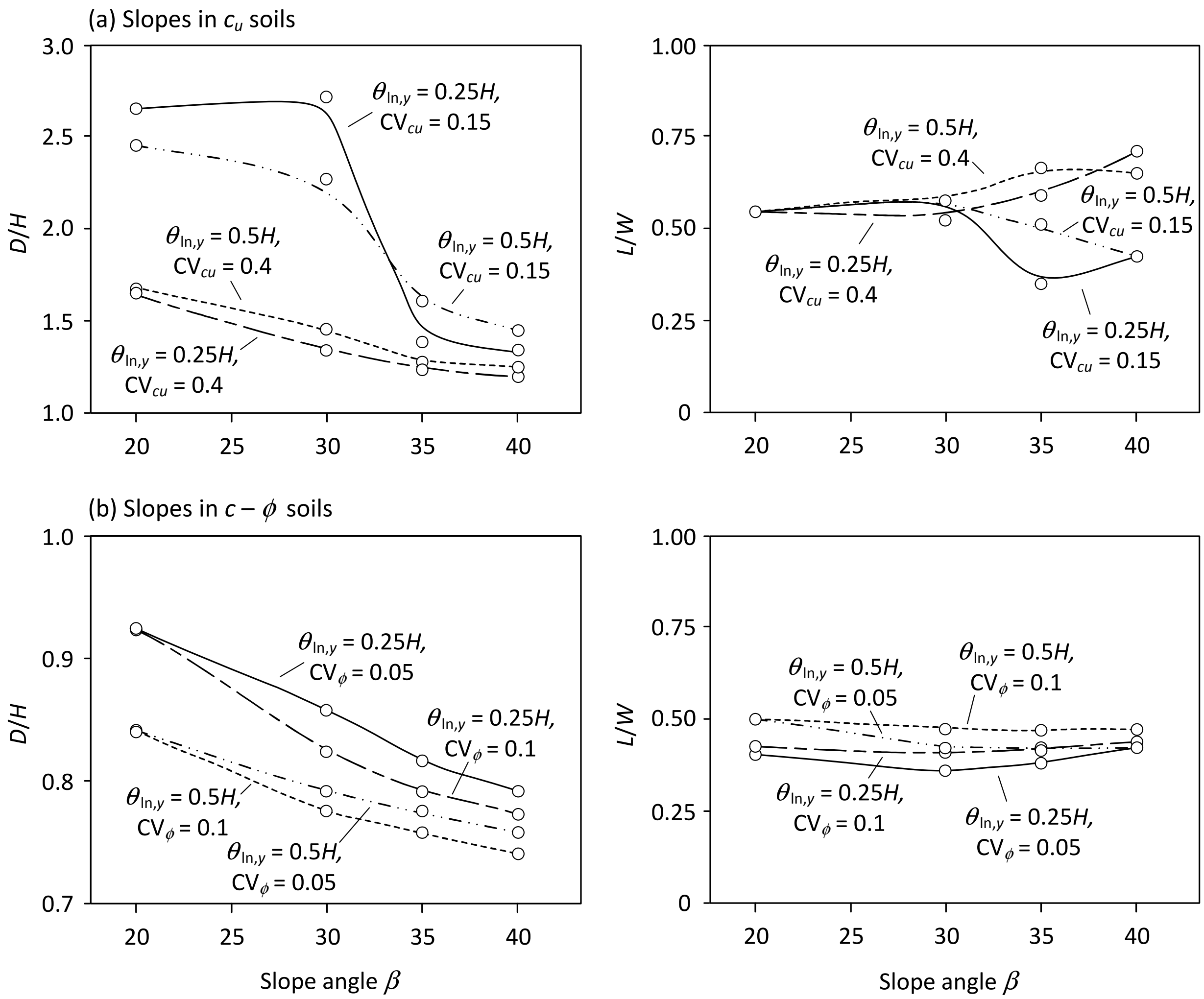

FIG. 2. Optimal sampling locations normalized by slope height (D/H) and width (L/W) for slopes in (a) c_u soils and (b) $c-\phi$ soils. Data points represent the results for four slope angles and the lines are trendlines

