Efficient simulation of multivariate three-dimensional cross-correlated random fields conditioning on non-lattice measurement data

Zhiyong Yang[1], Xueyou Li[2] and Xiaohui Qi*[3]

Abstract

It is challenging to simulate large-scale or fine-resolution multivariate three-dimensional (3D) cross-correlated conditional random fields because of computational issues such as inverting, storing or Cholesky decomposition of large correlation matrices. Recently, an efficient univariate 3D conditional random field simulation method was developed based on the separable assumption of the autocorrelation functions in the vertical and horizontal directions. The developed simulation method allows for Kronecker-product derivations of the large correlation matrices and thus does not need to invert and store large matrices. Moreover, it can handle univariate non-lattice data (e.g., all soundings measure the data of one soil property and there exists missing data at some depths at some soundings). It may be more common to see multivariate non-lattice data (e.g., all soundings measure the data of multiple soil properties and there exists missing data of some properties at some depths at some soundings) in practical site investigations. However, the proposed method is not applicable to multivariate non-lattice data because it cannot directly account for the cross-correlation among different variables. The purpose of the current paper is to extend the previous method to accommodate the multivariate non-lattice data. The extended method still takes advantage of the Kronecker-product derivations to avoid the mathematical operation of the large correlation matrices. A simulated example is adopted to illustrate the effectiveness of the extended method.

Keywords: 3D site characterization; spatial variability; non-lattice data; cross-correlated conditional random field; Gibbs sampler; Monte Carlo simulation.

[1] Assistant Professor, School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China. Email: yangzhy85@mail.sysu.edu.cn
[2] Professor, School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China. Email: lixueyou@mail.sysu.edu.cn
[3] (Corresponding Author) Lecturer, Department of Mechanical and Construction Engineering, Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK. Email: xiaohui.qi@northumbria.ac.uk

1. Introduction

Spatial variability of soil properties has a significant influence on the failure mechanism and reliability of geotechnical structures such as foundations (e.g., Fenton and Griffiths 2002; Popescu et al. 2005; El Haj et al. 2019), tunnels (e.g., Song et al. 2011; Cheng et al. 2019; Wu et al. 2021), retaining wall (e.g., Fenton et al. 2005; Hu and Ching 2014, 2015) and slopes (e.g., Jiang et al. 2015; Li et al. 2015; Qi and Li, 2018; Li et al. 2016a, 2019b; Gong et al. 2020). It is important to account for the spatial variability of soil properties when performing the geotechnical reliability and risk analysis. The conventional geotechnical reliability analysis often simplifies the spatial variability of soil property as a one-dimensional (1D) (Wang et al. 2011; Li et al. 2014) or two-dimensional (2D) problems (Griffiths and Fenton 2004; Huang et al. 2013; Ji et al. 2012) although soil properties generally exhibit a three-dimensional (3D) spatial variability. Some recent studies have recognized that the sophisticated 3D spatial variability cannot be fully represented by the simplified 1D or 2D spatial variability (e.g., Xiao et al. 2016; Liu et al. 2018; Varkey et al. 2019; Li et al. 2019b; Kawa and Puła 2019). To effectively capture the failure mechanism and accurately estimate the failure risk of geotechnical structure, it entails the modeling of the 3D spatial variability, which can be represented by a 3D random field. Moreover, site investigation for many realistic sites is generally performed by various test methods such as cone penetration test (CPT), standard penetration test (SPT), vane shear test (VST), and/or laboratory triaxial compression test. The associated sites often contain observed/measured data for multiple soil property parameters. It is also vital to simulate multivariate cross-correlated 3D random fields conditioning on the observed multivariate data.

Generally, the multivariate cross-correlated 3D conditional random field simulation might be subject to two difficulties: cross-correlated random field simulation and conditional random field simulation. In terms of the former, various cross-correlated random field simulation methods have been proposed. The conventional random field simulation method such as Cholesky decomposition, local average subdivision (e.g., Fenton and Vanmarcke 1990), Karhunen-Loève expansion (e.g.,

Phoon et al. 2002) and spectral representation (e.g., Jha and Ching 2012) methods can be readily extended to simulate the cross-correlated random field by taking the cross-correlated matrix into consideration during the random field generation process. For instance, Robin et al. (1993) proposed a Fourier transform-based cross-correlated random field simulation method. Vořechovský (2008) proposed a series expansion-based cross-correlated random field simulation method. Zhao and Wang (2018) proposed a Bayesian compressive sampling and Karhunen–Loève (KL) expansion-based cross-correlated random field simulation method. These methods are frequently used for simulating the 2D cross-correlated random fields in geotechnical engineering (e.g., Cho 2010; Fenton and Griffiths 2003; Jiang et al. 2015). These methods generally are efficient when simulating 3D cross-correlated unconditional random fields. However, they might be highly computationally intensive for simulating 3D cross-correlated conditional random fields especially for cases with large-scale domains. In terms of the latter, various conditional random field simulation methods such as Kriging methods (e.g., Liu et al. 2017; Lo and Leung 2017; Huang et al. 2019), patching algorithm-based method (Ou-Yang et al. 2021), Bayesian methods (e.g., Li et al. 2016b; Jiang et al. 2018) and some non-parametric methods (e.g., Zhao et al. 2020; Zhao and Wang 2020, 2021) and have been proposed. These methods are commonly utilized to simulate a 2D conditional random field although they might be extended to simulate a 3D conditional random field for a domain with a relatively small scale or with a low resolution (e.g., Li et al. 2016b; Li et al. 2016c; Xiao et al. 2018; Cai et al. 2019). Nevertheless, they are not applicable to the 3D conditional random field simulation for a domain with a large scale or with a fine resolution. In fact, due to the significant computational cost required in 3D problem, even the simulation of a 3D unconditional random field over a large domain with a high resolution is challenging with the existing random field simulation methods such as Cholesky decomposition, Karhunen-Loève expansion (e.g., Phoon et al. 2002) and spectral representation (e.g., Jha and Ching 2012). The local average subdivision (e.g., Fenton and Vanmarcke 1990) is computational efficiency for 3D problem; however, it requires a uniform grid to simulate the random field. Recently, this computational challenge of 3D unconditional random field simulation is

addressed by Cheng et al. (2019) and Li et al. (2019a). Cheng et al. (2019) suggested simulating each sub-domain at the 3D space sequentially by conditioning on neighboring sub-domains with distances less than a few times of the scale of fluctuation (SOF). This method still requires considerable computational cost if SOF is large. Li et al. (2019a) proposed a stepwise covariance matrix decomposition (CMDC) method with the separable assumption of the autocorrelation function in the horizontal and vertical directions. The CMDC method requires the 3D random field to be simulated follows a 3D lattice structure so that it can take advantages of the Kronecker-product derivations to decompose the correlation matrices. As a result, the CMDC does not need to invert and store large matrices and its computational cost is independent of the SOF. Noted that the above-mentioned studies (Cheng et al. 2019; Li et al. 2019a) did not address the simulation of a 3D conditional random field. Zhao and Wang (2020, 2021) developed non-parametric methods for simulating 3D conditional random field samples from measurement data. Although this method bypasses the usage of parametric correlation model, it is computationally inefficient. Ching et al. (2020) further proposed a method to address the computational inefficiency faced by the 3D conditional random field based on two assumptions of (a) the separable auto-correlation and (b) the regular-3D-lattice-distributed sounding data. The first assumption might be not strong while the second assumption is a little strong because the measured sounding data in geotechnical practice generally have a non-lattice 3D structure. To relax the second assumption, Yang and Ching (2020) further proposed an efficient univariate 3D conditional random field simulation method. Their method is admitted by the 3D non-lattice sounding data. However, both Zhao and Wang (2020, 2021), Ching et al. (2020) and Yang and Ching (2020) did not address the simulation of *multivariate 3D cross-correlated conditional* random fields.

In practice, it is not uncommon that the measured sounding data are multivariate and not with a regular 3D lattice structure. Consider a schematic diagram for sounding data at a typical site shown in Figure 1. The site investigations are performed by various tests including VST, SPT, CPT and laboratory triaxial compression test. As a result, multiple soil property parameters are measured at this site. If the random fields of multiple soil strength parameters such as undrained shear strength s_u,

cohesion c, friction angle ϕ and elastic modulus E are concerned, the measured multiple soil property data by VST, SPT, CPT and laboratory triaxial compression might be converted to the soil strength parameters through certain transformation relationships (e.g., Bjerrum 1973; Kulhawy and Mayne 1990; Ching and Phoon 2014). For illustration, it is supposed that the cross-correlated 3D conditional random fields of two soil property parameters X_1 and X_2 at this typical site are to be simulated. X_1 and X_2 can be the measured and/or transformed soil strength data based on the measured sounding data obtained from the VST, SPT, CPT or laboratory triaxial compression tests. Suppose that the sounding S1, S2, (S3, S4) and (S5, S6) contain the data collected from VST, laboratory triaxial compression test, CPT test and SPT test, respectively. It is not uncommon that different test methods (e.g., SPT, VST, CPT and laboratory triaxial compression tests) have different sampling depth intervals. Moreover, the same test method (e.g., CPT) may also have different sampling depths at different soundings (e.g., S3 and S4). When these non-lattice test data are converted to X_1 and X_2 data, X_1 and X_2 data are also not regularly distributed. That is to say, some X_1 and X_2 data are missing or have unequal spacing. For example, the missing data that cause the non-lattice structure of X_1 and X_2 are shown as open circles and open triangles in Figure 1. This non-lattice structure hinders the application of the Kronecker product in decomposing the auto-covariance matrix, hence prohibits an effective simulation of multivariate 3D cross-correlated conditional random fields.

This paper aims at extending the Yang and Ching (2020)'s method to simulate the multivariate 3D cross-correlated random field conditioning on irregularly distributed (non-lattice) sounding data. The extended method does not need to invert and store large matrices, either. This paper starts with the problem statement, followed by a brief review of the univariate 3D conditional random field simulation method proposed in Yang and Ching (2020). The modified method is then illustrated with the typical site data shown in Figure 1. A simulated example is employed to demonstrate the effectiveness of the modified method.

2. Problem statement

128 *2.1. Assumptions*

129 The study adopts the following four assumptions:

130 1. The cross-correlation and auto-correlation among various soil property parameters are separable.
131 This implies that the type of auto-correlation function (e.g., single exponential function) and scale
132 of fluctuation for all the soil property parameters are the same and the cross-correlation structure
133 between each pair of simulated random fields is simply defined by a cross-correlation coefficient
134 (Vořechovský 2008). Fenton and Griffiths (2003), Fenton et al. (2005) and Cho (2010) made
135 similar assumptions (identical auto-correlation structure for different soil parameters) by arguing
136 that the spatial correlation of a soil is governed largely by the spatial variability in its source
137 materials, weathering patterns, stress, sedimentation and formation history, etc. so that one would
138 expect that all soil properties will vary similarly between the two points. This assumption is
139 frequently used in the geotechnical field for simulating the multivariate cross-correlated random
140 field (Peng et al. 2017; Zhu et al. 2017; Li et al. 2019a). There is no compelling evidence to
141 support or negate this assumption, and this assumption is adopted in the current study to serve as
142 a baseline for further studies with more complicated auto-correlation models.
143 2. The auto-correlation functions in the vertical and horizontal directions are separable. This is also
144 a common assumption in 3D random field simulation (Vanmarcke 1977; Stuedlein et al. 2012;
145 Xiao et al. 2018; Li et al. 2019a; Yang and Ching 2020; Ching et al. 2020). There is no compelling
146 evidence to support or negate this assumption, either. This assumption is used mainly for reducing
147 the matrices' size and improving the computational efficiency.
148 3. The underlying unconditional random field is stationary and normally distributed. A non-
149 stationary random field with a trend can be transformed to a stationary random field by removing
150 the trend component (Jaksa et al. 1999; Uzielli et al. 2005). A non-normal random field can be
151 transformed from the normal random field with certain transformation methods such as Nataf
152 transformation.
153 4. The random field parameters such as auto-covariance parameters (e.g., standard deviation, SOF,

sample-path smoothness, etc.) and trend functions have been properly determined (i.e., they are prescribed) by suitable methods such as moment methods (e.g., Uzielli et al. 2005; Cami et al. 2020), maximum likelihood estimation (MLE) (e.g., DeGroot and Baecher 1993; Xiao et al. 2018; Qi and Liu 2019; Cami et al. 2020) or Bayesian methods (e.g., Cao and Wang 2014; Ching et al. 2020; Ching et al. 2021) before simulating the multivariate cross-correlated 3D random fields. The estimated random field parameters might be justified by comparing the estimated values with the empirical values summarized in literature and/or comparing the simulated random field with the measurement data.

The first assumption ensures that the auto-correlations of various soil properties can be expressed as the same auto-correlation matrix and the cross-correlation among the soil properties over the random field domain can be expressed as a matrix. This allows the correlation of all the soil parameters over the random field domain to be decomposed into a cross-correlated matrix and an auto-correlation matrix with the Kronecker product. The size of the auto-correlation matrix may still be large for the 3D problem with a large domain or a fine resolution. The second assumption is then introduced to decompose the auto-correlation matrix into a horizontal auto-correlation matrix and a vertical auto-correlation matrix, which have relatively small sizes. The first two assumptions are made mainly for taking full advantage of the Kronecker product to increase the computational efficiency. The last two assumptions are made mainly for simplifying the analysis. Nevertheless, the last two assumptions can be relaxed easily by combining the proposed method with the mentioned methods in the two assumptions, which is out of the scope of this study.

2.2. 3D lattice for the soundings

Let us take the 3D underground domain with sounding data in Figure 1 as an example. This 3D domain is discretized into a large number of random field elements (RFEs) (the light grey mesh in Figure 1). The RFEs play a role to partition the 3D domain and are used for simulating the multivariate cross-correlated 3D conditional random fields. When reliability analysis of geotechnical structures considering the 3D spatial variability is conducted with the finite-element method or finite-difference

method, the simulated multivariate cross-correlated 3D conditional random fields at the RFEs can be mapped into the 3D finite-element or 3D finite-difference mesh. Suppose that there are n_h soundings ($n_h = 6$ in Figure 1), and at each sounding there are n_v sounding depths ($n_v = 10$ in Figure 1). The data point at each sounding depth contains m number of soil property parameters (m =2 in Figure 1). There are in total $N = n_h \times n_v \times m$ sounding data (N = 120 in Figure 1). Let $\underline{\xi}_i$ contains the data for the i-th sounding. $\underline{\xi}_i = [\underline{X}_{i,1}\ \underline{X}_{i,2}\ \ldots\ \underline{X}_{i,nv}]^T$ is an $((m\times n_v)\times 1)$ column vector where $\underline{X}_{i,j} = [X_1\ X_2\ \ldots\ X_m]$ denotes the data point at the i-th sounding at the j-th depth. Let $\boldsymbol{\xi}$ contain all sounding data, including n_o observed data (filled markers in the figure, denoted by $\boldsymbol{\xi}^o$) and n_u un-observed data (open markers in the figure, denoted by $\boldsymbol{\xi}^u$). $\boldsymbol{\xi}$ is an (N×1) column vector denoting the union of $\boldsymbol{\xi}^o$ and $\boldsymbol{\xi}^u$. $\boldsymbol{\xi}$ can be assembled by stacking the soundings data vertically, $\boldsymbol{\xi} = [\underline{\xi}_1^T\ \underline{\xi}_2^T\ \ldots \underline{\xi}_{nh}^T]^T$. Let us denote the covariance matrix for $\boldsymbol{\xi}$ by $\mathrm{Var}(\boldsymbol{\xi}) \in \mathbf{R}^{N\times N}$. Base on the assumption of separability between the auto-correlation and the cross-correlation among the m number of parameters, $\mathrm{Var}(\boldsymbol{\xi})$ can be expressed as

$$\mathrm{Var}(\boldsymbol{\xi}) = \begin{bmatrix} \mathbf{R}\left(\underline{\xi}_1,\underline{\xi}_1\right) & \mathbf{R}\left(\underline{\xi}_1,\underline{\xi}_2\right) & \cdots & \mathbf{R}\left(\underline{\xi}_1,\underline{\xi}_{n_h}\right) \\ & \mathbf{R}\left(\underline{\xi}_2,\underline{\xi}_2\right) & & \\ & & \ddots & \mathbf{R}\left(\underline{\xi}_{n_h-1},\underline{\xi}_{n_h}\right) \\ \text{SYM.} & & & \mathbf{R}\left(\underline{\xi}_{n_h},\underline{\xi}_{n_h}\right) \end{bmatrix} \otimes \mathbf{C} \qquad (1.)$$

