Citation: Yang, Zhiyong, Li, Xueyou and Qi, Xiaohui (2022) Efficient simulation of multivariate three-dimensional cross-correlated random fields conditioning on non-lattice measurement data. Computer Methods in Applied Mechanics and Engineering, 388. p. 114208. ISSN 0045-7825

Published by: Elsevier

URL: https://doi.org/10.1016/j.cma.2021.114208
<https://doi.org/10.1016/j.cma.2021.114208>

This version was downloaded from Northumbria Research Link: https://nrl.northumbria.ac.uk/id/eprint/47658/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University’s research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher’s website (a subscription may be required.)
Efficient simulation of multivariate three-dimensional cross-correlated random fields conditioning on non-lattice measurement data

Zhiyong Yang\textsuperscript{1}, Xueyou Li\textsuperscript{2} and Xiaohui Qi\textsuperscript{*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.

\textsuperscript{1} Assistant Professor, School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China. Email: yangzhy85@mail.sysu.edu.cn
\textsuperscript{2} Professor, School of Civil Engineering, Sun Yat-Sen University, Guangzhou, China. Email: lixueyou@mail.sysu.edu.cn
\textsuperscript{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 Pula 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 $\phi$ 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 $S_1$, $S_2$, $(S_3, S_4)$ and $(S_5, S_6)$ 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., $S_3$ and $S_4$). 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
2.1. Assumptions

The study adopts the following four assumptions:

1. The cross-correlation and auto-correlation among various soil property parameters are separable. This implies that the type of auto-correlation function (e.g., single exponential function) and scale of fluctuation for all the soil property parameters are the same and the cross-correlation structure between each pair of simulated random fields is simply defined by a cross-correlation coefficient (Vořechovský 2008). Fenton and Griffiths (2003), Fenton et al. (2005) and Cho (2010) made similar assumptions (identical auto-correlation structure for different soil parameters) by arguing that the spatial correlation of a soil is governed largely by the spatial variability in its source materials, weathering patterns, stress, sedimentation and formation history, etc. so that one would expect that all soil properties will vary similarly between the two points. This assumption is frequently used in the geotechnical field for simulating the multivariate cross-correlated random field (Peng et al. 2017; Zhu et al. 2017; Li et al. 2019a). There is no compelling evidence to support or negate this assumption, and this assumption is adopted in the current study to serve as a baseline for further studies with more complicated auto-correlation models.

2. The auto-correlation functions in the vertical and horizontal directions are separable. This is also a common assumption in 3D random field simulation (Vanmarcke 1977; Stuedlein et al. 2012; Xiao et al. 2018; Li et al. 2019a; Yang and Ching 2020; Ching et al. 2020). There is no compelling evidence to support or negate this assumption, either. This assumption is used mainly for reducing the matrices’ size and improving the computational efficiency.

3. The underlying unconditional random field is stationary and normally distributed. A non-stationary random field with a trend can be transformed to a stationary random field by removing the trend component (Jaksa et al. 1999; Uzielli et al. 2005). A non-normal random field can be transformed from the normal random field with certain transformation methods such as Nataf transformation.

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 \(\xi_i\) contains the data for the \(i\)-th sounding. \(\xi_i = [X_{i,1} \ X_{i,2} \ldots X_{i,n_v}]^T\) is an \(((m \times n_v) \times 1)\) column vector where \(X_{i,j} = [X_1 \ X_2 \ldots X_m]\) denotes the data point at the \(i\)-th sounding at the \(j\)-th depth. Let \(\xi\) contain all sounding data, including \(n_o\) observed data (filled markers in the figure, denoted by \(\xi^o\)) and \(n_u\) un-observed data (open markers in the figure, denoted by \(\xi^u\)). \(\xi\) is an \((N \times 1)\) column vector denoting the union of \(\xi^o\) and \(\xi^u\). \(\xi\) can be assembled by stacking the soundings data vertically, \(\xi = [\xi_1^T \ \xi_2^T \ldots \xi_n^T]^T\). Let us denote the covariance matrix for \(\xi\) by \(\text{Var}(\xi) \in \mathbb{R}^{N \times N}\). Base on the assumption of separability between the auto-correlation and the cross-correlation among the \(m\) number of parameters, \(\text{Var}(\xi)\) can be expressed as