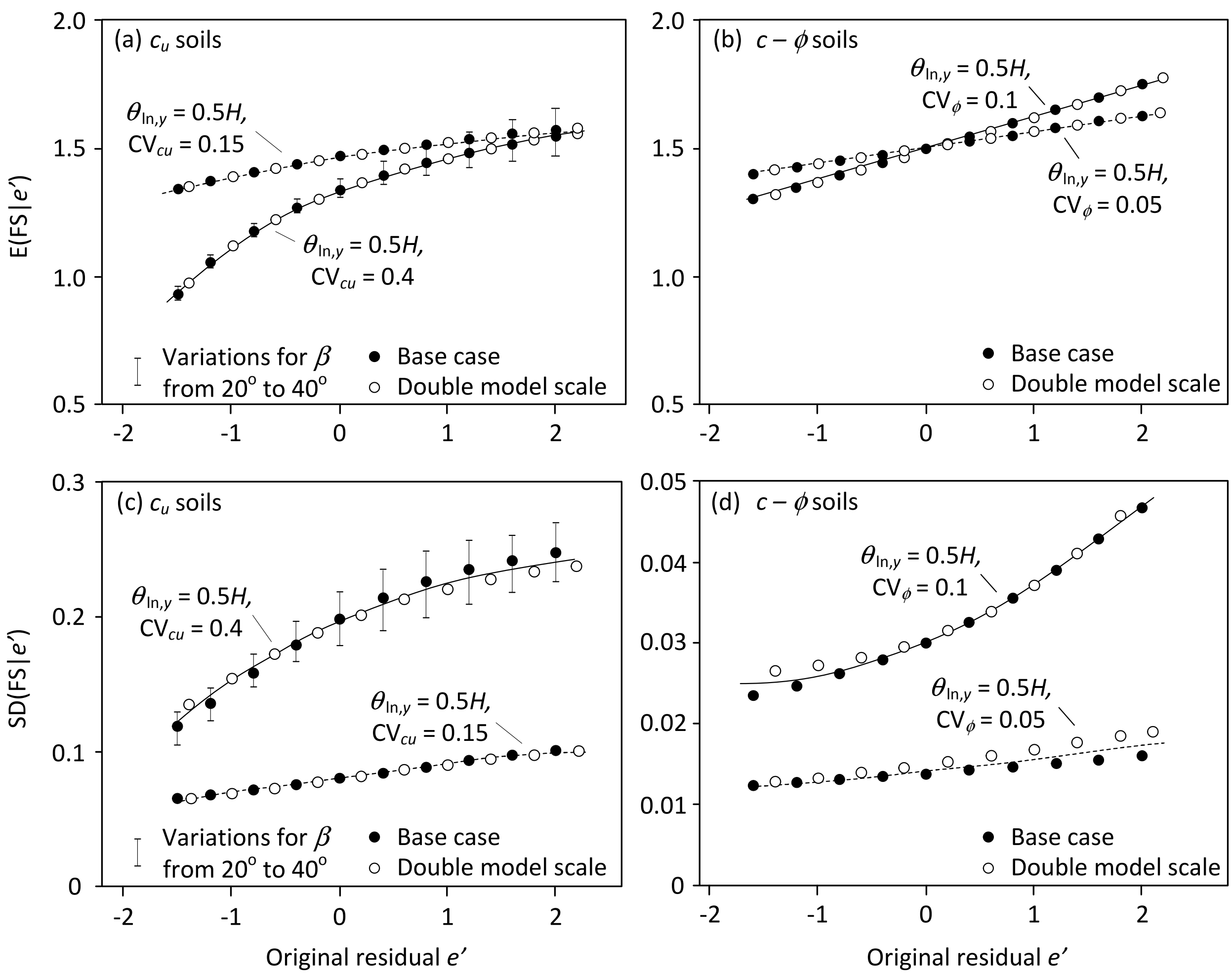

FIG. 3. Conditional mean and standard deviation functions with deterministic FS=1.5, slope angle varying from 20° to 40° and different model scales