where $\mathbf{A}\otimes\mathbf{B}$ denotes the Kronecker product between two matrices $\mathbf{A}$ and $\mathbf{B}$

$$\mathbf{A}_{p\times q} \otimes \mathbf{B}_{r\times s} = \begin{bmatrix} a_{11}\mathbf{B}_{r\times s} & a_{12}\mathbf{B}_{r\times s} & \mathsf{L} & a_{1q}\mathbf{B}_{r\times s} \\ a_{21}\mathbf{B}_{r\times s} & a_{22}\mathbf{B}_{r\times s} & \mathsf{L} & a_{2q}\mathbf{B}_{r\times s} \\ \mathsf{M} & \mathsf{M} & \mathsf{O} & \mathsf{M} \\ a_{p1}\mathbf{B}_{r\times s} & a_{p2}\mathbf{B}_{r\times s} & \mathsf{L} & a_{pq}\mathbf{B}_{r\times s} \end{bmatrix}_{pr\times qs} \qquad (2.)$$

and $\mathbf{R}(\underline{\xi}_i,\underline{\xi}_j) \in \mathbf{R}^{nv\times nv}$ is the correlation matrix between $\underline{\xi}_i$ and $\underline{\xi}_j$; $\mathbf{C}$ is the cross-covariance matrix between various soil property parameters

$$\mathbf{C}=\begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,m}\sigma_1\sigma_m \\ & \sigma_2^2 & & \\ & & \ddots & \rho_{m-1,m}\sigma_{m-1}\sigma_m \\ \text{SYM.} & & & \sigma_m^2 \end{bmatrix} \tag{3.}$$

where σ_i is the standard deviation of the i-th soil property parameter; $\rho_{i,j}$ is the correlation coefficient between the i-th and the j-th soil property parameters. Following Yang and Ching (2020), the current study also adopts the separable assumption of auto-correlation function in the horizontal and vertical directions, i.e.,

$$\rho(\Delta x,\Delta y,\Delta z)=\rho_h(\Delta x,\Delta y)\times\rho_v(\Delta z) \tag{4.}$$

where $\rho(\Delta x, \Delta y, \Delta z)$ represents the 3D auto-correlation function; $\rho_h(\Delta x, \Delta y)$ represents the 2D auto-correlation function for the horizontal direction (i.e., x-y plane); $\rho_v(\Delta z)$ represents the 1D auto-correlation function for the vertical direction (i.e., z direction). According to the separability assumption in auto-correlation, Eq. (1) can be further expressed as the following Kronecker product:

$$\mathrm{Var}(\xi)=\left[\mathbf{R}_{h,\xi}\otimes\mathbf{R}_{v,\xi}\right]\otimes\mathbf{C} \tag{5.}$$

where $\mathbf{R}_{h,\xi}\in\mathbf{R}^{nh\times nh}$ and $\mathbf{R}_{v,\xi}\in\mathbf{R}^{nv\times nv}$ denote the horizontal and vertical correlation matrices for ξ, respectively:

$$\mathbf{R}_{h,\xi}=\begin{bmatrix} 1 & \rho_h(x_1-x_2,y_1-y_2) & \rho_h(x_1-x_3,y_1-y_3) & \cdots & \rho_h(x_1-x_{n_h},y_1-y_{n_h}) \\ & 1 & \rho_h(x_2-x_3,y_2-y_3) & \cdots & \rho_h(x_2-x_{n_h},y_2-y_{n_h}) \\ & & 1 & & \\ & & & \ddots & \vdots \\ \text{SYM.} & & & & 1 \end{bmatrix} \tag{6.}$$

$$\mathbf{R}_{v,\xi}=\begin{bmatrix} 1 & \rho_v(z_1-z_2) & \rho_v(z_1-z_3) & \cdots & \rho_v(z_1-z_{n_v}) \\ & 1 & \rho_v(z_2-z_3) & \cdots & \rho_v(z_2-z_{n_v}) \\ & & 1 & & \\ & & & \ddots & \vdots \\ \text{SYM.} & & & & 1 \end{bmatrix} \tag{7.}$$

(x_i, y_i) are the x- and y-coordinates for the i-th sounding (i = 1, 2, …, n_h), and z_j is the z-coordinate

for the j-th depth at each sounding (j = 1, 2, …, n_v).

2.3. 3D lattice for random field elements

Let $\boldsymbol{\xi}'$ contain the soil properties at the centroids of all random field elements (RFEs). Taking Figure 1 as an example, there are n'_h elements in the x-y plane ($n'_h = 5\times5 = 25$ in Figure 1) and n'_v elements in the z direction ($n'_v = 10$ in Figure 1). Each RFE contains m soil property parameters. As a result, there are $N' = n'_h \times n'_v \times m$ number of soil data for all the RFEs ($N' = 500$ in Figure 1). Note that (n'_h, n'_v) are not necessary to be the same as (n_h, n_v). $\boldsymbol{\xi}'$ can be represented as an ($N'\times1$) column vector by stacking the data at each column of RFEs, i.e., $\boldsymbol{\xi}' = [\xi'_1{}^T\ \xi'_2{}^T\ \ldots\xi'_{n'h}{}^T]^T$, where ξ'_i is an $((m\times n'_v)\times 1)$ column vector that contains the ξ values for the RFEs at the i-th column in Figure 1. Let us denote the covariance matrix for $\boldsymbol{\xi}'$ by $\mathrm{Var}(\boldsymbol{\xi}')\in\mathbf{R}^{N'\times N'}$. Due to the separable assumption of auto-correlation in the horizontal and vertical directions, $\mathrm{Var}(\boldsymbol{\xi}')$ can be expressed as the following Kronecker product:

$$\mathrm{Var}(\boldsymbol{\xi}') = \left[\mathbf{R}_{h,\xi'} \otimes \mathbf{R}_{v,\xi'}\right] \otimes \mathbf{C} \tag{8.}$$

where $\mathbf{R}_{h,\xi'}\in\mathbf{R}^{n'h\times n'h}$ and $\mathbf{R}_{v,\xi'}\in\mathbf{R}^{n'v\times n'v}$ denote the horizontal and vertical correlation matrices of $\boldsymbol{\xi}'$, respectively. $\mathbf{R}_{h,\xi'}$ and $\mathbf{R}_{v,\xi'}$ can be estimated by Eq. (6) and Eq. (7) by replacing (x_i, y_i) with (x'_i, y'_i), where (x'_i, y'_i) are the x- and y-coordinates for the i-th column of the RFEs (i = 1, 2, …, n'_h), and z'_j is the z-coordinate for the j-th element at each column (j = 1, 2, …, n'_v).

2.4. Covariance between soundings and RFEs

Let us denote the covariance matrix between $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ by $\mathrm{Cov}(\boldsymbol{\xi}',\boldsymbol{\xi})\in\mathbf{R}^{N'\times N}$. If the auto-correlation is separable in the horizontal and vertical directions, $\mathrm{Cov}(\boldsymbol{\xi}',\boldsymbol{\xi})$ can be expressed as the following Kronecker product:

$$\mathrm{Cov}(\boldsymbol{\xi}',\boldsymbol{\xi}) = \left[\mathbf{R}_{h,\xi'\xi} \otimes \mathbf{R}_{v,\xi'\xi}\right] \otimes \mathbf{C} \tag{9.}$$

where $\mathbf{R}_{h,\xi'\xi}\in\mathbf{R}^{n'h\times nh}$ and $\mathbf{R}_{v,\xi'\xi}\in\mathbf{R}^{n'v\times nv}$ denote the horizontal and vertical correlation matrices between $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$, respectively:

$$\mathbf{R}_{\mathrm{h},\xi'\xi}=\begin{bmatrix}\rho_{\mathrm{h}}\left(x'_1-x_1,y'_1-y_1\right) & \rho_{\mathrm{h}}\left(x'_1-x_2,y'_1-y_2\right) & \cdots & \rho_{\mathrm{h}}\left(x'_1-x_{n_{\mathrm{h}}},y'_1-y_{n_{\mathrm{h}}}\right)\\ \rho_{\mathrm{h}}\left(x'_2-x_1,y'_2-y_1\right) & \rho_{\mathrm{h}}\left(x'_2-x_2,y'_2-y_2\right) & \cdots & \rho_{\mathrm{h}}\left(x'_2-x_{n_{\mathrm{h}}},y'_2-y_{n_{\mathrm{h}}}\right)\\ & & \ddots & \vdots\\ \rho_{\mathrm{h}}\left(x'_{n'_{\mathrm{h}}}-x_1,y'_{n'_{\mathrm{h}}}-y_1\right) & \rho_{\mathrm{h}}\left(x'_{n'_{\mathrm{h}}}-x_2,y'_{n'_{\mathrm{h}}}-y_2\right) & \cdots & \rho_{\mathrm{h}}\left(x'_{n'_{\mathrm{h}}}-x_{n_{\mathrm{h}}},y'_{n'_{\mathrm{h}}}-y_{n_{\mathrm{h}}}\right)\end{bmatrix} \tag{10.}$$

$$\mathbf{R}_{\mathrm{v},\xi'\xi}=\begin{bmatrix}\rho_{\mathrm{v}}\left(z'_1-z_1\right) & \rho_{\mathrm{v}}\left(z'_1-z_2\right) & \cdots & \rho_{\mathrm{v}}\left(z'_1-z_{n_{\mathrm{v}}}\right)\\ \rho_{\mathrm{v}}\left(z'_2-z_1\right) & \rho_{\mathrm{v}}\left(z'_2-z_2\right) & \cdots & \rho_{\mathrm{v}}\left(z'_2-z_{n_{\mathrm{v}}}\right)\\ & & \ddots & \vdots\\ \rho_{\mathrm{v}}\left(z'_{n'_{\mathrm{v}}}-z_1\right) & \rho_{\mathrm{v}}\left(z'_{n'_{\mathrm{v}}}-z_2\right) & \cdots & \rho_{\mathrm{v}}\left(z'_{n'_{\mathrm{v}}}-z_{n_{\mathrm{v}}}\right)\end{bmatrix} \tag{11.}$$

2.5. Brute-force method

Simulating the multivariate cross-correlated 3D conditional random field is equivalent to sample ξ' conditioning on the observed sounding data ξ^{o}. Suppose that the mean vector of the m number of variables contained in ξ is $\underline{\mu}_s = [\mu_1\ \mu_2\ \ldots\ \mu_m]^T$. Let $\underline{\mu}$ denote the mean vector of ξ, including the mean values $\underline{\mu}^{o}$ for ξ^{o} and unconditional mean values $\underline{\mu}^{u}$ for ξ^{u}. $\underline{\mu}$ is formed by stacking n number of $\underline{\mu}_s$ vertically into a new column vector. Let $\underline{\mu}'$ denote the unconditional mean vector of ξ' by stacking n′ number of $\underline{\mu}_s$ vertically. It is noted that the ξ^{o} and ξ' follows a multivariate normal distribution. According to the multivariate normal distribution theory, the conditional probability density function (PDF) $f(\xi'|\xi^{o})$ still follows a multivariate normal distribution with the following mean vector and covariance matrix:

$$\begin{aligned}&\mathrm{E}\left(\xi'\mid\xi^{o}\right)=\mathrm{E}\left(\xi'\right)+\mathrm{Cov}\left(\xi',\xi^{o}\right)\mathrm{Var}\left(\xi^{o}\right)^{-1}\left[\xi^{o}-\mathrm{E}\left(\xi^{o}\right)\right]\\&=\underline{\mu}'+\mathrm{Cov}\left(\xi',\xi^{o}\right)\mathrm{Var}\left(\xi^{o}\right)^{-1}\left[\xi^{o}-\underline{\mu}^{o}\right]\end{aligned} \tag{12.}$$

$$\mathrm{Var}\left(\xi'\mid\xi^{o}\right)=\mathrm{Var}\left(\xi'\right)-\mathrm{Cov}\left(\xi',\xi^{o}\right)\mathrm{Var}\left(\xi^{o}\right)^{-1}\mathrm{Cov}\left(\xi',\xi^{o}\right)^{\mathrm{T}} \tag{13.}$$

where $\mathrm{Cov}(\xi', \xi^{o}) \in \mathbf{R}^{N'\times no}$ denotes the covariance matrix between ξ' and ξ^{o}; $\mathrm{Var}(\xi') \in \mathbf{R}^{N'\times N'}$ is the covariance matrix of ξ'; $\mathrm{Var}(\xi^{o}) \in \mathbf{R}^{no\times no}$ is the covariance matrices of ξ^{o}. The brute-force method can simulate the conditional sample of ξ' directly using the Cholesky decomposition method when $E(\xi'|\xi^{o})$ and $\mathrm{Var}(\xi'|\xi^{o})$ are computed. To be specific, $\xi' = \underline{\mu}' + \mathbf{LZ}$ where $\mathbf{LL}^{T} = \mathrm{Var}(\xi'|\xi^{o})$ and $\mathbf{Z}$ is an

independent standard normal vector with size of N′. The simulation of $\boldsymbol{\xi}'$ require the Cholesky decomposition of Var($\boldsymbol{\xi}'|\boldsymbol{\xi}^o$), which can be an extremely large matrix. The size of this extremely large matrix may be larger than the storage capacity of the current desk computer. As a result, the brute-force method may become infeasible to simulate $\boldsymbol{\xi}'$ (To be seen in subsection 5.4 “Comparison of computational efficiency among different methods” in the illustrative example).

3. Review of the Yang and Ching (2020)’s method

Yang and Ching (2020) proposed a novel method to efficiently simulate $\boldsymbol{\xi}'$ base on the assumption of separability between horizontal and vertical auto-correlations. The method mainly contains two steps:

1. Simulate $\boldsymbol{\xi}^u$ by conditioning on $\boldsymbol{\xi}^o$. Two novel sampling techniques are proposed for implementing this step by Yang and Ching (2020): the sounding-wise Gibbs sampler (GS) and the depth-wise Monte Carlo simulation (MCS). MCS is a numerical process of repeatedly drawing samples from prescribed probability distributions while GS generates a sample from the distribution of each variable in turn, conditional on the current values of the other variables. The adjacent samples are independent in MCS while they are not in GS.
2. Simulates $\boldsymbol{\xi}'$ based on the make-complete $\boldsymbol{\xi}$ (the union of $\boldsymbol{\xi}^o$ and the simulated $\boldsymbol{\xi}^u$ at the last step). The $\boldsymbol{\xi}'$ can be efficiently simulated with the following formula (Yang and Ching 2020; Ching et al. 2020):

$$\boldsymbol{\xi}' = \mathrm{vec}\left[\mathbf{R}_{\mathrm{v},\xi'\xi}\cdot\mathbf{R}_{\mathrm{v},\xi}^{-1}\cdot\mathrm{mat}\left(\boldsymbol{\xi}\right)\cdot\mathbf{R}_{\mathrm{h},\xi}^{-1}\cdot\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]+\sigma\cdot\mathrm{vec}\left[\mathbf{L}_{\mathrm{v},\xi'}\cdot\mathrm{mat}\left(\mathbf{Z}\right)\cdot\mathbf{L}_{\mathrm{h},\xi'}^{\mathrm{T}}\right] \quad (14.)$$

where mat($\mathbf{Z}$) is an ($n'_v \times n'_h$) matrix containing independent standard normal random variables; mat($\boldsymbol{\xi}$) = [$\underline{\xi}_1\ \underline{\xi}_2 \ldots \underline{\xi}_{nh}$] is an ($n_v \times n_h$) matrix containing all sounding data; vec[$\mathbf{A}$] is a vector formed by stacking the columns in $\mathbf{A}$ vertically; $\mathbf{L}_{h,\xi'}$ and $\mathbf{L}_{v,\xi'}$ are the lower triangular Cholesky decompositions of ($\mathbf{R}_{h,\xi'} - \mathbf{R}_{h,\xi'\xi}\mathbf{R}_{h,\xi}^{-1}\mathbf{R}_{h,\xi'\xi}^{T}$) and $\mathbf{R}_{v,\xi'}$, respectively.