\[
\text{Var}(\xi) = \begin{bmatrix}
\text{R}(\xi_1, \xi_1) & \text{R}(\xi_1, \xi_2) & \ldots & \text{R}(\xi_1, \xi_{n_h}) \\
\text{R}(\xi_2, \xi_1) & \text{R}(\xi_2, \xi_2) & \ldots & \text{R}(\xi_2, \xi_{n_h}) \\
\vdots & \vdots & \ddots & \vdots \\
\text{R}(\xi_{n_h}, \xi_1) & \text{R}(\xi_{n_h}, \xi_2) & \ldots & \text{R}(\xi_{n_h}, \xi_{n_h})
\end{bmatrix} \otimes C
\]

where \(A \otimes B\) denotes the Kronecker product between two matrices \(A\) and \(B\)

\[
A_{p_1q_1} \otimes B_{rs} = \begin{bmatrix}
a_{11}B_{rs} & a_{12}B_{rs} & \ldots & a_{1q}B_{rs} \\
a_{21}B_{rs} & a_{22}B_{rs} & \ldots & a_{2q}B_{rs} \\
\vdots & \vdots & \ddots & \vdots \\
a_{p1}B_{rs} & a_{p2}B_{rs} & \ldots & a_{pq}B_{rs}
\end{bmatrix}
\]

and \(\text{R}(\xi_i, \xi_j) \in \mathbb{R}^{n_v \times n_v}\) is the correlation matrix between \(\xi_i\) and \(\xi_j\); \(C\) is the cross-covariance matrix between various soil property parameters.
\[
C = \begin{bmatrix}
\sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,m}\sigma_1\sigma_m \\
\rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{m-1,m}\sigma_{m-1}\sigma_m \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{1,m}\sigma_1\sigma_m & \rho_{m-1,m}\sigma_{m-1}\sigma_m & \cdots & \sigma_m^2
\end{bmatrix}
\]

(3.)

where \(\sigma_i\) is the standard deviation of the \(i\)-th soil property parameter; \(\rho_{ij}\) 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.,

\[
\rho(\Delta x, \Delta y, \Delta z) = \rho_x(\Delta x, \Delta y) \times \rho_v(\Delta z)
\]

(4.)

where \(\rho(\Delta x, \Delta y, \Delta z)\) represents the 3D auto-correlation function; \(\rho_x(\Delta x, \Delta y)\) 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:

\[
\text{Var}(\xi) = [R_{h,\xi} \otimes R_{v,\xi}] \otimes C
\]

(5.)

where \(R_{h,\xi} \in \mathbb{R}^{nh \times nh}\) and \(R_{v,\xi} \in \mathbb{R}^{nv \times nv}\) denote the horizontal and vertical correlation matrices for \(\xi\), respectively:

\[
R_{h,\xi} = \begin{bmatrix}
1 & \rho_h(x_1-x_2, y_1-y_2) & \rho_h(x_1-x_3, y_1-y_3) & \cdots & \rho_h(x_1-x_\bar{n}, y_1-y_\bar{n}) \\
1 & \rho_h(x_2-x_3, y_2-y_3) & \cdots & \rho_h(x_2-x_\bar{n}, y_2-y_\bar{n}) \\
\vdots & \vdots & \ddots & \vdots \\
1 & \rho_h(x_{\bar{n}}-x_\bar{n}, y_{\bar{n}}-y_\bar{n}) & \cdots & 1
\end{bmatrix}
\]

(6.)

\[
R_{v,\xi} = \begin{bmatrix}
1 & \rho_v(z_1-z_2) & \rho_v(z_1-z_3) & \cdots & \rho_v(z_1-z_\bar{n}) \\
1 & \rho_v(z_2-z_3) & \cdots & \rho_v(z_2-z_\bar{n}) \\
\vdots & \vdots & \ddots & \vdots \\
1 & \rho_v(z_{\bar{n}}-z_\bar{n}) & \cdots & 1
\end{bmatrix}
\]

(7.)