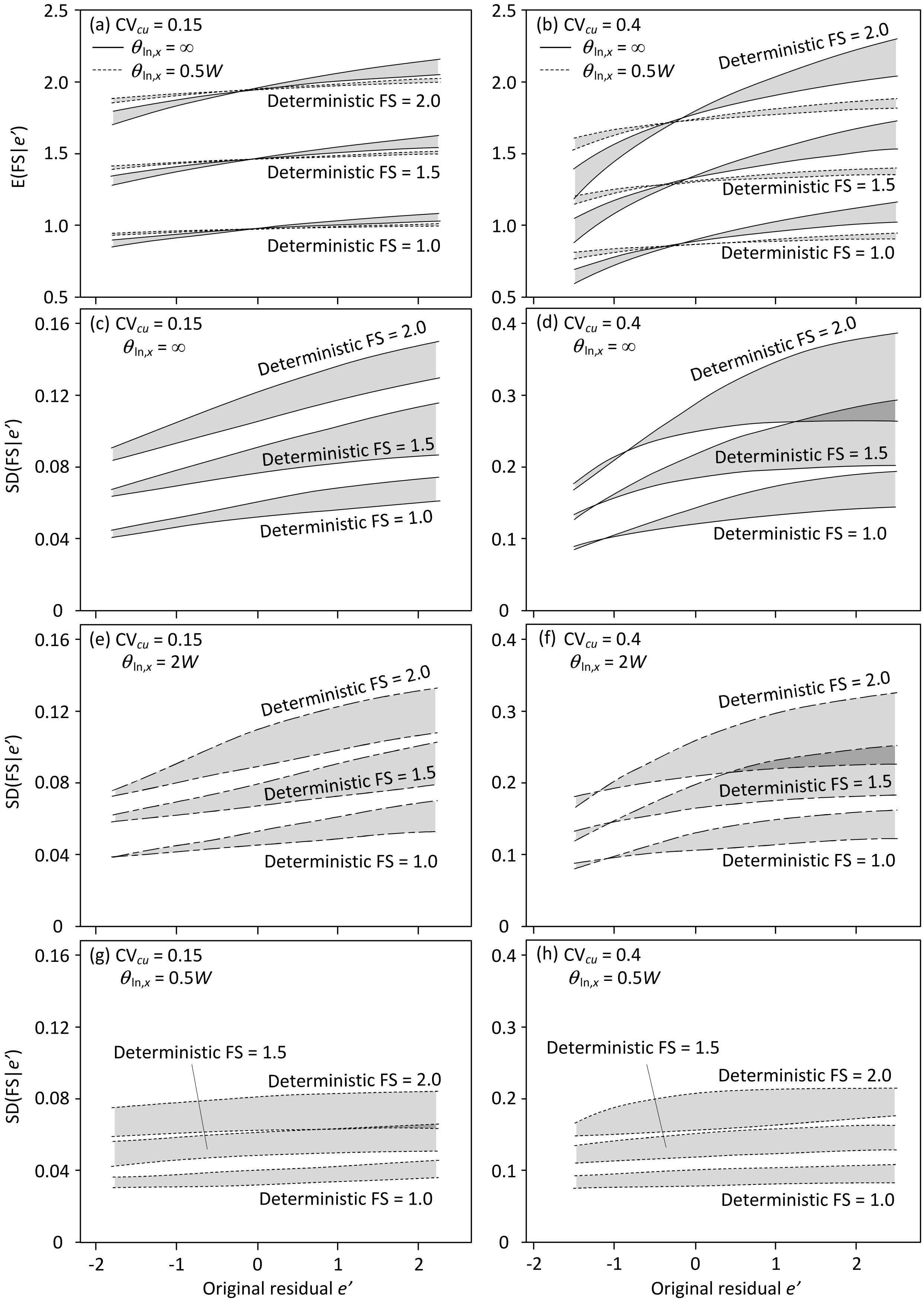

FIG 4. Conditional mean and standard deviation functions for slopes in c_u soils, with gray areas bounded by $\theta_{\ln,y}$ = 0.25*H* and 0.5*H*. $\theta_{\ln,y}$ = 0.5*H* results in higher SD and mean FS for positive *e*' (*x*-axes are identical for sub-figures)