This method works quite well for a univariate 3D conditional random field, i.e., $\boldsymbol{\xi}'$ only contains one soil property parameter (m = 1). However, it cannot be directly applied to multivariate cross-

correlated 3D conditional random field, i.e., $\boldsymbol{\xi}'$ contains multiple soil property parameters (m > 1). The extended method is to accommodate the multivariate non-lattice data by accounting for the cross-correlation among different variables at Step 1 and Step 2. The extended method is elaborated in the following section.

4. Method

4.1. Step 1 – simulating $\boldsymbol{\xi}^u$ conditioning on $\boldsymbol{\xi}^o$

Recall that $\boldsymbol{\xi}^u$ and $\boldsymbol{\xi}^o$ follows a multivariate normal distribution. According to the multivariate normal distribution theory, the conditional PDF $f(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ is still a multivariate normal distribution with the following mean vector $E(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ and covariance matrix $Var(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$. The $E(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ and $Var(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ can be computed by Eqs. (12) and (13) by replacing $\underline{\mu}'$ and $\boldsymbol{\xi}'$ with $\underline{\mu}^u$ and $\boldsymbol{\xi}^u$, where $\underline{\mu}^u$ is the mean vector of $\boldsymbol{\xi}^u$ with size of $n_u \times 1$. Once $E(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ and $Var(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ are computed, the conditional sample of $\boldsymbol{\xi}^u$ can be readily simulated using the Cholesky decomposition method. However, both $Var(\boldsymbol{\xi}^o)$ and $Var(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ can be large matrices for a 3D problem especially when there are many explored soundings. The calculation of both $Var(\boldsymbol{\xi}^o)^{-1}$ and $Var(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$ can be computationally expensive and may be prone to numerical errors. The sounding-wise GS and depth-wise MCS is applied to avoiding the computational issues associated with large matrices in simulating the sounding data $\boldsymbol{\xi}^u$.

4.1.1 Sounding-wise GS

The sounding-wise GS simulates the un-observed data at each sounding sequentially. Recall that the sounding data $\boldsymbol{\xi} = [\underline{\xi}_1^T \underline{\xi}_2^T \dots \underline{\xi}_{nh}^T]^T$ where $\underline{\xi}_i$ denotes the complete data for the i-th sounding. $\boldsymbol{\xi}$ is the union of observed sounding data $\boldsymbol{\xi}^o$ and un-observed sounding data $\boldsymbol{\xi}^u$. Let $\underline{\xi}_i^o$ and $\underline{\xi}_i^u$ denote the observed and unobserved parts in $\underline{\xi}_i$, respectively. Let $\boldsymbol{\xi}_{\sim i} = [\underline{\xi}_1^T \dots \underline{\xi}_{i-1}^T \underline{\xi}_{i+1}^T \dots \underline{\xi}_{nh}^T]^T$ contains all data except the i-th sounding.

4.1.1.1. Implementation procedures

The sounding wise GS starts with assigning $\boldsymbol{\xi}^u$ with any values (e.g., mean values) so that all open markers in Figure 1 are filled with these assigned values (i.e., $\boldsymbol{\xi}$ becomes complete with a lattice

304 structure). Its implementation procedure is summarized as follows:

305 (1) (Simulate the 1st $\boldsymbol{\xi}^u$ sample) Do the sounding-wise sampling successively for soundings #1,
306 #2, …, and #nh:

307 a. Draw $\underline{\xi}_1^u$ from $f(\underline{\xi}_1^u|\underline{\xi}_1^o, \boldsymbol{\xi}_{\sim 1})$ and update the $\underline{\xi}_1^u$ values in $\boldsymbol{\xi}$ with the sampled values (i.e.,
308 these $\underline{\xi}_1^u$ values are used for simulating the $\underline{\xi}_2^u$ at the subsequent step).

309 b. Draw $\underline{\xi}_2^u, \cdots, \underline{\xi}_{nh}^u$, in a way similar to that of $\underline{\xi}_1^u$. In other words, draw $\underline{\xi}_i^u$ from $f(\underline{\xi}_i^u|\underline{\xi}_i^o, \boldsymbol{\xi}_{\sim i})$
310 and update the $\underline{\xi}_i^u$ values in $\boldsymbol{\xi}$ with the sampled values (i.e., these $\underline{\xi}_i^u$ values are used for
311 simulating the $\underline{\xi}_{i+1}^u$).

312 c. Combine the samples $(\underline{\xi}_1^u, \underline{\xi}_2^u, \ldots, \underline{\xi}_{nh}^u)$ to form the 1st sample of $\boldsymbol{\xi}^u$. This sample is used
313 for simulating the next $\boldsymbol{\xi}^u$ sample.

314 (2) Repeat step (1) for t = 2, …, T to obtain another (T-1) samples of $\boldsymbol{\xi}^u$. The initial $\boldsymbol{\xi}^u$ samples should
315 be discarded because they cannot represent the target PDF $f(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$. The nearby samples are
316 correlated, so the $\boldsymbol{\xi}^u$ samples should be re-sampled by a suitable interval to reduce this correlation.
317 Suppose that there are N_{GS} samples of $\boldsymbol{\xi}^u$ (i = 1, 2, …, N_{GS}) after discarding the preliminary
318 samples and re-sampling.

319 *4.1.1.2. Simulating $\underline{\xi}_k^u$ from $f(\underline{\xi}_k^u|\underline{\xi}_k^o, \boldsymbol{\xi}_{\sim k})$*

320 The simulation of $\underline{\xi}_k^u \sim f(\underline{\xi}_k^u|\underline{\xi}_k^o, \boldsymbol{\xi}_{\sim k})$ is a key step of the sounding-wise GS. To avoid the large
321 matrix issues, the sounding-wise GS decomposes this key step into two sub-steps: (a) compute the
322 $f(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{\sim k})$ and (b) compute the $f(\underline{\xi}_k^u|\underline{\xi}_k^o, \boldsymbol{\xi}_{\sim k})$. Figure 2 shows the schematic diagram of the two sub-
323 steps. Because $\boldsymbol{\xi}$ is multivariate normal, $f(\boldsymbol{\xi}_k|\boldsymbol{\xi}_{\sim k})$ is a multivariate normal PDF with the following
324 mean vector and covariance matrix:

325 $$E\left(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k}\right) = \underline{\mu}_k + \mathrm{Cov}\left(\underline{\xi}_k, \boldsymbol{\xi}_{\sim k}\right)\mathrm{Var}\left(\boldsymbol{\xi}_{\sim k}\right)^{-1}\left(\boldsymbol{\xi}_{\sim k} - \underline{\mu}_{\sim k}\right) \quad (15.)$$

326 $$\mathrm{Var}\left(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k}\right) = \mathrm{Var}\left(\underline{\xi}_k\right) - \mathrm{Cov}(\underline{\xi}_k, \boldsymbol{\xi}_{\sim k})\mathrm{Var}(\boldsymbol{\xi}_{\sim k})^{-1}\mathrm{Cov}(\underline{\xi}_k, \boldsymbol{\xi}_{\sim k})^T \quad (16.)$$

327 where $\underline{\mu}_k$ and $\underline{\mu}_{\sim k}$ are the unconditional mean vectors for $\underline{\xi}_k$ and $\boldsymbol{\xi}_{\sim k}$. They are the sub-vectors in $\underline{\mu}$.

Because $\boldsymbol{\xi}_{\sim k}$ is filled with assigning/simulated values (the first sample) or simulated values (after the first sample), $\boldsymbol{\xi}_{\sim k}$ satisfies the 3D lattice structure. Based on the separable auto-correlation function, the covariance matrices for $\underline{\xi}_k$ and $\boldsymbol{\xi}_{\sim k}$ allow the following Kronecker product representations:

$$\mathrm{Var}\left(\boldsymbol{\xi}_{\sim k}\right)=\left(\mathbf{R}_{h,\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)\qquad \mathrm{Var}\left(\underline{\xi}_k\right)=\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\qquad \mathrm{Cov}(\underline{\xi}_k,\boldsymbol{\xi}_{\sim k})=\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right) \tag{17.}$$

where $\mathbf{R}_{h,\boldsymbol{\xi}_{\sim k}}\in\mathbf{R}^{(nh-1)\times(nh-1)}$ denotes the horizontal auto-correlation matrix for $\boldsymbol{\xi}_{\sim k}$ and is the sub-matrix of $\mathbf{R}_{h,\boldsymbol{\xi}}$ by excluding the k-th row and k-th column; $\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\in\mathbf{R}^{1\times(nh-1)}$ denotes the horizontal auto-correlation vector between $\underline{\xi}_k$ and $\boldsymbol{\xi}_{\sim k}$

$$\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}=\left[\rho_h\left(x_k-x_1,y_k-y_1\right)\ \cdots\ \rho_h\left(x_k-x_{k-1},y_k-y_{k-1}\right)\ \rho_h\left(x_k-x_{k+1},y_k-y_{k+1}\right)\ \cdots\ \rho_h\left(x_k-x_{n_h},y_k-y_{n_h}\right)\right] \tag{18.}$$

In Eq. (17), the vertical auto-correlation matrix for $\underline{\xi}_k$ and $\boldsymbol{\xi}_{\sim k}$ is the same (i.e., $\mathbf{R}_{v,\boldsymbol{\xi}}\in\mathbf{R}^{nv\times nv}$) because $\underline{\xi}_k$ and $\boldsymbol{\xi}_{\sim k}$ are with same sounding depths ($z_1,z_2,..,z_{nv}$). Considering the following matrix identities $(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$, $(\mathbf{A}\otimes\mathbf{B})^{T}=\mathbf{A}^{T}\otimes\mathbf{B}^{T}$, and $(\mathbf{A}^{T}\otimes\mathbf{B})\times\mathrm{vec}(\mathbf{C})=\mathrm{vec}(\mathbf{BCA})$ and inserting Eq. (17) into Eqs. (15) and (16) yields:

$$\begin{aligned}
&\mathrm{E}\left(\underline{\xi}_k\left|\boldsymbol{\xi}_{\sim k}\right.\right)=\underline{\mu}_k+\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)\cdot\left(\mathbf{R}_{h,\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)^{-1}\cdot\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\\
&=\underline{\mu}_k+\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)\cdot\left(\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}^{-1}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}^{-1}\right)\cdot\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\\
&=\left[\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\cdot\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\right)\otimes\left(\mathbf{R}_{v,\boldsymbol{\xi}}\cdot\mathbf{R}^{-1}_{v,\boldsymbol{\xi}}\right)\otimes\left(\mathbf{C}\cdot\mathbf{C}^{-1}\right)\right]\cdot\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\\
&=\underline{\mu}_k+\left[\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\cdot\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\right)\otimes\mathbf{I}_v\otimes\mathbf{I}_s\right]\cdot\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\\
&=\underline{\mu}_k+\mathrm{vec}\left[\left(\mathbf{I}_v\otimes\mathbf{I}_s\right)\cdot\mathrm{mat}\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\cdot\left(\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\underline{\mathrm{R}}^{T}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\right)\right]\\
&=\underline{\mu}_k+\mathrm{vec}\left[\mathrm{mat}\left(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k}\right)\cdot\left(\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\underline{\mathrm{R}}^{T}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\right)\right]
\end{aligned} \tag{19.}$$

$$\begin{aligned}
&\mathrm{Var}\left(\underline{\xi}_k\left|\boldsymbol{\xi}_{\sim k}\right.\right)=\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}-\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)\cdot\left(\mathbf{R}_{h,\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)^{-1}\cdot\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}\right)^{T}\\
&=\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}-\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\xi}\otimes\mathbf{C}\right)\cdot\left(\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}^{-1}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}^{-1}\right)\cdot\left(\underline{\mathrm{R}}^{T}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\otimes\mathbf{R}_{v,\xi}\otimes\mathbf{C}\ \right)\\
&=\mathbf{R}_{v,\boldsymbol{\xi}}\otimes\mathbf{C}-\left(\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\mathbf{R}^{-1}_{h,\boldsymbol{\xi}_{\sim k}}\underline{\mathrm{R}}^{T}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\right)\otimes\mathbf{R}_{v,\xi}\otimes\mathbf{C}
\end{aligned} \tag{20.}$$

where $\mathbf{I}_v$ denotes the identity matrix with size of $n_v\times n_v$; $\mathbf{I}_s$ denote the identity matrix with size of $m\times m$; $\mathrm{mat}(\boldsymbol{\xi}_{\sim k}-\underline{\mu}_{\sim k})=[\underline{\xi}_1-\underline{\mu}_1\ldots\underline{\xi}_{k-1}-\underline{\mu}_{k-1}\ \underline{\xi}_{k+1}-\underline{\mu}_{k+1}\ \ldots\underline{\xi}_{nh}-\underline{\mu}_{nh}]$, an $(m\times n_v)\times(n_h-1)$ matrix. Note that the term $\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}\mathbf{R}_{h,\boldsymbol{\xi}_{\sim k}}{}^{-1}\underline{\mathrm{R}}_{h,\underline{\xi}_k\boldsymbol{\xi}_{\sim k}}{}^{T}$ in Eq. (20) is a scaler, hence

$$\mathrm{Var}\left(\underline{\xi}_k \left| \boldsymbol{\xi}_{\sim k} \right.\right) = \left(1 - \underline{\mathrm{R}}_{\mathrm{h},\underline{\xi}_k \boldsymbol{\xi}_{\sim k}} \mathbf{R}^{-1}_{\mathrm{h},\boldsymbol{\xi}_{\sim k}} \underline{\mathrm{R}}^{\mathrm{T}}_{\mathrm{h},\underline{\xi}_k \boldsymbol{\xi}_{\sim k}}\right)\left(\mathbf{R}_{\mathrm{v},\boldsymbol{\xi}} \otimes \mathbf{C}\right) \tag{21.}$$

Because the $\underline{\xi}_k$ follows a multivariate normal distribution with the mean vector of $\mathrm{E}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$ and covariance of $\mathrm{Var}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$, the PDF $f(\underline{\xi}_k^{\mathrm{u}}|\underline{\xi}_k^{\mathrm{o}}, \boldsymbol{\xi}_{\sim k})$ is also a multivariate normal PDF with the following mean vector and covariance matrix

$$\mathrm{E}\left(\underline{\xi}_k^{\mathrm{u}} \left| \underline{\xi}_k^{\mathrm{o}}, \boldsymbol{\xi}_{\sim k} \right.\right) = \mathrm{E}\left(\underline{\xi}_k^{\mathrm{u}} \left| \boldsymbol{\xi}_{\sim k} \right.\right) + \mathrm{Cov}\left(\underline{\xi}_k^{\mathrm{u}}, \underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right) \mathrm{Var}\left(\underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right)^{-1} \left[\underline{\xi}_k^{\mathrm{o}} - \mathrm{E}\left(\underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right)\right] \tag{22.}$$

$$\mathrm{Var}\left(\underline{\xi}_k^{\mathrm{u}} \left| \underline{\xi}_k^{\mathrm{o}}, \boldsymbol{\xi}_{\sim k} \right.\right) = \mathrm{Var}\left(\underline{\xi}_k^{\mathrm{u}} \left| \boldsymbol{\xi}_{\sim k} \right.\right) - \mathrm{Cov}\left(\underline{\xi}_k^{\mathrm{u}}, \underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right) \mathrm{Var}\left(\underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right)^{-1} \mathrm{Cov}\left(\underline{\xi}_k^{\mathrm{u}}, \underline{\xi}_k^{\mathrm{o}} \left| \boldsymbol{\xi}_{\sim k} \right.\right)^{\mathrm{T}} \tag{23.}$$

where $\mathrm{E}(\underline{\xi}_k^{\mathrm{u}}|\boldsymbol{\xi}_{\sim k})$ and $\mathrm{E}(\underline{\xi}_k^{\mathrm{u}}|\boldsymbol{\xi}_{\sim k})$ are sub-vectors of $\mathrm{E}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$, and $\mathrm{Var}(\underline{\xi}_k^{\mathrm{o}}|\boldsymbol{\xi}_{\sim k})$, $\mathrm{Var}(\underline{\xi}_k^{\mathrm{o}}|\boldsymbol{\xi}_{\sim k})$, and $\mathrm{Cov}(\underline{\xi}_k^{\mathrm{u}}, \underline{\xi}_k^{\mathrm{o}}|\boldsymbol{\xi}_{\sim k})$ are sub-matrices of $\mathrm{Var}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$. These sub-vectors and sub-matrices can be extracted from $\mathrm{E}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$ and $\mathrm{Var}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$ directly. The evaluation of $\mathrm{E}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$ and $\mathrm{Var}(\underline{\xi}_k|\boldsymbol{\xi}_{\sim k})$ in Eqs. (19) and (21) does not require inversion and storage of larger matrices. The simulation of $\underline{\xi}_k^{\mathrm{u}}$ sample with Eqs. (22) and (23) also does not need to make Cholesky decomposition of larger matrices.