\((x_i, y_i)\) are the x- and y-coordinates for the \(i\)-th sounding (\(i = 1, 2, \ldots, n_h\)), 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 \( \xi' \) 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 (\( n'_h = 5 \times 5 = 25 \) 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 \( (n_h, n_v) \).

\( \xi' \) can be represented as an \( (N' \times 1) \) column vector by stacking the data at each column of RFEs, i.e., \( \xi' = [\xi'_1^T \xi'_2^T \ldots \xi'_{n'_h}^T]^T \), where \( \xi'_i \) is an \( ((m \times n'_v) \times 1) \) column vector that contains the \( \xi \) values for the RFEs at the i-th column in Figure 1. Let us denote the covariance matrix for \( \xi' \) by \( \text{Var}(\xi') \in \mathbb{R}^{N' \times N'}. \) Due to the separable assumption of auto-correlation in the horizontal and vertical directions, \( \text{Var}(\xi') \) can be expressed as the following Kronecker product:

\[
\text{Var}(\xi') = \left[ R_{h,\xi'} \otimes R_{v,\xi'} \right] \otimes C
\] (8.)

where \( R_{h,\xi'} \in \mathbb{R}^{n'_h \times n'h} \) and \( R_{v,\xi'} \in \mathbb{R}^{n'_v \times n'v} \) denote the horizontal and vertical correlation matrices of \( \xi' \), respectively. \( R_{h,\xi} \) and \( R_{v,\xi} \) 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 (\( i = 1, 2, \ldots, n'_h \)), and \( z'_j \) is the z-coordinate for the j-th element at each column (\( j = 1, 2, \ldots, n'_v \)).

2.4. Covariance between soundings and RFEs

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

\[
\text{Cov}(\xi', \xi) = \left[ R_{h,\xi} \otimes R_{v,\xi} \right] \otimes C
\] (9.)

where \( R_{h,\xi} \in \mathbb{R}^{n'h \times n'h} \) and \( R_{v,\xi} \in \mathbb{R}^{n'v \times n'v} \) denote the horizontal and vertical correlation matrices between \( \xi' \) and \( \xi \), respectively:
Simulating the multivariate cross-correlated 3D conditional random field is equivalent to sample $\xi'$ conditioning on the observed sounding data $\xi^o$. Suppose that the mean vector of the $m$ number of variables contained in $\xi$ is $\mu_s = [\mu_1 \mu_2 \ldots \mu_m]^T$. Let $\mu$ denote the mean vector of $\xi$, including the mean values $\mu^o$ for $\xi^o$ and unconditional mean values $\mu^u$ for $\xi^u$. $\mu$ is formed by stacking $n$ number of $\mu_s$ vertically into a new column vector. Let $\mu'$ denote the unconditional mean vector of $\xi'$ by stacking $n'$ number of $\mu_s$ vertically. It is noted that the $\xi^o$ and $\xi'$ 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:

$$E(\xi'|\xi^o) = E(\xi^o) + \text{Cov}(\xi',\xi^o) \text{Var}(\xi^o)^{-1} [\xi^o - E(\xi^o)]$$

$$= \mu' + \text{Cov}(\xi',\xi^o) \text{Var}(\xi^o)^{-1} [\xi^o - \mu^o]$$

(12.)