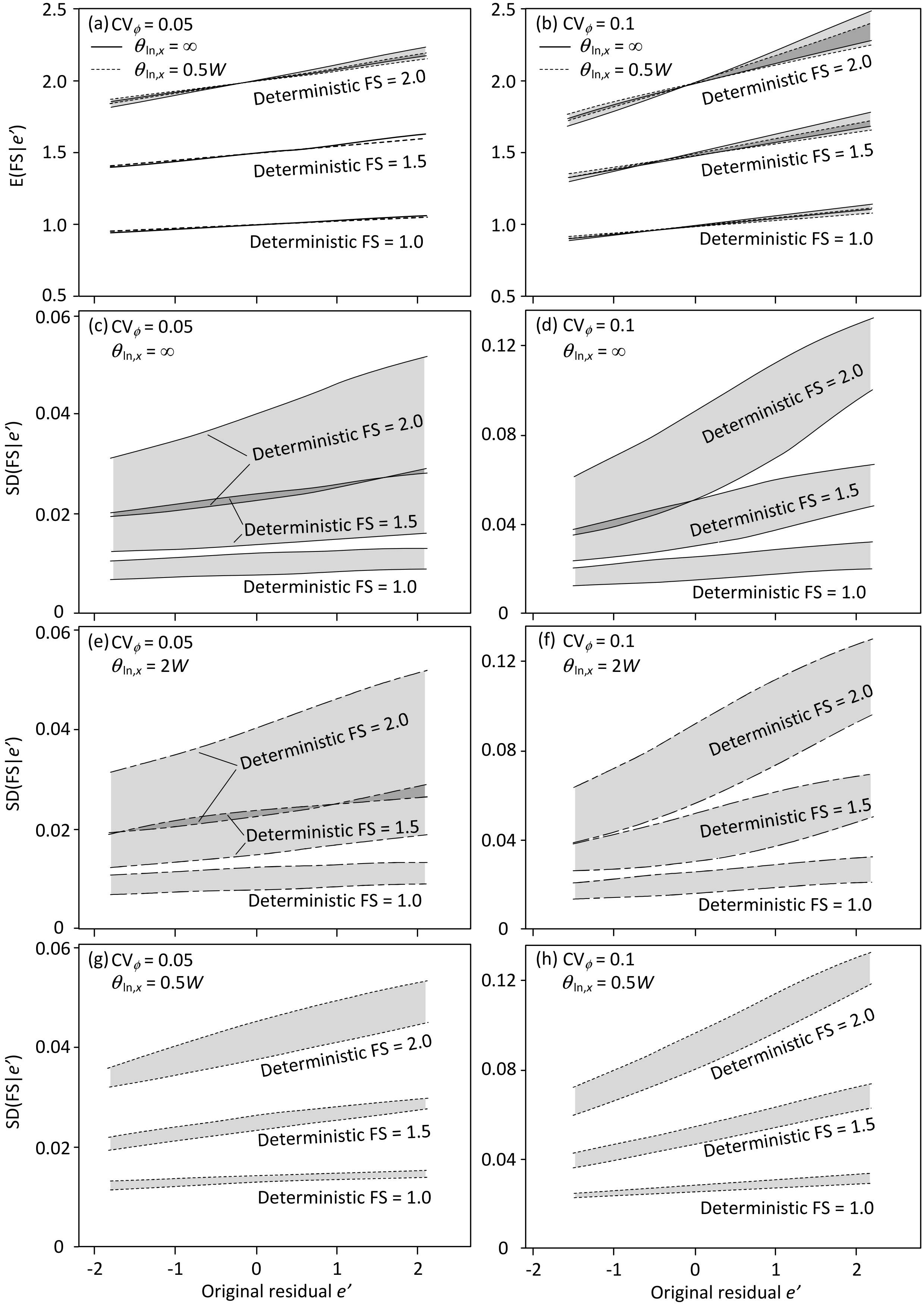

FIG 5. Conditional mean and standard deviation functions for slopes in $c-\phi$ soils, with gray areas bounded by $\theta_{\ln,y}$ = 0.25*H* and 0.5*H*. $\theta_{\ln,y}$ = 0.5*H* results in lower SD and higher mean FS for positive *e*' (*x*-axes are identical for sub-figures)