4.1.2 Depth-wise MCS

Let $\underline{\xi}_{zi} \in \mathbf{R}^{\mathrm{nh}\times 1}$ contain all sounding data at the z_i depth. It is evident that $\underline{\xi}_{zi}$ is a $\mathrm{m}\times\mathrm{n_h}$ column vector since the data at each depth at each sounding contain m number of variables. Let $\underline{\xi}_{zi}^{\mathrm{o}}$ and $\underline{\xi}_{zi}^{\mathrm{u}}$ denote the observed and unobserved parts, respectively, in $\underline{\xi}_{zi}$. The depth-wise MCS considers that $\underline{\xi}_{zi}$ depends on $\underline{\xi}_z$ only if z is within the depth range of $[z_i-\alpha\delta_v, z_i+\alpha\delta_v]$, where δ_v denotes the vertical SOF and α denote the coefficient that defines the depth range. It is based on the fact that soil properties at a certain depth are mainly correlated with the soil properties at depths with a distance less than a few times of vertical SOFs. Therefore, the α should take a value to ensure that the data points with a distance equal to or larger than $\alpha\delta_v$ are generally independent. A larger value of α might reduce the computational efficiency of depth-wise MCS (discussed later in subsection 5.4).

Figure 3 shows the schematic diagram of the depth-wise MCS. The implementation procedure of the depth-wise MCS is summarized as follows (Yang and Ching 2020):

(1) (Simulate the 1st $\boldsymbol{\xi}^{\mathrm{u}}$ sample) Do the depth-wise sampling for depths z_1, z_2, …, and z_{nv}:

a. Simulate $\underline{\xi}_{z1}^{u} \sim f(\underline{\xi}_{z1}^{u}|\underline{\xi}_{z1}^{o}, \ldots, \underline{\xi}_{zmax}^{o})$, where z_{max} denotes the maximum conditioning depth (z_{max} is either $z_i+\alpha\delta_v$ or z_{nv}; the smaller one is taken)

b. Successively simulate $\underline{\xi}_{z2}^{u}, \cdots, \underline{\xi}_{znv}^{u}$ in a way similar to that of $\underline{\xi}_{z1}^{u}$. In other words, Simulate $\underline{\xi}_{zi}^{u} \sim f(\underline{\xi}_{zi}^{u}|\underline{\xi}_{zmin}^{o}, \ldots, \underline{\xi}_{zmax}^{o}, \underline{\xi}_{zmin}^{u}, \ldots, \underline{\xi}_{zi-1}^{u})$, where z_{min} denotes the minimum conditioning depth (z_{min} is either $z_i-\alpha\delta_v$ or z_1; the larger one is taken). Here, as shown in Figure 3, ($\underline{\xi}_{zmin}^{u} \ldots \underline{\xi}_{zi-1}^{u}$) take the values that were simulated in previous steps.

c. Combine the samples ($\underline{\xi}_{z1}^{u}, \underline{\xi}_{z2}^{u}, \ldots, \underline{\xi}_{znv}^{u}$) to form a sample of $\boldsymbol{\xi}^{u}$. This sample is NOT used for simulating the next $\boldsymbol{\xi}^{u}$ sample.

(2) Repeat step (1) for t = 2, …, N_{MCS} to obtain N_{MCS} samples of $\boldsymbol{\xi}^{u}$. Note that the simulating order for depths in step (1) is not important, it just provides a simple order that is easy to code.

Because $\boldsymbol{\xi}$ are normal variables, $f(\underline{\xi}_{z1}^{u}|\underline{\xi}_{z1}^{o}, \ldots, \underline{\xi}_{zmax}^{o})$, $f(\underline{\xi}_{zi}^{u}|\underline{\xi}_{zmin}^{o}, \ldots, \underline{\xi}_{zmax}^{o}, \underline{\xi}_{zmin}^{u}, \ldots, \underline{\xi}_{zi-1}^{u})$ and $f(\underline{\xi}_{znv}^{u}|\underline{\xi}_{zmin}^{o}, \ldots, \underline{\xi}_{znv}^{o}, \underline{\xi}_{zmin}^{u}, \ldots, \underline{\xi}_{znv-1}^{u})$ also follow multivariate normal distributions. Their mean vectors and covariance matrices can be estimated using Eqs (12) and (13).

4.2. Step 2 – simulating ξ' conditioning on ξ

It is noted that $\boldsymbol{\xi}$ is make-complete and containing no missing data after Step 1. The make-complete $\boldsymbol{\xi}$ is then used for simulating $\boldsymbol{\xi}'$. Both $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ are with a lattice structure. As a result, it can take advantages of the Kronecker product to efficiently simulate $\boldsymbol{\xi}$. Because $\boldsymbol{\xi}'$ and $\boldsymbol{\xi}$ are multivariate normal, $\boldsymbol{\xi}'$ conditioning on $\boldsymbol{\xi}$ is still multivariate normal with the following mean vector and covariance matrix:

$$\mathrm{E}\left(\boldsymbol{\xi}' \mid \boldsymbol{\xi}\right) = \boldsymbol{\mu}' + \mathrm{Cov}\left(\boldsymbol{\xi}', \boldsymbol{\xi}\right) \mathrm{Var}\left(\boldsymbol{\xi}\right)^{-1} \left(\boldsymbol{\xi} - \boldsymbol{\mu}\right) \tag{24.}$$

$$\mathrm{Var}\left(\boldsymbol{\xi}' \mid \boldsymbol{\xi}\right) = \mathrm{Var}\left(\boldsymbol{\xi}'\right) - \mathrm{Cov}\left(\boldsymbol{\xi}', \boldsymbol{\xi}\right) \mathrm{Var}\left(\boldsymbol{\xi}\right)^{-1} \mathrm{Cov}\left(\boldsymbol{\xi}', \boldsymbol{\xi}\right)^{\mathrm{T}} \tag{25.}$$

Note that $\mathrm{Cov}(\boldsymbol{\xi}',\boldsymbol{\xi}) = [\mathbf{R}_{h,\xi'\xi} \otimes \mathbf{R}_{v,\xi'\xi} \otimes \mathbf{C}]$, $\mathrm{Var}(\boldsymbol{\xi}) = [\mathbf{R}_{h,\xi} \otimes \mathbf{R}_{v,\xi} \otimes \mathbf{C}]$, and $\mathrm{Var}(\boldsymbol{\xi}') = [\mathbf{R}_{h,\xi'} \otimes \mathbf{R}_{v,\xi'} \otimes \mathbf{C}]$. Considering the matrix identities $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$, $(\mathbf{A}\otimes\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\otimes\mathbf{B}^{\mathrm{T}}$, and $(\mathbf{A}^{\mathrm{T}}\otimes\mathbf{B})\times\mathrm{vec}(\mathbf{C}) = \mathrm{vec}(\mathbf{BCA})$, Eqs. (24) and (25) can be further expressed as

393
$$\begin{aligned}&\mathrm{E}\left(\boldsymbol{\xi}'\mid\boldsymbol{\xi}\right)=\boldsymbol{\mu}'+\left(\mathbf{R}_{\mathrm{h},\xi'\xi}\otimes\mathbf{R}_{\mathrm{v},\xi'\xi}\otimes\mathbf{C}\right)\left(\mathbf{R}_{\mathrm{h},\xi}\otimes\mathbf{R}_{\mathrm{v},\xi}\otimes\mathbf{C}\right)^{-1}\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\\&=\boldsymbol{\mu}'+\left[\left(\mathbf{R}_{\mathrm{h},\xi'\xi}\mathbf{R}_{\mathrm{h},\xi}^{-1}\right)\otimes\left(\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\right)\otimes\mathbf{I}_{\mathrm{s}}\right]\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)=\mathrm{vec}\left[\left(\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\otimes\mathbf{I}_{\mathrm{s}}\right)\mathrm{mat}\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]\end{aligned}\tag{26.}$$

394
$$\begin{aligned}&\mathrm{Var}\left(\boldsymbol{\xi}'\mid\boldsymbol{\xi}\right)=\left[\mathbf{R}_{\mathrm{h},\xi'}\otimes\mathbf{R}_{\mathrm{v},\xi'}\otimes\mathbf{C}-\left(\mathbf{R}_{\mathrm{h},\xi'\xi}\otimes\mathbf{R}_{\mathrm{h},\xi'\xi}\otimes\mathbf{C}\right)\left(\mathbf{R}_{\mathrm{h},\xi}\otimes\mathbf{R}_{\mathrm{v},\xi}\otimes\mathbf{C}\right)^{-1}\left(\mathbf{R}_{\mathrm{h},\xi'\xi}\otimes\mathbf{R}_{\mathrm{v},\xi'\xi}\otimes\mathbf{C}\right)^{\mathrm{T}}\right]\\&=\left[\mathbf{R}_{\mathrm{h},\xi'}\otimes\mathbf{R}_{\mathrm{v},\xi'}\otimes\mathbf{C}-\left(\left[\mathbf{R}_{\mathrm{h},\xi'\xi}\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]\otimes\left[\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\mathbf{R}_{\mathrm{v},\xi'\xi}^{\mathrm{T}}\right]\right)\otimes\mathbf{C}\right]\end{aligned}\tag{27.}$$

395 where mat($\boldsymbol{\xi}$) = [ξ_1 ξ_2 …ξ_{nh}], an (m×n_v)×n_h matrix; vec[**A**] is a vector formed by stacking the columns
396 in **A** vertically; $\mathbf{I}_s$ is a m×m identity matrix. Ching et al. (2020) found that if the "densely covers"
397 conditions holds, it has $\mathbf{R}_{v,\xi'} \approx \mathbf{R}_{v,\xi'\xi}\mathbf{R}_{v,\xi}^{-1}\mathbf{R}_{v,\xi'\xi}^{T}$. The "densely covers" condition means that the
398 sampling intervals for the sounding data (z_1,z_2,..,z_{nv}) should significantly smaller than the vertical
399 SOF and the sounding depth range [z_1, z_{nv}] should cover the random field element depth range [z'_1,
400 $z'_{n'v}$]. Yang and Ching (2020) further pointed out that the "densely covers" conditions can be satisfied
401 by inserting additional un-observed depths into the sample intervals for the sounding data. The
402 sounding data at the additional un-observed depths can be simulated at Step 1 so that the sounding
403 data is make-complete and satisfies the "densely covers" conditions. When the make-complete
404 sounding data $\boldsymbol{\xi}$ are used for simulating $\boldsymbol{\xi}'$, Eq. (27) reduces to the following equation

405
$$\mathrm{Var}\left(\boldsymbol{\xi}'\mid\boldsymbol{\xi}\right)=\left(\mathbf{R}_{\mathrm{h},\xi'}-\mathbf{R}_{\mathrm{h},\xi'\xi}\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right)\otimes\mathbf{R}_{\mathrm{v},\xi'}\otimes\mathbf{C}\tag{28.}$$

406 Let $\mathbf{L}_{h,\xi'}$, $\mathbf{L}_{v,\xi'}$ and $\mathbf{L}_c$ be the lower triangular Cholesky decompositions of [$\mathbf{R}_{h,\xi'} - \mathbf{R}_{h,\xi'\xi}\mathbf{R}_{h,\xi}^{-1}\mathbf{R}_{h,\xi'\xi}^{T}$]
407 and $\mathbf{R}_{v,\xi'}$, respectively. The lower triangular Cholesky decomposition of [$\mathbf{R}_{h,\xi'} - \mathbf{R}_{h,\xi'\xi}\mathbf{R}_{h,\xi}^{-1}\mathbf{R}_{h,\xi'\xi}^{T}$] ⊗
408 $\mathbf{R}_{v,\xi'}$ ⊗ **C** is equal to $\mathbf{L}_{h,\xi'}$ ⊗ $\mathbf{L}_{v,\xi'}$ ⊗ $\mathbf{L}_c$. With the Eqs. (26) and (28), a conditional sample of $\boldsymbol{\xi}'$ can
409 then be simulated by the following equation:

410
$$\begin{aligned}&\boldsymbol{\xi}'=\mathrm{vec}\left[\left(\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\otimes\mathbf{I}_{\mathrm{s}}\right)\mathrm{mat}\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]+\left(\mathbf{L}_{\mathrm{h},\xi'}\otimes\mathbf{L}_{\mathrm{v},\xi'}\otimes\mathbf{L}_{\mathrm{c}}\right)\mathbf{Z}\\&=\mathrm{vec}\left[\left(\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\otimes\mathbf{I}_{\mathrm{s}}\right)\mathrm{mat}\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]+\left(\mathbf{L}_{\mathrm{h},\xi'}\otimes\mathbf{L}_{\mathrm{v},\xi'}\otimes\mathbf{L}_{\mathrm{c}}\right)\mathrm{vec}\left(\mathbf{Z}\right)\\&=\mathrm{vec}\left[\left(\mathbf{R}_{\mathrm{v},\xi'\xi}\mathbf{R}_{\mathrm{v},\xi}^{-1}\otimes\mathbf{I}_{\mathrm{s}}\right)\mathrm{mat}\left(\boldsymbol{\xi}-\boldsymbol{\mu}\right)\mathbf{R}_{\mathrm{h},\xi}^{-1}\mathbf{R}_{\mathrm{h},\xi'\xi}^{\mathrm{T}}\right]+\mathrm{vec}\left[\left(\mathbf{L}_{\mathrm{v},\xi'}\otimes\mathbf{L}_{\mathrm{c}}\right)\mathbf{Z}\mathbf{L}_{\mathrm{h},\xi'}^{\mathrm{T}}\right]\end{aligned}\tag{29.}$$

411 where mat(**Z**) is an (m×n′$_v$)×n′$_h$ matrix containing independent standard normal variables.

412 The proposed method is a little more complicated in theory compared with the traditional

methods. However, it can be programmed as a black box for practical application. As a result, the users do not need to know the theory of the method can use it with simple operation. The program is available from the first author upon reasonable request.

5. Simulated example

To illustrate the proposed method, a simulated virtual clay site is considered. The size of the virtual site is 50m × 50m × 15m in the x, y and z directions. Suppose that the virtual site is discretized into a 3D mesh of size 1m × 1m × 0.1m, respectively. As a result, there are 51 × 51 = 2601 RFEs in the x-y plane (n'_h = 2601) and 151 elements in the z-direction (n'_v = 151). In total, there are N' = 2601 × 151 = 392,751 RFEs. Suppose that the simulated virtual site merely considers three CPTu parameters (m = 3), including pore pressure ratio $B_q = (u_2-u_0)/(q_t-\sigma_v)$, normalized cone resistance $Q_t = (q_t-\sigma_v)/\sigma'_v$, and normalized effective cone resistance $Q_e = (q_t-u_2)/\sigma'_v$, where u_2 is pore pressure behind the cone; u_0 is static pore pressure; q_t is corrected cone resistance; σ_v is vertical total stress; and σ'_v is vertical effective stress. As a result, ξ' containing the values at the centroids of the 392,751 RFEs is a column vector with 392,751 × 3 = 1,178,253 number of elements.