And

$$\text{Var}(\xi'|\xi^o) = \text{Var}(\xi^o) - \text{Cov}(\xi',\xi^o) \text{Var}(\xi^o)^{-1} \text{Cov}(\xi',\xi^o)^T$$

(13.)

where $\text{Cov}(\xi',\xi^o) \in \mathbb{R}^{N' \times no}$ denotes the covariance matrix between $\xi'$ and $\xi^o$; $\text{Var}(\xi^o) \in \mathbb{R}^{N \times N}$ is the covariance matrix of $\xi^o$; $\text{Var}(\xi^o) \in \mathbb{R}^{no \times no}$ is the covariance matrices of $\xi^o$. The brute-force method can simulate the conditional sample of $\xi'$ directly using the Cholesky decomposition method when $E(\xi'|\xi^o)$ and $\text{Var}(\xi'|\xi^o)$ are computed. To be specific, $\xi' = \mu' + LZ$ where $LL^T = \text{Var}(\xi'|\xi^o)$ and $Z$ is an
independent standard normal vector with size of \( N' \). The simulation of \( \xi' \) require the Cholesky decomposition of \( \text{Var}(\xi'|\xi^o) \), 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 \( \xi' \) (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 \( \xi' \) base on the assumption of separability between horizontal and vertical auto-correlations. The method mainly contains two steps:

1. Simulate \( \xi^u \) by conditioning on \( \xi^o \). 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 \( \xi' \) based on the make-complete \( \xi \) (the union of \( \xi^o \) and the simulated \( \xi^u \) at the last step). The \( \xi' \) can be efficiently simulated with the following formula (Yang and Ching 2020; Ching et al. 2020):

\[
\xi' = \text{vec}[\mathbf{R}_{v,\xi} \cdot \mathbf{R}_{v,\xi}^{-1} \cdot \text{mat}(\xi) \cdot \mathbf{R}_{h,\xi}^{-1} \cdot \mathbf{R}_{h,\xi}^\top] + \sigma \cdot \text{vec}[(\mathbf{L}_{v,\xi}^\top \cdot \text{mat}(\mathbf{Z}) \cdot \mathbf{L}_{h,\xi}^\top)]
\]

(14.)

where \( \text{mat}(\mathbf{Z}) \) is an \((n' \times n')\) matrix containing independent standard normal random variables; \( \text{mat}(\xi) = [\xi_1 \xi_2 \ldots \xi_{nh}] \) is an \((n' \times nh)\) matrix containing all sounding data; \( \text{vec}[\mathbf{A}] \) 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 \( (\mathbf{R}_{h,\xi} - \mathbf{R}_{h,\xi}^{-1} \mathbf{R}_{h,\xi} \mathbf{R}_{h,\xi}^\top) \) and \( \mathbf{R}_{v,\xi} \), respectively.

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

Recall that $\xi^u$ and $\xi^o$ follows a multivariate normal distribution. According to the multivariate normal distribution theory, the conditional PDF $f(\xi^u|\xi^o)$ is still a multivariate normal distribution with the following mean vector $E(\xi^u|\xi^o)$ and covariance matrix $\text{Var}(\xi^u|\xi^o)$. The $E(\xi^u|\xi^o)$ and $\text{Var}(\xi^u|\xi^o)$ can be computed by Eqs. (12) and (13) by replacing $\mu'$ and $\xi'$ with $\mu^u$ and $\xi^u$, where $\mu^u$ is the mean vector of $\xi^u$ with size of $n_u \times 1$. Once $E(\xi^u|\xi^o)$ and $\text{Var}(\xi^u|\xi^o)$ are computed, the conditional sample of $\xi^u$ can be readily simulated using the Cholesky decomposition method. However, both $\text{Var}(\xi^o)$ and $\text{Var}(\xi^u|\xi^o)$ can be large matrices for a 3D problem especially when there are many explored soundings. The calculation of both $\text{Var}(\xi^o)^{-1}$ and $\text{Var}(\xi^u|\xi^o)$ 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 $\xi^u$.

4.1.1 Sounding-wise GS

The sounding-wise GS simulates the un-observed data at each sounding sequentially. Recall that the sounding data $\xi = [\xi_1^T \xi_2^T \ldots \xi_{nh}^T]^T$ where $\xi_i$ denotes the complete data for the $i$-th sounding. $\xi$ is the union of observed sounding data $\xi^o$ and un-observed sounding data $\xi^u$. Let $\xi_i^o$ and $\xi_i^u$ denote the observed and unobserved parts in $\xi_i$, respectively. Let $\xi_{-i} = [\xi_1^T \ldots \xi_{i-1}^T \xi_{i+1}^T \ldots \xi_{nh}^T]^T$ contains all data except the $i$-th sounding.