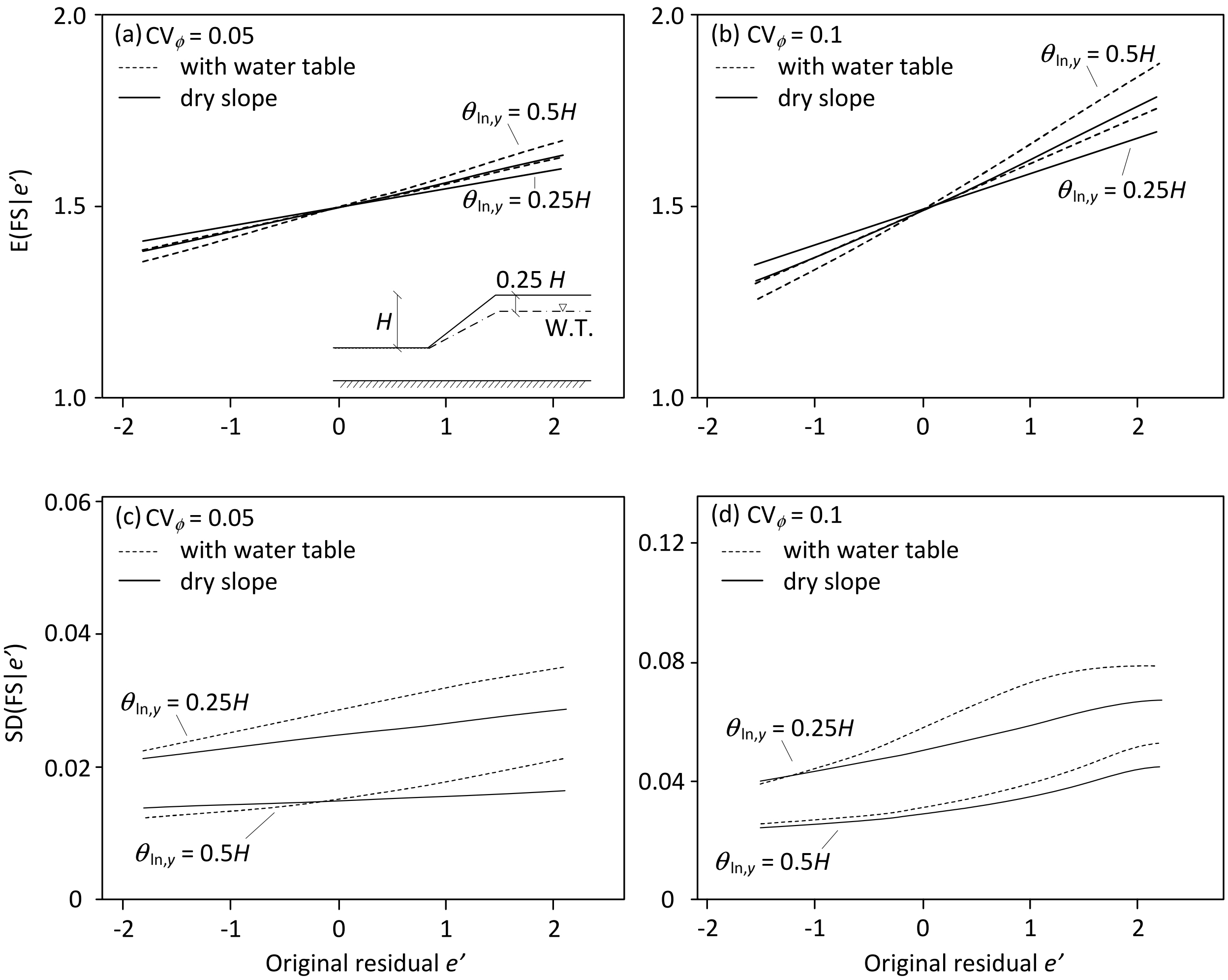

FIG. 6. Influence of water table on sensitivity functions

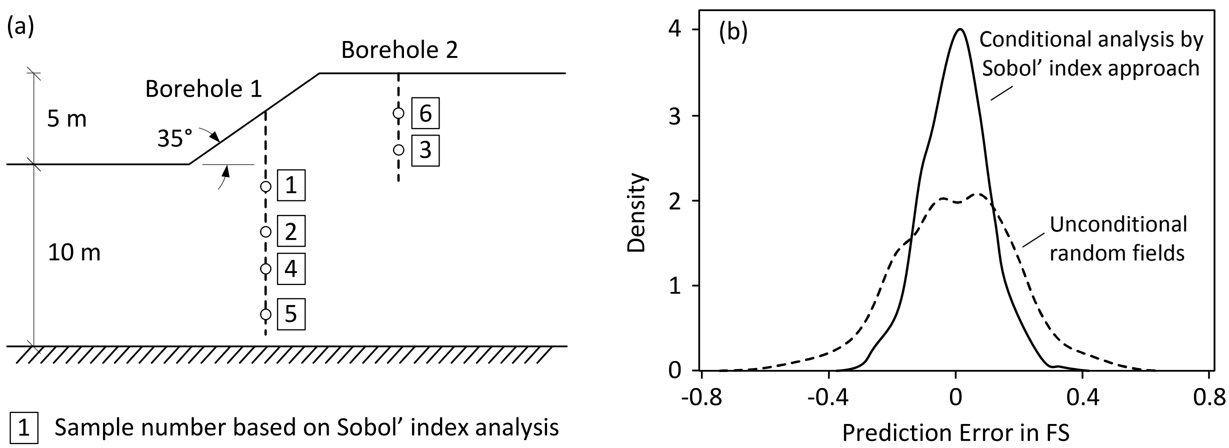

FIG. 7. (a) Locations of six samples along two boreholes; (b) Prediction errors in FS comparing 1,000 realizations with unconditional and conditional mean estimates