5.1. Generation of the virtual site and sounding data

The virtual site is modelled by trivariate random fields of B_q, Q_t and Q_e. Li et al. (2019a) proposed a CMDC method to simulate trivariate unconditional random fields of B_q, Q_t and Q_e. This method is adopted to generate the virtual site. During the generation of the virtual site, the distributions of B_q, $\ln(Q_t)$ and $\ln(Q_e)$ are assumed to be unbound Johnson distributions, which follows the multivariate statistical results of a global clay database in Ching and Phoon (2014). Let $Y_1 = B_q$, $Y_2 = \ln(Q_t)$ and $Y_3 = \ln(Q_e)$ denotes the Johnson random variables. Let X_1, X_2 and X_3 denote the corresponding standard normal variables of Y_1, Y_2 and Y_3. When a variable Y (i.e., Y_1, Y_2 and Y_3) follows the unbound Johnson distribution, it can be transformed from a standard normal random variable X (i.e., X_1, X_2 and X_3) by

$$Y = \sinh\left[\frac{X - b_X}{a_X}\right] a_Y + b_Y \quad (30.)$$

where $[a_X, b_X, a_Y, b_Y]$ are four Johnson distribution parameters. The Johnson distribution parameters take the values of $[a_X, b_X, a_Y, b_Y]$ = [2.676, 0.161, 0.513, 0.615], [1.340, -0.572, 0.659, 1.476], and [2.134, -1.102, 1.154, 0.657] for Y_1, Y_2 and Y_3, respectively. The spatial variability of Y_1, Y_2 and Y_3 is modeled by first modeling the trivariate standard normal random fields of X_1, X_2 and X_3. The cross-correlation matrix among X_1, X_2 and X_3 is **C** = [1.00, -0.45, -0.63; -0.45, 1.00, 0.74; -0.63, 0.74, 1.00]. The values of $[a_X, b_X, a_Y, b_Y]$ and **C** are calibrated by Ching and Phoon (2014) based on the database CLAY/10/7490. The random fields of Y_1, Y_2 and Y_3 can be simulated by substituting the simulated standard normal random fields of X_1, X_2 and X_3 into the Eq. (30). The standard normal random fields of X_1, X_2 and X_3 share the same auto-correlation model and SOFs. The auto-correlation is modeled by the single exponential model (SExp) with δ_h = 20 m and δ_v = 1 m:

$$\rho_h(\Delta x, \Delta y) = \exp\left(-2\sqrt{(\Delta x)^2 + (\Delta y)^2} \Big/ \delta_h\right) \quad (31.)$$

$$\rho_v(\Delta z) = \exp\left(-2\Delta z / \delta_v\right) \quad (32.)$$

Suppose that there are five explored soundings (n_h = 5) at this virtual site. The locations of the five soundings are denoted as S1-S5 in Figure 4. Suppose that the maximum exploration depth of each sounding is 15 m. A single realization of trivariate unconditional random fields of B_q, Q_t and Q_e is used to generate the observed B_q, Q_t and Q_e data at the sounding depths. Figure 5 shows the generated B_q and $\ln(Q_t)$ data at S1-S5. The $\ln(Q_e)$ data are not shown for simplification. The filled circles and open circles denote the observed and un-observed B_q data, respectively, whereas the filled squares and open squares denote the observed and un-observed $\ln(Q_t)$ data, respectively. The following sampling strategy is adopted for the observed data:

1. The vertical sampling interval of B_q is 1 m at S1-S4. There is no sampling at S5 for B_q. The sampling depth at S1 and S2 is 15 m, whereas it is 10 m at S3 and S4. As a result, there is 2×16 + 2×10 = 52 number of observed B_q data at S1-S5.

2. The vertical sampling interval of $\ln(Q_t)$ is 0.5 m at S1-S5. The sampling depths at S1, S2, S3, S4 and S5 are 15 m, 15 m, 12 m, 4 m and 10 m, respectively. As a result, there is 31+31+24+8+20 = 114 number of observed $\ln(Q_t)$ data at S1-S5.
3. The vertical sampling interval of $\ln(Q_e)$ is 0.2 m at S1, S3, S4 and S5. There is no sampling at S2 for $\ln(Q_e)$. The sampling depths at S1, S3, S4 and S5 are 15 m, 2 m, 7 m, and 15 m, respectively. As a result, there is 76 + 10 + 35+ 76 = 197 number of observed $\ln(Q_e)$ data at S1-S5.

In total, the virtual site contains 52+114+197 = 363 observed B_q, $\ln(Q_t)$ and $\ln(Q_e)$ data. It is noted that the CPTu data (i.e., B_q, $\ln(Q_t)$ and $\ln(Q_t)$) at different soundings might be with different sampling depths but they should share the same sampling interval. Nevertheless, the simulated virtual site adopts various sampling intervals for B_q, Q_t and Q_e at various soundings for better demonstration. The adoption of various sampling intervals for B_q, Q_t and Q_e is to model the possible practical site investigation where other soil data such as cohesion c, friction angle ϕ and elastic modulus E measured at different depth intervals at S1-S5 are also of interest. The observed B_q, $\ln(Q_t)$ and $\ln(Q_e)$ data at the virtual site are then converted to the zero-mean normal random data ξ^o by the inversion of Eq. (30). The purpose of this simulated example is to simulate the data $\boldsymbol{\xi}'$ at the centroids of the 392,751 RFEs ($\boldsymbol{\xi}' \in \mathbf{R}^{1,178,253\times 1}$) by conditioning on the observed sounding data ($\boldsymbol{\xi}^o \in \mathbf{R}^{363\times 1}$). The simulated $\boldsymbol{\xi}'$ values for the RFEs can be converted to B_q, $\ln(Q_t)$ and $\ln(Q_e)$ values by Eq. (30). The un-observed data ($\boldsymbol{\xi}^u$) at the soundings were also simulated (open markers in Figure 5) for the investigated example, they are treated as unknown during the analysis and merely used for verifying the proposed method. 4. It is assumed that the mean, cross-covariance matrix and random field parameters of B_q, $\ln(Q_t)$ and $\ln(Q_e)$ have been properly estimated (i.e., they are assumed to be the real values of the virtual site) by suitable methods such as moment methods (e.g., Uzielli et al. 2005; Cami et al. 2020), maximum likelihood estimation (MLE) (e.g., DeGroot and Baecher 1993; Xiao et al. 2018; Qi and Liu 2019; Cami et al. 2020) or Bayesian methods (e.g., Cao and Wang 2014; Ching et al. 2020; Ching et al.2021).

5.2. Step 1: simulate $\boldsymbol{\xi}^u$ conditioning on $\boldsymbol{\xi}^o$

It is clear that the observed data $\boldsymbol{\xi}^o$ do not satisfy the dense-cover condition: (a) the sampling interval is not always significantly smaller than δ_v (e.g., the sampling interval for B_q at S1-S4 is 1 m, whereas δ_v is also 1 m), and (b) the sounding depth range $[z_1, z_{nv}]$ does not cover the RFE's depth range $[z'_1, z'_{n'v}]$ (e.g., the maximum sounding depth for $\ln(Q_t)$ at S4 is 4 m, far less than the maximum RFE's depth 15 m). It is also clear that the observed data $\boldsymbol{\xi}^o$ do not have a 3D lattice structure: all the five soundings do not share the same observed depths. The proposed method requires $\boldsymbol{\xi}$ to fulfill the above condition. This is achieved by augmenting the observed depths for $\boldsymbol{\xi}^o$ by the un-observed depths for $\boldsymbol{\xi}^u$. One possible augmentation is shown in Figure 5 where the open markers denote the augmented un-observed depths. After the augmentation, the sounding data contain $n_v = 151$ depths [0m, 0.1m, 0.2m, …, 14.9m, 15m] with sounding interval of 0.1 m, which is significantly less than $\delta_v = 1$ m. The depth range for all soundings covers the RFE's depth range [0m, 15m]. Let $\underline{\xi}_i$ denote the augmented sounding data at the i-th sounding. It is clear that $\underline{\xi}_1$, $\underline{\xi}_2$, $\underline{\xi}_3$, $\underline{\xi}_4$ and $\underline{\xi}_5$ are all $(151\times3)\times1$ column vectors since each depth contains m = 3 data (i.e., X_1, X_2 and X_3) at each sounding. It is also clear that $\boldsymbol{\xi}$ is a $(5\times151\times3)\times1$ column vector. $\boldsymbol{\xi}$ contain 363 observed data (i.e., $\boldsymbol{\xi}^o$) and 1902 unobserved data (i.e., $\boldsymbol{\xi}^u$).

The sounding-wise GS first simulate $T = 4\times10^4+1000$ samples of $\boldsymbol{\xi}^u$. The initial 1000 samples are discarded because they cannot represent the target distribution $f(\boldsymbol{\xi}^u|\boldsymbol{\xi}^o)$. To mitigate the correlation issues between the adjacent samples, the remaining 4×10^4 samples are re-sampled with an interval of 4 to obtained $N_{GS} = 1\times10^4$ samples of $\boldsymbol{\xi}^u$. The depth-wise MCS does not need to discard the initial samples and the samples simulated by depth-wise MCS are independent. Therefore, the depth-wise MCS can simulate $N_{MCS} = 1\times10^4$ samples of $\boldsymbol{\xi}^u$ directly. When the $\underline{\xi}^u$ samples are simulated, they are converted to B_q, $\ln(Q_t)$ and $\ln(Q_e)$ samples by Eq. (30). Based on the 1×10^4 simulated samples of B_q, $\ln(Q_t)$ and $\ln(Q_e)$ samples, the conditional median and 95% confidential interval (CI) can be estimated

with the MATLAB function prctile(). The analytical results can be calculated using the brute-force method by a revised version of Eqs. (12) and (13) with $\boldsymbol{\xi}'$ replaced by $\boldsymbol{\xi}^u$. The analytical results are treated as the reference solutions that are used to validate the results of Step 1. The estimation error for the sounding-wise GS and the depth-wise MCS is defined by the absolute value of the difference between the results of the concerned methods (i.e., sounding-wise GS and depth-wise MCS) and the reference solutions. Figure 6 shows the estimation error of sounding-wise GS and depth-wise MCS based on B_q at S1-S2. Using other data at other soundings gives similar comparison results. Therefore, they are not shown here. It is seen that the estimation errors of the median and 95% CI of the proposed methods are less than 0.02, indicating the results of the proposed methods are quite close to the reference solution. Figure 6 only can give the validation results of median and 95% CI of the simulated data. It is also needed to testify the auto-correlation and cross-correlation results. To examine the auto-correlation results, consider the auto-correlations between the ξ^u value at the first un-observed depth at a sounding and the ξ^u value at the remaining un-observed depths at this sounding. The correlations can be estimated by the 1×10^4 simulated samples of $\boldsymbol{\xi}^u$ with the MATLAB function cov(). The reference solution is also provided by the brute-force method. Figure 7 shows that the estimation error of the sounding-wise GS and depth-wise MCS are less than 0.04. To examine the cross-correlation, consider the cross-correlations between the simulated X_1, X_2 and X_3 values at each sounding. The simulated X_1, X_2 and X_3 values can be extracted from the 1×10^4 simulated samples of $\boldsymbol{\xi}^u$, based on which the cross-correlations among X_1, X_2 and X_3 at each sounding can be estimated. The reference solution is also provided by the brute-force method. Figure 8 shows the comparison results. It can be seen that the estimated results by the sounding-wise GS and depth-wise MCS are close to the reference solution. Figure 8 also shows the prescribed unconditional cross-correlation coefficient among X_1, X_2 and X_3. The conditional cross-correlation is reduced compared with the unconditional cross-correlation at the depths where its adjacent depths contain observed data. The detailed reasons for this reduction are provided in Appendix.

5.3. Step 2: simulate ξ' conditioning on ξ^o

It is indicated in Step 1 that both the sounding-wise GS and the depth-wise MCS can accurately simulate the $\boldsymbol{\xi}^u$ conditioning on $\boldsymbol{\xi}^o$. Therefore, this section only presents the Step 2 results based on the 1×10^4 samples of $\boldsymbol{\xi}^u$ simulated by the sounding-wise GS. In Step 2, each $\boldsymbol{\xi}^u$ sample simulated by the sounding-wise GS is used to simulate a $\boldsymbol{\xi}'$ sample. As a result, Step 2 finally obtains 1×10^4 samples of $\boldsymbol{\xi}'$. Similar to Step 1, $\boldsymbol{\xi}'$ samples can be converted to the realizations of the multivariate 3D conditional random field of B_q, $\ln(Q_t)$ and $\ln(Q_e)$ with the Eq. (30). Figure 9 shows a typical realization of the obtained trivariate 3D cross-correlated conditional random fields of B_q, $\ln(Q_t)$ and $\ln(Q_e)$. It is difficult to visualize and validate the 3D cross-correlated conditional random fields. For clarification, the visualization and validation are conducted for the five sounding locations (S1, S2, S3, S4, S5) and another five un-explored locations (T1, T2, T3, T4, T5) (see Figure 5).

Based on the 1×10^4 realizations of the trivariate 3D cross-correlated conditional random fields, the conditional median and 95% confidence interval (CI) for B_q, $\ln(Q_t)$ and $\ln(Q_e)$ can be estimated. Figure 10 shows the conditional mean and 95% CI of B_q and $\ln(Q_t)$ at the (S1, S2, S3, S4 and S5) locations along the depth direction. The results for $\ln(Q_e)$ are similar to the results for B_q, $\ln(Q_t)$, they thus are not shown for simplification. For comparison, 10 realizations of B_q and $\ln(Q_t)$ are shown as grey lines in Figure 10. The depth interval in Figure 10 is 0.1 m because B_q, $\ln(Q_t)$ and $\ln(Q_e)$ are simulated according to the RFE mesh that has a vertical interval of 0.1 m. Note that the values of B_q and $\ln(Q_t)$ at some depths are actually observed. For these depths, the B_q and $\ln(Q_t)$ values are plotted as filled markers for comparison. For the depths where the values of B_q and $\ln(Q_t)$ are not observed, the values of B_q and $\ln(Q_t)$ (treated as unknown during analysis) are plotted as open markers. Finally, the reference analytical solution calculated by the brute-force method is also shown in Figure 10, which shows that the Step 2 results are close to the reference solution. Figure 11 shows the results for the conditional median, 95% CI, 10 realizations of B_q and $\ln(Q_t)$, and the reference solution for the un-explored locations T1-T5. Again, the Step 2 results are close to the reference solution. It is

560 interesting to observe how the width of the 95% CI depends on the distance to a nearby sounding.
561 The 95% CI in Figure 11 for B_q and $\ln(Q_t)$ at T1 and T5 is relatively narrow compared with that at
562 T2-T4. This is because T1 and T5 are very close to the explored soundings S1 and S2, respectively.
563 The uncertainty at T1 and T5 is effectively reduced by the nearby observed data. In contrast, such
564 uncertainty reduction decreases as the un-explored locations are farther away from the explored
565 soundings. This can be further observed from Figure 12. Figure 12 shows the mean and standard
566 deviations of the simulated B_q, $\ln(Q_t)$ and $\ln(Q_e)$ at A-A section (see Figure 4). The mean values of
567 the simulated data at the observed locations are well constrained to the observed data. The uncertainty
568 of the simulated data reduces significantly at the adjacent locations of the observed locations. This
569 indicates that the extended method can properly simulate the multivariate 3D cross-correlated
570 conditional random fields.