4.1.1.1. Implementation procedures

The sounding wise GS starts with assigning $\xi^u$ with any values (e.g., mean values) so that all open markers in Figure 1 are filled with these assigned values (i.e., $\xi$ becomes complete with a lattice
structure). Its implementation procedure is summarized as follows:

1. (Simulate the 1st $\xi^u$ sample) Do the sounding-wise sampling successively for soundings #1, #2, …, and #nh:
   a. Draw $\xi^u_1$ from $f(\xi^u_1|\xi^o, \xi_-)$ and update the $\xi^u$ values in $\xi$ with the sampled values (i.e., these $\xi^u$ values are used for simulating the $\xi^u$ at the subsequent step).
   b. Draw $\xi^u_2, \ldots, \xi^u_n$, in a way similar to that of $\xi^u_1$. In other words, draw $\xi^u_i$ from $f(\xi^u_i|\xi^o, \xi_-)$ and update the $\xi^u$ values in $\xi$ with the sampled values (i.e., these $\xi^u$ values are used for simulating the $\xi_{i-1}^u$).
   c. Combine the samples ($\xi^u_1, \xi^u_2, \ldots, \xi^u_n$) to form the 1st sample of $\xi^u$. This sample is used for simulating the next $\xi^u$ sample.

2. Repeat step (1) for $t = 2, \ldots, T$ to obtain another (T-1) samples of $\xi^u$. The initial $\xi^u$ samples should be discarded because they cannot represent the target PDF $f(\xi^u|\xi^o)$. The nearby samples are correlated, so the $\xi^u$ samples should be re-sampled by a suitable interval to reduce this correlation.

Suppose that there are NGS samples of $\xi^u$ (i = 1, 2, …, NGS) after discarding the preliminary samples and re-sampling.

4.1.1.2. Simulating $\xi^u_k$ from $f(\xi^u_k|\xi^o, \xi_k)$

The simulation of $\xi^u_k \sim f(\xi^u_k|\xi^o, \xi_k)$ is a key step of the sounding-wise GS. To avoid the large matrix issues, the sounding-wise GS decomposes this key step into two sub-steps: (a) compute the $f(\xi_k|\xi_k)$ and (b) compute the $f(\xi^u_k|\xi^o, \xi_k)$. Figure 2 shows the schematic diagram of the two sub-steps. Because $\xi$ is multivariate normal, $f(\xi_k|\xi_k)$ is a multivariate normal PDF with the following mean vector and covariance matrix:

$$E(\xi_k|\xi_k) = \mu_k + \text{Cov}(\xi_k, \xi_k)\text{Var}(\xi_k)^{-1}(\xi_k - \mu_k)$$

(15.)

$$\text{Var}(\xi_k|\xi_k) = \text{Var}(\xi_k) - \text{Cov}(\xi_k, \xi_k)\text{Var}(\xi_k)^{-1}\text{Cov}(\xi_k, \xi_k)^T$$

(16.)

where $\mu_k$ and $\mu_{-k}$ are the unconditional mean vectors for $\xi_k$ and $\xi_{-k}$. They are the sub-vectors in $\mu$. 
Because $\xi_k$ is filled with assigning/simulated values (the first sample) or simulated values (after the first sample), $\xi_k$ satisfies the 3D lattice structure. Based on the separable auto-correlation function, the covariance matrices for $\xi_k$ and $\xi_k$ allow the following Kronecker product representations:

\[
\text{Var}(\bar{\xi}_k) = (R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C) \text{ Var}(\bar{\xi}_k) = R_{v,\xi_k} \otimes C \quad \text{Cov}(\bar{\xi}_k, \bar{\xi}_k) = (R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C) \quad (17.)
\]

where $R_{h,\xi_k} \in \mathbb{R}^{(nh-1) \times (nh-1)}$ denotes the horizontal auto-correlation matrix for $\xi_k$ and is the sub-matrix of $R_{h,\xi_k}$ by excluding the $k$-th row and $k$-th column; $R_{h,\xi_k} \in \mathbb{R}^{1 \times (nh-1)}$ denotes the horizontal auto-correlation vector between $\xi_k$ and $\xi_k$.