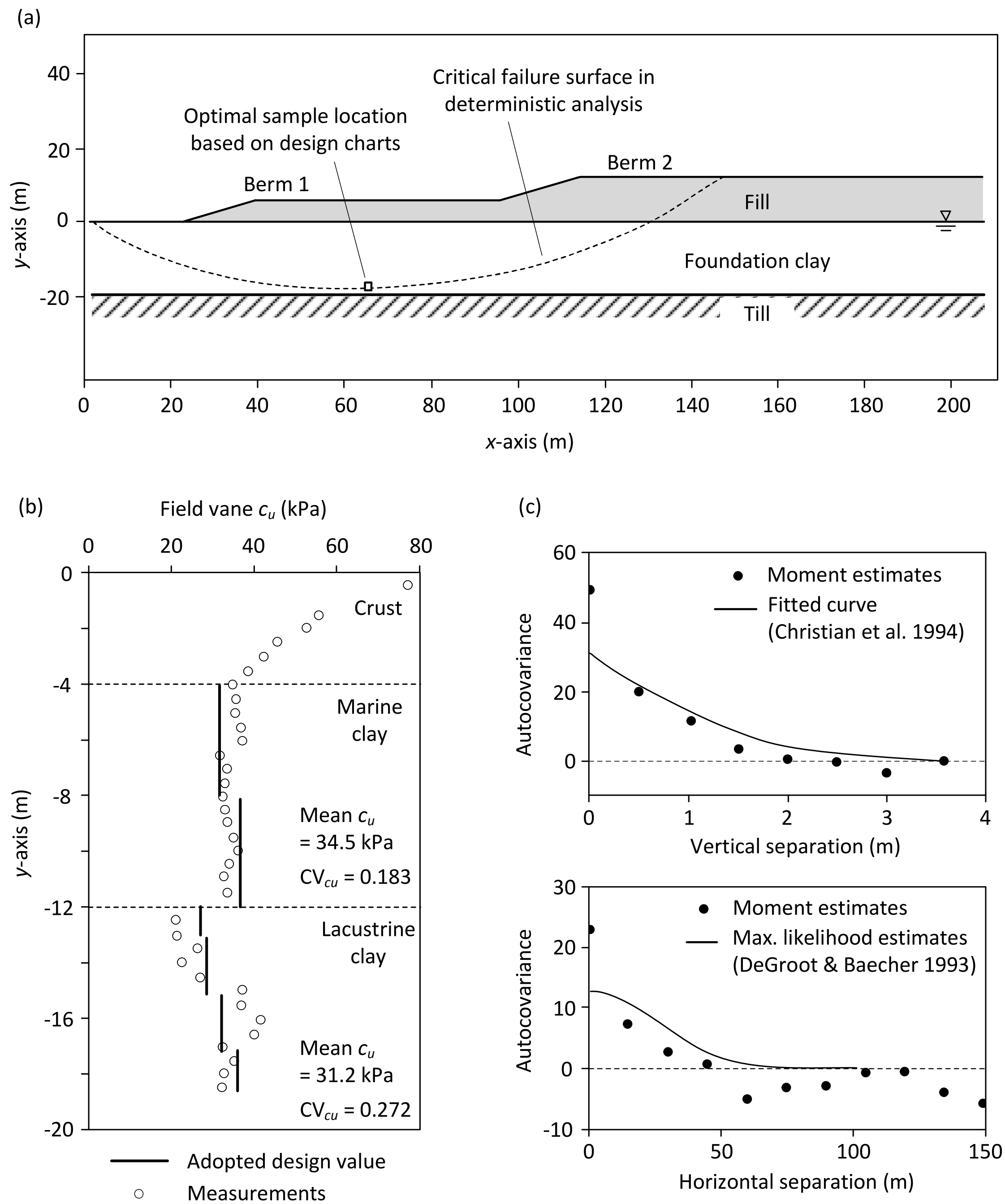

FIG. 8. (a) Cross section of embankment in James Bay project; (b) Variations of c_u in foundation clay; (c) Autocovariance functions for c_u in vertical and horizontal directions (adapted from Christian et al. 1994 and DeGroot and Baecher 1993)