571 5.4. Comparison of computational efficiency among different methods

572 Table 1 shows the comparison results of the proposed method and the brute-force method.
573 Compared with the brute-force method, both sounding-wise GS and depth-wise MCS can reduce the
574 size of the matrices that need to be inverted or decomposed with Cholesky. For instance, the brute-
575 force method requires the inversion of the covariance matrix $\mathrm{Var}(\boldsymbol{\xi}^o)$ (with a size of 363×363 for the
576 investigated example) and Cholesky decomposition of the covariance matrix $\mathrm{Var}(\boldsymbol{\xi}'|\boldsymbol{\xi}^o)$ (with a size
577 up to 1177890 × 1177890). Nevertheless, the sounding-wise GS only requires five evaluations of the
578 inversions for $\mathbf{R}_{h,\boldsymbol{\xi}_{\sim 1}}$, $\mathbf{R}_{h,\boldsymbol{\xi}_{\sim 2}}$, …, $\mathbf{R}_{h,\boldsymbol{\xi}_{\sim nh}}$ (with a size of $(n_h\text{-}1)\times(n_h\text{-}1)=4\times4$) and $\mathrm{Var}(\xi_1^o|\boldsymbol{\xi}_{\sim 1})$,
579 $\mathrm{Var}(\xi_2^o|\boldsymbol{\xi}_{\sim 2})$, …, $\mathrm{Var}(\xi_{nh}^o|\boldsymbol{\xi}_{\sim nh})$ (with a size up to $[m\times n_v]\times[m\times n_v] = 123\times123$). In addition, five
580 Cholesky decompositions need to be performed for covariance matrix $\mathrm{Var}(\xi_k^u|\xi_k^o,\boldsymbol{\xi}_{\sim k})$ (with a size up
581 to $[m\times n_v]\times [m\times n_v] = 409\times409$). The depth-wise MCS requires 151 evaluations for the inversion of
582 the covariance matrix $\mathrm{Var}(\xi_{zmin}^o, \mathbf{\ldots}, \xi_{zmax}^o, \xi_{zmin}^u, \ldots, \xi_{zi\text{-}1}^u)$ (with a size up to 369×369 for the
583 investigated example) and Cholesky decomposition of the covariance matrix $\mathrm{Var}(\xi_{zi}^u|\xi_{zmin}^o, \mathbf{\ldots},$
584 $\xi_{zmax}^o, \xi_{zmin}^u, \ldots, \xi_{zi\text{-}1}^u)$ (with a size up to 15×15). Step 2 of the proposed method requires one

585 evaluation of the inversion of the covariance matrix $\mathbf{R}_{v,\xi}$ (with a size up to 151×151) and Cholesky
586 decomposition of the covariance matrix $\mathbf{R}_{v,\xi'}$ (with a size up to 2601×2601). Table 1 also compared
587 the computational time required by the proposed method and the brute-force method. The
588 computation time is based on the MATLAB platform in a desktop computer with 64 GB RAM and
589 Intel Core i9-10900K CPU clocked at 3.70GHz. To simulate 1×10^4 number of $\boldsymbol{\xi}^u$ samples, the
590 sounding-wise GS takes 1.87 minutes, which is favorably comparable with that of 0.98 minutes taken
591 by the depth-wise MCS. Step 2 takes 10.99 minutes. In total, the proposed method takes about 12
592 minutes to generate 1×10^4 realizations of cross-correlated conditional random fields of B_q, $\ln(Q_t)$ and
593 $\ln(Q_e)$. The brute-force method is not able to simulate the trivariate 3D cross-correlated conditional
594 random fields because the storage space needed to operate the matrix $\mathrm{Var}(\boldsymbol{\xi}'|\boldsymbol{\xi}^o)$ (with a size up to
595 1177890×1177890) for the brute-force method exceeds the memory of the used desk computer.

596 It is noted that, for the depth-wise MCS, ($\xi_{z\min}^o$, **…**, $\xi_{z\max}^o$, $\xi_{z\min}^u$, …, $\xi_{z_{i-1}}^u$) collectively does not
597 have a lattice structure in general, because ξ^u at depths of z_{i+1}, …, z_{max} has not yet been simulated
598 when simulating the $\xi_{z_i}^u$, hence it is in general not feasible to represent the covariance matrix
599 $\mathrm{Var}(\xi_{z\min}^o$, **…**, $\xi_{z\max}^o$, $\xi_{z\min}^u$, …, $\xi_{z_{i-1}}^u)$ as a Kronecker product. The size of this covariance matrix is
600 $(m\times n_h\times n_{cd})\times(m\times n_h\times n_{cd})$, where n_{cd} denotes the number of conditioning depths, i.e., the number of
601 sounding depths within the range $[z_i-\alpha\delta_v, z_i+\alpha\delta_v]$. n_{cd} is roughly equal to $2\alpha\delta_v$/(sounding depth
602 interval). When δ_v is large or α takes a very large value, the range $[z_i-\alpha\delta_v, z_i+\alpha\delta_v]$ might cover the
603 whole depth range $[z_1, z_{nv}]$, the depth-wise MCS's computational efficiency might be reduced since
604 the size of the covariance matrix to be inverted becomes large. For instance, if α takes a value of 80,
605 simulating the data at the i-th depth needs to condition on the data at all the depths. As a result, the
606 matrices need to be inverted or make Cholesky decomposition larger. The advantage of depth-wise
607 MCS is that it can generate independent samples. On the contrary, the sounding-wise GS method
608 requires to use the Gibbs sampler to simulate the un-observed data at each sounding. When the
609 horizontal scale fluctuation is very large, the data at each sounding have a strong correlation. The re-

sampler interval at the sounding-wise GS should be large enough to avoid the correlation between samples. This may lead to the reduction of the computational efficiency of the sounding-wise GS.

5.5. Sites with layered soils

The sites in practice may contain soil layers involving multiple soil types. When a site with layered soils is concerned, one needs to simulate the conditional random fields of each type of soil over the 3D domain and then map them onto the random field mesh according to the boundaries of these soil layers. The simulation procedure for the conditional random fields of each type of soil is similar to that of the investigated virtual site. The only difference is the augmenting depths.

6. Conclusions

This paper proposes a novel method for efficiently simulating the multivariate 3D conditional cross-correlated random fields based on a recently developed univariate three-dimensional (3D) conditional random field method. The proposed method is more versatile in the sense that it can handle the multivariate incomplete sounding data. The proposed method contains two steps: Step 1 simulates the missing sounding data ξ^u to make the sounding data complete and Step 2 simulates the conditional random fields based on the make-complete sounding data ξ. A simulated example of trivariate cross-correlated random fields is adopted to illustrate the effectiveness of the proposed method. It is found that the proposed method can effectively estimate the median, 95% CI, auto-correlation and cross-correlation of the multivariate conditional random fields, producing results that are close to the analytical reference solution. The proposed method can simulate multivariate 3D cross-correlated conditional random fields without handling the large matrices (e.g., matrix inversions or Cholesky decompositions). It also does not need to store large matrices. It may provide an effective tool to simulate the multivariate 3D cross-correlated conditional random fields for a large 3D domain or with a fine resolution.

Acknowledgements

636 This work was supported by the National Natural Science Foundation of China (No. 51909288), the
637 Guangdong Provincial Department of Science and Technology (2019ZT08G090) and the Open
638 Innovation Fund of Changjiang Institute of Survey, Planning, Design and Research (No.
639 CX2020K07).

640

641 **Appendix A. Proof of reduction of cross-correlations in conditional random fields**

642 Consider the data at the i-th depth and j-th for example. It is assumed that each depth contains two
643 soil property parameters: a and b. Suppose that cross-covariance matrix between a and b is [σ_a^2
644 $\rho\sigma_a\sigma_b$;$\rho\rho\sigma_a\sigma_b$ σ_b^2] and the auto-correlation matrix between the i-th depth and j-th depth is [1 r_{ij}; r_{ji} 1].
645 The data at the two depths follow the following normal distribution:

646
$$f\left(x_{ai},x_{bi},x_{aj},x_{bj}\right)=N\left(\left[\mu_a,\mu_b,\mu_a,\mu_b\right],\begin{bmatrix}\sigma_a^{\,2} & \rho\sigma_a\sigma_b & r_{ij}\sigma_a^{\,2} & r_{ij}\rho\sigma_a\sigma_b\\ \rho\sigma_a\sigma_b & \sigma_b^{\,2} & r_{ij}\rho\sigma_a\sigma_b & r_{ij}\sigma_b^{\,2}\\ r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b & \sigma_a^{\,2} & \rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2} & \rho\sigma_a\sigma_b & \sigma_b^{\,2}\end{bmatrix}\right) \tag{A1}$$

647 If the data at the i-th depth are observed, the covariance between the data at the j-th depth is estimated
648 as

649
$$\begin{aligned}
\mathrm{Cov}\left(x_{aj},x_{bj}\mid x_{ai},x_{bi}\right)&=\begin{bmatrix}\sigma_a^{\,2} & \rho\sigma_a\sigma_b\\ \rho\sigma_a\sigma_b & \sigma_b^{\,2}\end{bmatrix}-\begin{bmatrix}r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2}\end{bmatrix}\begin{bmatrix}\sigma_a^{\,2} & \rho\sigma_a\sigma_b\\ \rho\sigma_a\sigma_b & \sigma_b^{\,2}\end{bmatrix}^{-1}\begin{bmatrix}r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2}\end{bmatrix}^{\mathrm{T}}\\
&=\begin{bmatrix}\sigma_a^{\,2} & \rho\sigma_a\sigma_b\\ \rho\sigma_a\sigma_b & \sigma_b^{\,2}\end{bmatrix}-\begin{bmatrix}r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2}\end{bmatrix}\begin{bmatrix}\sigma_b^{\,2} & -\rho\sigma_a\sigma_b\\ -\rho\sigma_a\sigma_b & \sigma_a^{\,2}\end{bmatrix}\begin{bmatrix}r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2}\end{bmatrix}\frac{1}{\sigma_a^{\,2}\sigma_b^{\,2}-\left(\rho\sigma_a\sigma_b\right)^2}\\
&=\begin{bmatrix}\sigma_a^{\,2} & \rho\sigma_a\sigma_b\\ \rho\sigma_a\sigma_b & \sigma_b^{\,2}\end{bmatrix}-\begin{bmatrix}r_{ji}\sigma_a^{\,2}\sigma_b^{\,2}-r_{ji}\left(\rho\sigma_a\sigma_b\right)^2 & 0\\ 0 & r_{ji}\sigma_b^{\,2}\sigma_a^{\,2}-r_{ji}\left(\rho\sigma_a\sigma_b\right)^2\end{bmatrix}\begin{bmatrix}r_{ji}\sigma_a^{\,2} & r_{ji}\rho\sigma_a\sigma_b\\ r_{ji}\rho\sigma_a\sigma_b & r_{ji}\sigma_b^{\,2}\end{bmatrix}\frac{1}{\sigma_a^{\,2}\sigma_b^{\,2}-\left(\rho\sigma_a\sigma_b\right)^2}\\
&=\begin{bmatrix}\cdots & \rho\sigma_a\sigma_b-r_{ji}^2\rho\sigma_a\sigma_b\\ \cdots & \cdots\end{bmatrix}=\begin{bmatrix}\cdots & \rho\sigma_a\sigma_b\left(1-r_{ji}^{\,2}\right)\\ \cdots & \cdots\end{bmatrix}
\end{aligned} \tag{A2}$$

650 It is seen that the conditional cross-correlation between a and b at the j-th depth depends on the
651 distance between the i-the depth and j-th depth. The closer distance gives the larger auto-correlation
652 coefficient r_{ji}, and further results in larger reduction of the cross-correlation.

653

654 **References**

655 Bjerrum L. (1973). Problems of soil mechanics and construction on soft clays and structurally
656 unstable soils (collapsible, expansive and others). In: Proceedings of the 8th International
657 Conference on Soil Mechanics and Foundation Engineering, Moscow.

658 Cai, Y., Li, J., Li, X., Li, D., and Zhang, L. (2019). Estimating soil resistance at unsampled locations
659 based on limited CPT data. Bulletin of Engineering Geology and the Environment, 78(5), 3637-
660 3648.

661 Cao, Z., and Wang, Y. (2014). Bayesian model comparison and selection of spatial correlation
662 functions for soil parameters. Structural Safety, 49, 10-17.

663 Cami, B., Javankhoshdel, S., Phoon, K. K., and Ching, J. (2020). Scale of fluctuation for spatially
664 varying soils: Estimation methods and values. ASCE-ASME Journal of Risk and Uncertainty in
665 Engineering Systems, Part A: Civil Engineering, 6(4), 03120002.

666 Cheng, H., Chen, J., Chen, R., Huang, J., and Li, J. (2019). Three-dimensional analysis of tunnel face
667 stability in spatially variable soils. Computers and Geotechnics, 111, 76-88.

668 Ching, J. Y., and Phoon, K. K. (2014). Correlations among some clay parameters - the multivariate
669 distribution. Canadian Geotechnical Journal, 51(6), 686-704.

670 Ching, J., Huang, W. H., Phoon, K. K. (2020). 3D Probabilistic site characterization by sparse
671 bayesian learning. Journal of Engineering Mechanics, 146(12), 04020134.

672 Ching, J., Yang, Z., Phoon, K. K. (2021). Dealing with nonlattice data in three-dimensional

673 probabilistic site characterization. Journal of Engineering Mechanics, 147(5), 06021003.
674 Cho, S. E. (2010). Probabilistic assessment of slope stability that considers the spatial variability of
675 soil properties. Journal of geotechnical and geoenvironmental engineering, 136(7), 975-984.
676 DeGroot, D. J., and Baecher, G. B. (1993). Estimating autocovariance of in-situ soil properties.
677 Journal of Geotechnical Engineering, 119(1), 147-166.
678 El Haj, A. K., Soubra, A. H., and Fajoui, J. (2019). Probabilistic analysis of an offshore monopile
679 foundation taking into account the soil spatial variability. Computers and Geotechnics, 106, 205-
680 216.
681 Fenton, G. A., and Griffiths, D. V. (2002). Probabilistic foundation settlement on spatially random
682 soil. Journal of Geotechnical & Geoenvironmental Engineering, 128(5), 381-390.
683 Fenton, G. A., and Griffiths, D. V. (2003). Bearing-capacity prediction of spatially random c φ soils.
684 Canadian geotechnical journal, 40(1), 54-65.
685 Fenton, G. A., and Vanmarcke, E. H. (1990). Simulation of random fields via local average
686 subdivision. Journal of Engineering Mechanics, 116(8), 1733-1749.
687 Fenton, G. A., Griffiths, D. V., and Williams, M. B. (2005). Reliability of traditional retaining wall
688 design. Géotechnique, 55(1), 55-62.
689 Gong, W., Tang, H., Juang, C. H., and Wang, L. (2020). Optimization design of stabilizing piles in
690 slopes considering spatial variability. Acta Geotechnica (online).
691 Griffiths, D. V., and Fenton, G. A. (2004). Probabilistic slope stability analysis by finite elements.
692 Journal of geotechnical and geoenvironmental engineering, 130(5), 507-518.

693 Hu, Y. G., and Ching, J. (2014). The critical scale of fluctuation for active lateral forces in spatially
694 variable undrained clays. Computers and Geotechnics, 57, 24-29.
695 Hu, Y. G., and Ching, J. (2015). Impact of spatial variability in undrained shear strength on active
696 lateral force in clay. Structural Safety, 52, 121-131.
697 Huang, J., Lyamin, A. V., Griffiths, D. V., Krabbenhoft, K., and Sloan, S. W. (2013). Quantitative risk
698 assessment of landslide by limit analysis and random fields. Computers and Geotechnics, 53,
699 60-67.
700 Huang, L., Cheng, Y. M., Leung, Y. F., and Li, L. (2019). Influence of rotated anisotropy on slope
701 reliability evaluation using conditional random field. Computers and Geotechnics, 115, 103133.
702 Jaksa, M., Kaggwa, W.S., and Brooker, P.I. (1999). Experimental evaluation of the scale of fluctuation
703 of a stiff clay. Proceedings of the 8th International Conference on Application of Statistics and
704 Probability, A.A. Balkema, Rotterdam, 415-422.
705 Jha, S. K., and Ching, J. (2012). Simulating spatial averages of stationary random field using the
706 Fourier series method. Journal of Engineering Mechanics, 139(5), 594-605.
707 Ji, J., Liao, H. J., and Low, B. K. (2012). Modeling 2-D spatial variation in slope reliability analysis
708 using interpolated autocorrelations. Computers and Geotechnics, 40, 135-146.
709 Jiang, S. H., Huang, J., Huang, F., Yang, J., Yao, C., and Zhou, C. B. (2018). Modelling of spatial
710 variability of soil undrained shear strength by conditional random fields for slope reliability
711 analysis. Applied Mathematical Modelling, 63, 374-389.
712 Jiang, S. H., Li, D. Q., Cao, Z. J., Zhou, C. B., and Phoon, K. K. (2015). Efficient system reliability

713 analysis of slope stability in spatially variable soils using Monte Carlo simulation. Journal of
714 geotechnical and geoenvironmental engineering, 141(2), 04014096.
715 Kulhawy, F. H., and Mayne, P. W. (1990). Manual on estimating soil properties for foundation design
716 (No. EPRI-EL-6800). Electric Power Research Inst., Palo Alto, CA (USA); Cornell Univ., Ithaca,
717 NY (USA). Geotechnical Engineering Group.
718 Kawa, M., and Puła, W. (2019). 3D bearing capacity probabilistic analyses of footings on spatially
719 variable c–φ soil. Acta Geotechnica, 1-14.
720 Li, D. Q., Xiao, T., Cao, Z. J., Zhou, C. B., and Zhang, L. M. (2016a). Enhancement of random finite
721 element method in reliability analysis and risk assessment of soil slopes using Subset Simulation.
722 Landslides, 13(2), 293-303.
723 Li, D. Q., Xiao, T., Zhang, L. M., and Cao, Z. J. (2019a). Stepwise covariance matrix decomposition
724 for efficient simulation of multivariate large-scale three-dimensional random fields. Applied
725 Mathematical Modelling, 68, 169-181.
726 Li, D. Q., Yang, Z. Y., Cao, Z. J., amd Zhang, L. M. (2019b). Area failure probability method for
727 slope system failure risk assessment. Computers and Geotechnics, 107, 36-44.
728 Li, L., Wang, Y., and Cao, Z. (2014). Probabilistic slope stability analysis by risk aggregation.
729 Engineering Geology, 176, 57-65.
730 Li, X. Y., Zhang, L. M., and Li, J. H. (2016b). Using conditioned random field to characterize the
731 variability of geologic profiles. Journal of Geotechnical and Geoenvironmental Engineering,
732 142(4), 04015096.

733 Li, Y. J., Hicks, M. A., and Nuttall, J. D. (2015). Comparative analyses of slope reliability in 3D.
734 Engineering Geology, 196, 12-23.
735 Li, Y. J., Hicks, M. A., and Vardon, P. J. (2016c). Uncertainty reduction and sampling efficiency in
736 slope designs using 3D conditional random fields. Computers and Geotechnics, 79, 159-172.
737 Liu, L. L., Cheng, Y. M., and Zhang, S. H. (2017). Conditional random field reliability analysis of a
738 cohesion-frictional slope. Computers and Geotechnics, 82, 173-186.
739 Liu, Y., Zhang, W., Zhang, L., Zhu, Z., Hu, J., and Wei, H. (2018). Probabilistic stability analyses of
740 undrained slopes by 3D random fields and finite element methods. Geoscience Frontiers, 9(6),
741 1657-1664.
742 Lo, M. K., and Leung, Y. F. (2017). Probabilistic analyses of slopes and footings with spatially
743 variable soils considering cross-correlation and conditioned random field. Journal of
744 Geotechnical and Geoenvironmental Engineering, 143(9), 04017044.
745 Ou-Yang, J. Y., Li, D. Q., Tang, X. S., Liu, Y. (2021). A patching algorithm for conditional random
746 fields in modeling material properties. Computer Methods in Applied Mechanics and
747 Engineering, 377, 113719.
748 Phoon, K. K., Huang, S. P., and Quek, S. T. (2002). Implementation of Karhunen-Loeve expansion
749 for simulation using a wavelet-Galerkin scheme. Probabilistic Engineering Mechanics, 17(3),
750 293-303.
751 Popescu, R., Deodatis, G., and Nobahar, A. (2005). Effects of random heterogeneity of soil properties
752 on bearing capacity. Probabilistic Engineering Mechanics, 20(4), 324-341.

753 Peng, X. Y., Zhang, L. L., Jeng, D. S., Chen, L. H., Liao, C. C., and Yang, H. Q. (2017). Effects of
754 cross-correlated multiple spatially random soil properties on wave-induced oscillatory seabed
755 response. Applied Ocean Research, 62, 57-69.
756 Qi, X. H., and Li, D. Q. (2018). Effect of spatial variability of shear strength parameters on critical
757 slip surfaces of slopes. Engineering Geology, 239, 41-49.
758 Qi, X. H., and Liu, H. X. (2019). Estimation of autocorrelation distances for in-situ geotechnical
759 properties using limited data. Structural Safety, 79, 26-38.
760 Robin, M. J. L., Gutjahr, A. L., Sudicky, E. A., and Wilson, J. L. (1993). Cross-correlated random
761 field generation with the direct Fourier Transform Method. Water Resources Research, 29(7),
762 2385-2397.
763 Song, K. I., Cho, G. C., and Lee, S. W. (2011). Effects of spatially variable weathered rock properties
764 on tunnel behavior. Probabilistic Engineering Mechanics, 26(3), 413-426.
765 Stuedlein, A.W., Kramer, S.L., Arduino, P., and Holtz, R.D. (2012). Reliability of Spread Footing
766 Performance in Desiccated Clay. Journal of Geotechnical and Geoenvironmental Engineering,
767 ASCE, 138(11), 1314-1325.
768 Uzielli, M., Vannucchi, G., and Phoon, K. K. (2005). Random field characterisation of stress-
769 nomalised cone penetration testing parameters. Geotechnique, 55(1), 3-20.
770 Varkey, D., Hicks, M. A., and Vardon, P. J. (2019). An improved semi-analytical method for 3D slope
771 reliability assessments. Computers and Geotechnics, 111, 181-190.
772 Vořechovský, M. (2008). Simulation of simply cross correlated random fields by series expansion

773 methods. Structural safety, 30(4), 337-363.
774 Vanmarcke, E. H. (1977). Probabilistic modeling of soil profiles. ASCE Journal of Geotechnical
775 Engineering, GT11, 1227-1246.
776 Wang, Y., Cao, Z., and Au, S. K. (2011). Practical reliability analysis of slope stability by advanced
777 Monte Carlo simulations in a spreadsheet. Canadian Geotechnical Journal, 48(1), 162-172.
778 Wu, G., Zhao, H., Zhao, M., and Zhu, Z. (2021). Stochastic analysis of dual tunnels in spatially
779 random soil. Computers and Geotechnics, 129, 103861.
780 Xiao, T., Li, D. Q., Cao, Z. J., and Zhang, L. M. (2018). CPT-based probabilistic characterization of
781 three-dimensional spatial variability using MLE. Journal of Geotechnical and Geoenvironmental
782 Engineering, 144(5), 04018023.
783 Xiao, T., Li, D. Q., Cao, Z. J., Au, S. K., and Phoon, K. K. (2016). Three-dimensional slope reliability
784 and risk assessment using auxiliary random finite element method. Computers and Geotechnics,
785 79, 146-158.
786 Yang, Z. Y., and Ching, J. Y. (2020). Simulation of three-dimensional random field conditioning on
787 incomplete site data. Engineering Geology. https://doi.org/10.1016/j.enggeo.2020.105987.
788 Zhao, T., and Wang, Y. (2018). Simulation of cross-correlated random field samples from sparse
789 measurements using Bayesian compressive sensing. Mechanical Systems and Signal Processing,
790 112, 384-400.
791 Zhao, T., and Wang, Y. (2020). Non-parametric simulation of non-stationary non-gaussian 3D random
792 field samples directly from sparse measurements using signal decomposition and Markov Chain

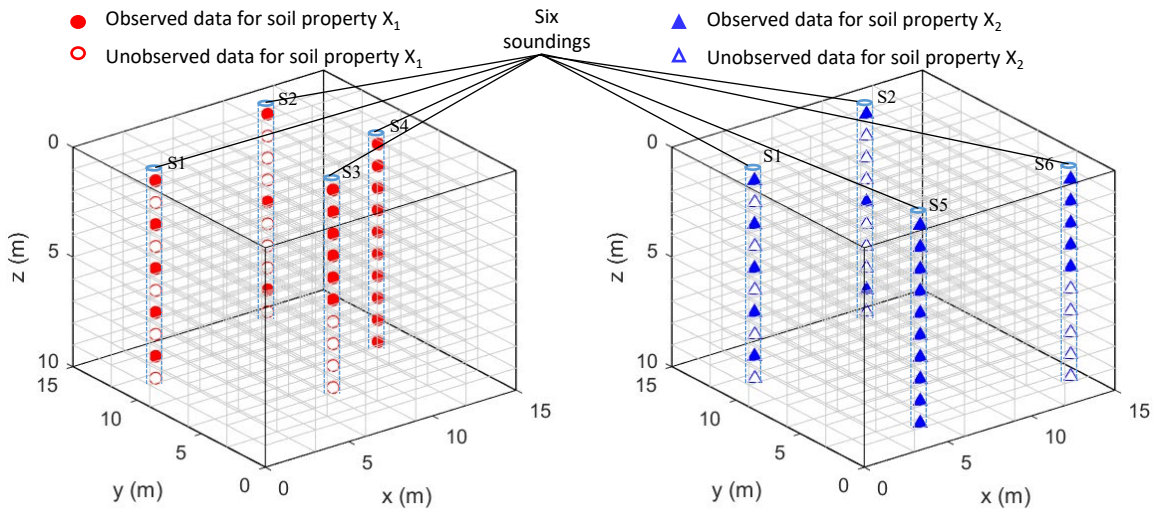
793 Monte Carlo (MCMC) simulation. Reliability Engineering & System Safety, 203, 107087.
794 Zhao, T., and Wang, Y. (2021). Statistical interpolation of spatially varying but sparsely measured 3D
795 Geo-Data using compressive sensing and variational bayesian inference. Mathematical
796 Geosciences, 1-29.
797 Zhao, T., Xu, L., and Wang, Y. (2020). Fast non-parametric simulation of 2D multi-layer cone
798 penetration test (CPT) data without pre-stratification using Markov Chain Monte Carlo
799 simulation. Engineering Geology, 273, 105670.
800 Zhu, H., Zhang, L. M., Xiao, T., Li, X. Y. (2017). Generation of multivariate cross-correlated
801 geotechnical random fields. Computers and Geotechnics, 86, 95-107.
802

803

804 Table 1 Summary of the computational cost

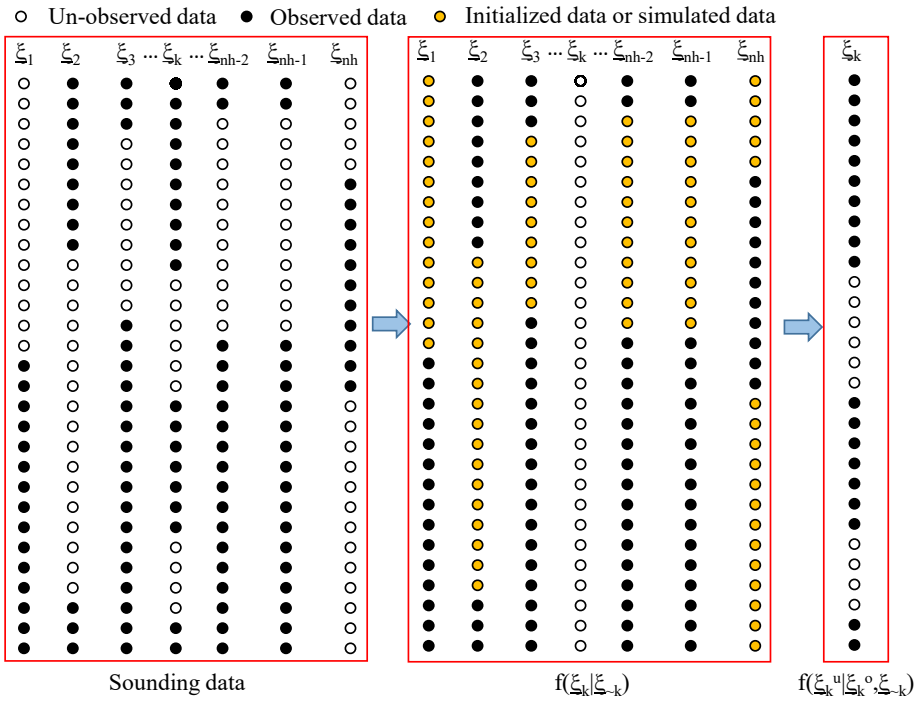
	Proposed method			Brute-force method
	Step 1		Step 2	
	Sounding-wise GS	Depth-wise MCS		
Max. size for matrix inversions (# of evaluations)	123×123 (5)	369×369 (151)	151×151 (1)	363×363 (1)
Max. size for Cholesky decomposition (# of evaluations)	409×409 (5)	15×15 (151)	2601×2601 (1)	1177890×1177890 (1)
Max. size of stored matrices	453×453	369×369	2601×2601	1177890×1177890
Computational time	1.87[a] mins	0.98[b] mins	10.99[c] mins	-[d]

805 Notes:
806 [a] Time for simulating 4×10^4+1000 GS samples of $\boldsymbol{\xi}^u$
807 [b] Time for simulating 1×10^4 MCS samples of $\boldsymbol{\xi}^u$
808 [c] Time for simulating 1×10^4 samples of $\boldsymbol{\xi}'$
809 [d] Result is not available due to the large computational cost
810
811



813 Figure 1 Schematic diagram for sounding data at a typical site (6 soundings in total) and RFEs (5,
814 5 and 10 elements along the x-, y- and z-directions) generated for simulating spatial variability of
815 soil properties X_1 and X_2 of this site

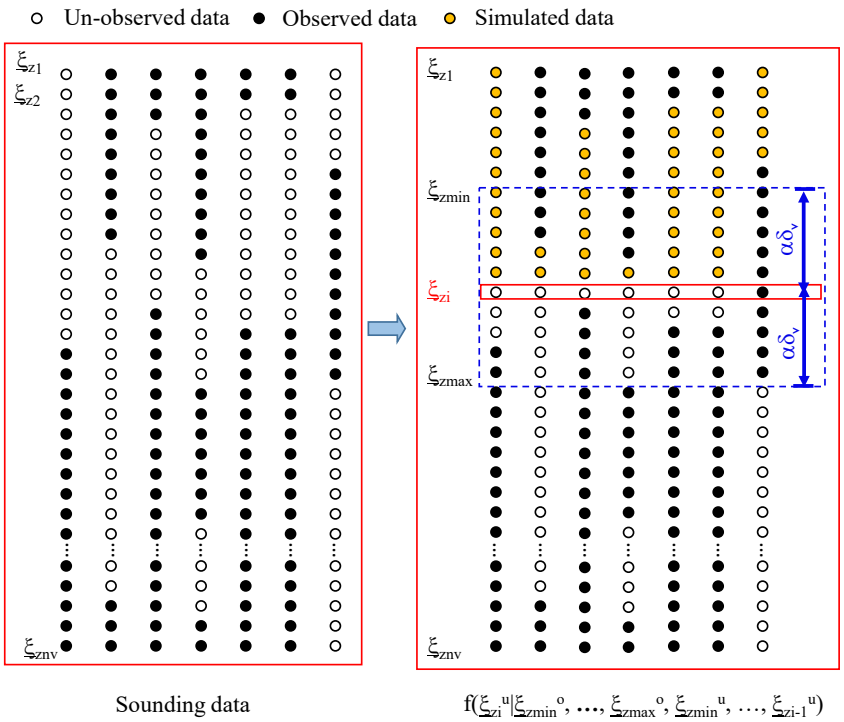
816



818 Figure 2 Schematic diagram of sounding-wise GS for simulating the un-observed data at the k-th
819 sounding (i.e., ξ_k^u)

820

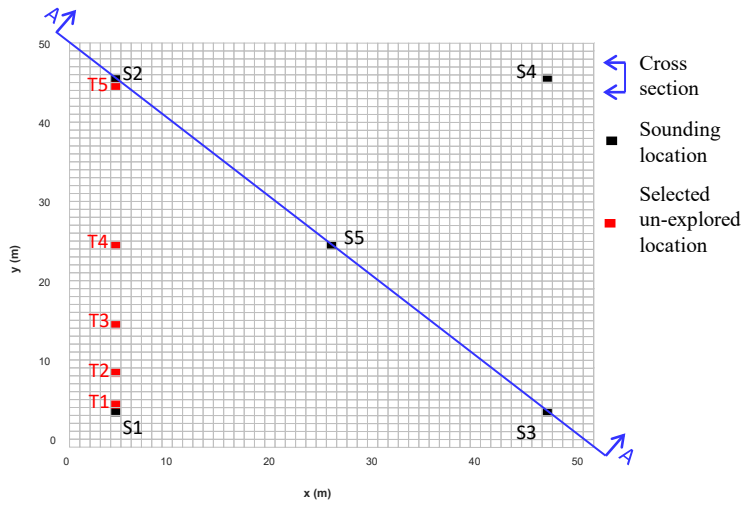
821



823 Figure 3 Schematic diagram of depth-wise MCS for simulating the un-observed data at the i-th

824 depth (i.e., ξ_{zi}^{u})

825



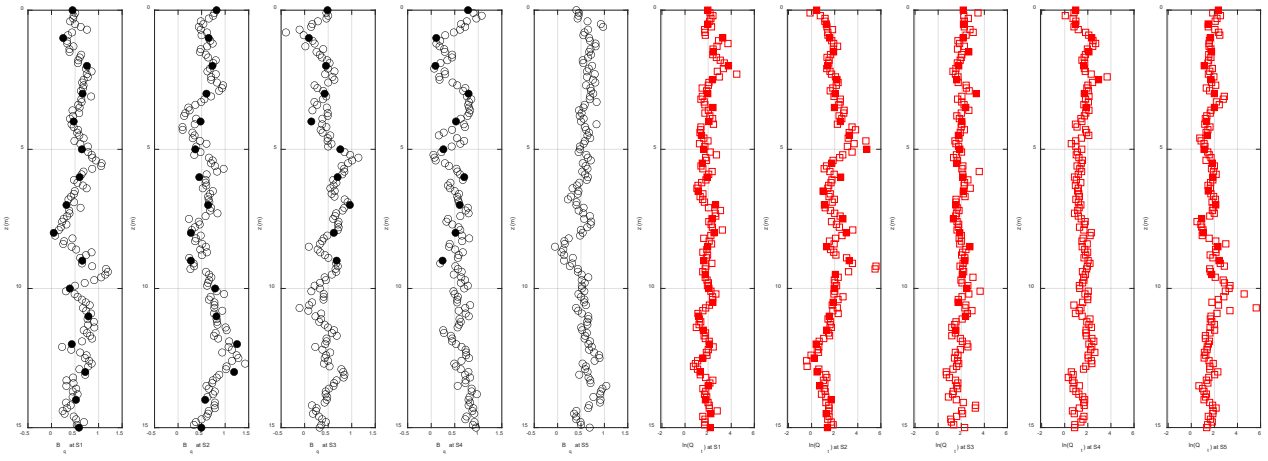
827 Figure 4 Top view for A-A cross section, soundings S1-S5, and selected un-explored location T1-

828 T5

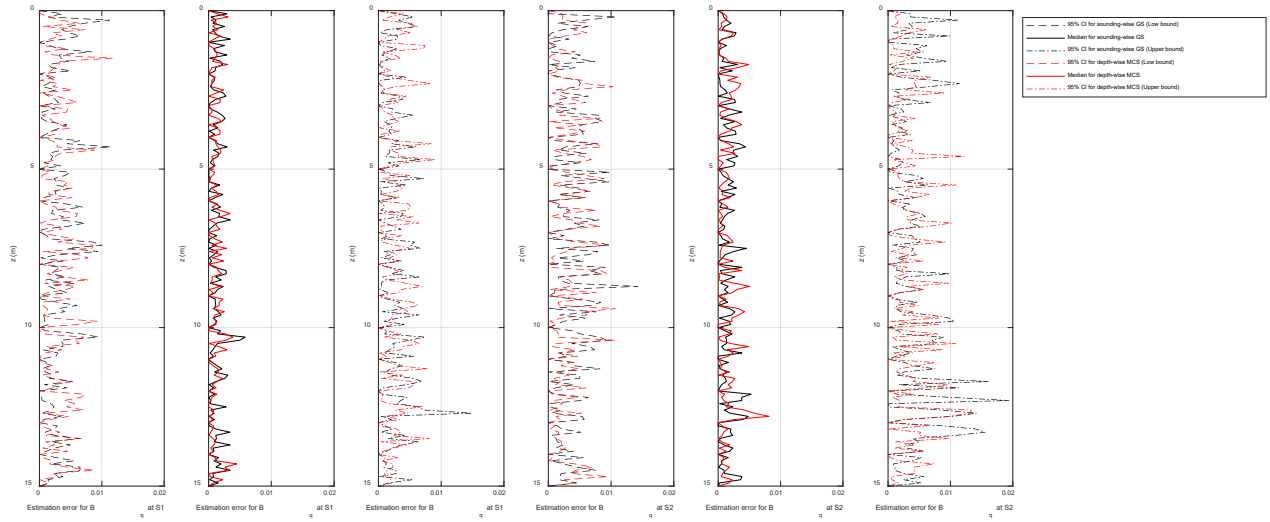
829

830

831

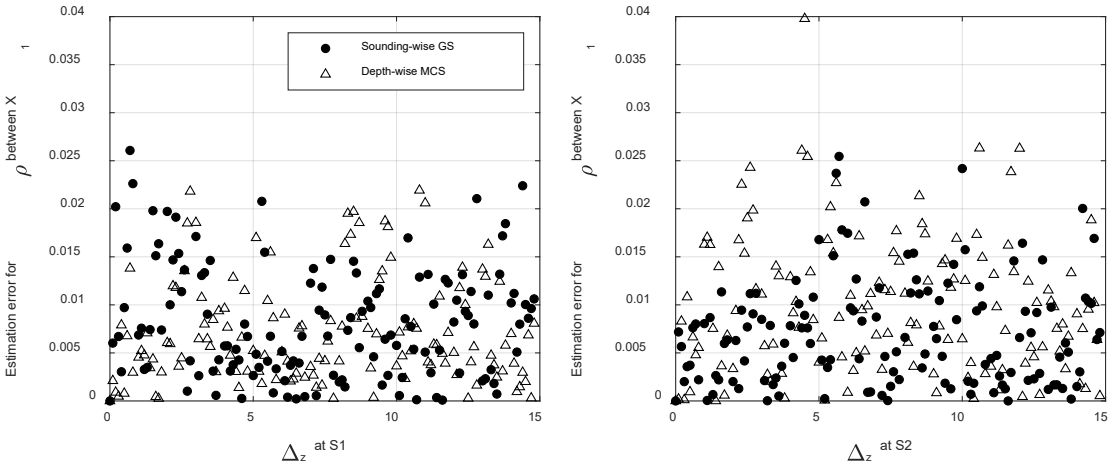


833 Figure 5 Observed (filled marker) and un-observed (open marker) data at S1-S5

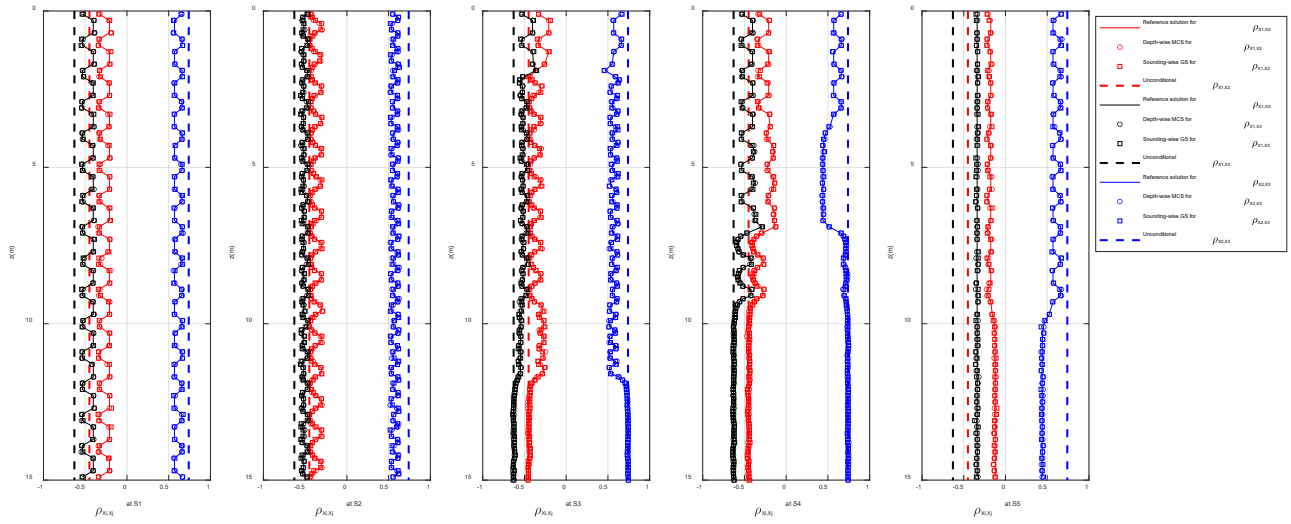


835 Figure 6 Estimation error of sounding-wise GS and depth-wise MCS based on B_q at S1-S2

836



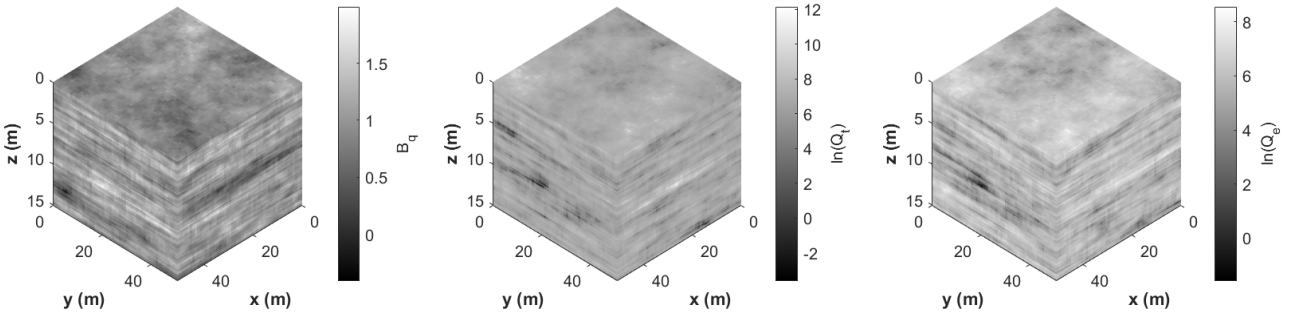
838 Figure 7 Estimation error for the auto-correlations of X_1 at S1-S2



840 Figure 8 Estimation results for the cross-correlations among X_1, X_2 and X_3 at S1-S5

841

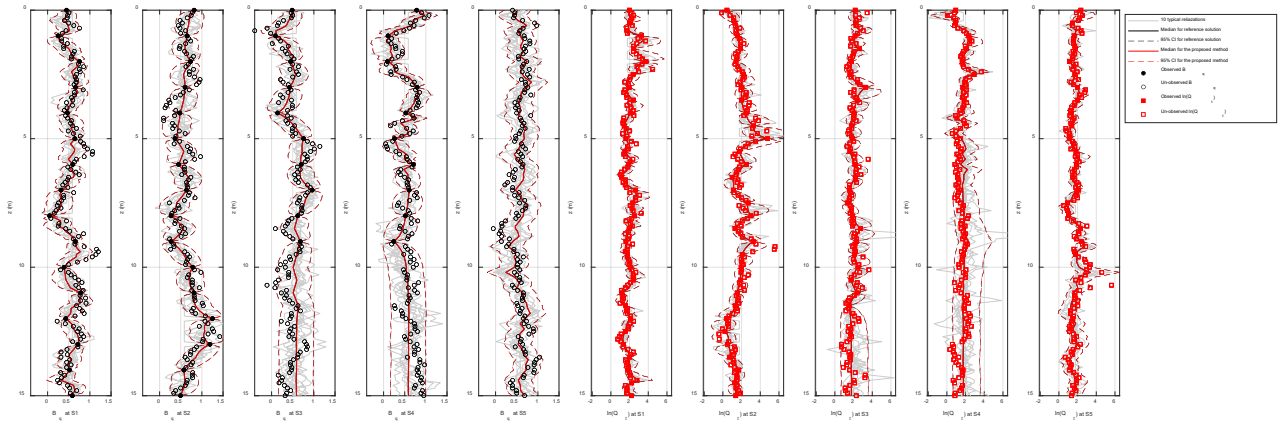
842



844 Figure 9 One typical realization of multivariate 3D cross-correlated conditional random fields of

845 B_q, $\ln(Q_t)$ and $\ln(Q_e)$

846



848 Figure 10 Estimated median and 95% CI profiles of B_q and $\ln(Q_t)$ at S1-S5

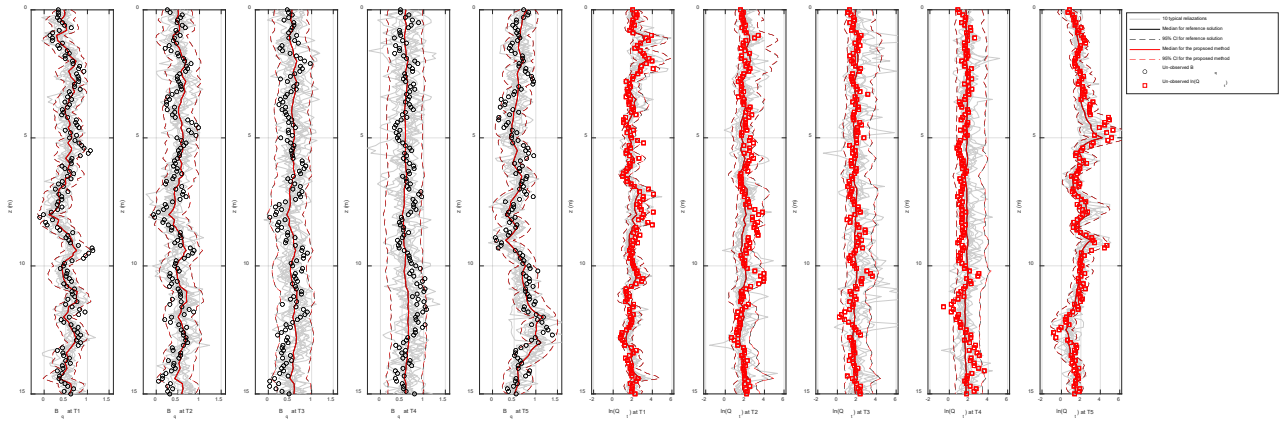


Figure 11 Estimated median and 95% CI profiles of B_q and $\ln(Q_t)$ at T1-T5

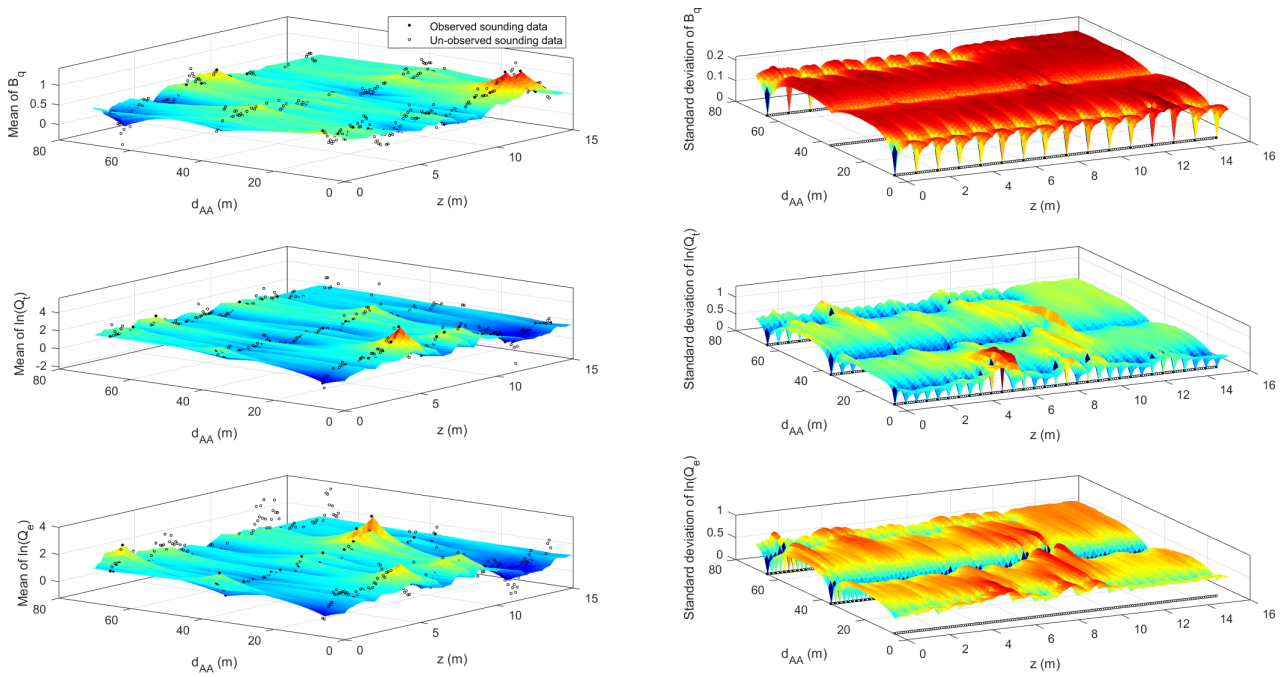


Figure 12 Mean and standard deviation of simulated data at A-A section