In Eq. (17), the vertical auto-correlation matrix for $\xi_k$ and $\xi_k$ is the same (i.e., $R_{v,\xi_k} \in \mathbb{R}^{nv \times nv}$) because $\xi_k$ and $\xi_k$ are with same sounding depths $(z_1, z_2, ..., z_m)$. Considering the following matrix identities:

\[
(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}, \quad (A \otimes B)^T = A^T \otimes B^T, \quad (A^T \otimes B) \times \text{vec}(C) = \text{vec}(BCA)
\]

and inserting Eq. (17) into Eqs. (15) and (16) yields:

\[
\begin{align*}
\text{E}(\xi_k|\xi_k) &= \mu_k + \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot (\xi_k - \mu_k) \\
&= \mu_k + \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot (\xi_k - \mu_k) \\
&= \left[ \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \right] \cdot (\xi_k - \mu_k) \\
&= \mu_k + \text{vec} \left[ \left( I_v \otimes I_h \right) \cdot \text{mat} (\xi_k - \mu_k) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \right] \\
&= \mu_k + \text{vec} \left[ \text{mat} (\xi_k - \mu_k) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \right] \\
\text{Var}(\xi_k|\xi_k) &= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^T \\
&= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot R_{v,\xi_k} \otimes C \\
&= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \otimes R_{v,\xi_k} \otimes C \\
\end{align*}
\]

where $I_v$ denotes the identity matrix with size of $nv \times nv$; $I_h$ denote the identity matrix with size of $m \times m$; $\text{mat}(\xi_k - \mu_k) = [\bar{\xi}_1 - \mu_1 \ldots \bar{\xi}_{k-1} - \mu_{k-1} \bar{\xi}_k - \mu_k \ldots \bar{\xi}_{nh} - \mu_{nh}]$, an $(m \times nv) \times (nh-1)$ matrix. Note that the term $\text{vec}(BCA)$ and inserting Eq. (17) into Eqs. (15) and (16) yields:

\[
\begin{align*}
\text{Var}(\xi_k|\xi_k) &= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^T \\
&= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \cdot \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right)^{-1} \cdot R_{v,\xi_k} \otimes C \\
&= R_{v,\xi_k} \otimes C - \left( R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes R_{h,\xi_k} \otimes R_{v,\xi_k} \otimes C \right) \otimes R_{v,\xi_k} \otimes C \\
\end{align*}
\]
Because the $\xi_k$ follows a multivariate normal distribution with the mean vector of $E(\xi_k|\xi_{-k})$ and covariance of $\text{Var}(\xi_k|\xi_{-k})$, the PDF $f(\xi_k|\xi_{-k}, \xi_{-k})$ is also a multivariate normal PDF with the following mean vector and covariance matrix:

$$E\left(\xi_k|\xi_{-k}\right) = E\left(\xi_k|\xi_{-k}\right) + \text{Cov}\left(\xi_k|\xi_{-k}\right) \text{Var}\left(\xi_k|\xi_{-k}\right)^{-1} \left[\xi_k - E\left(\xi_k|\xi_{-k}\right)\right]$$

(22.)

$$\text{Var}\left(\xi_k|\xi_{-k}\right) = \text{Var}\left(\xi_k|\xi_{-k}\right) - \text{Cov}\left(\xi_k|\xi_{-k}\right) \text{Var}\left(\xi_k|\xi_{-k}\right)^{-1} \text{Cov}\left(\xi_k|\xi_{-k}\right)^T$$

(23.)

where $E(\xi_k|\xi_{-k})$ and $E(\xi_k|\xi_{-k})$ are sub-vectors of $E(\xi_k|\xi_{-k})$, and $\text{Var}(\xi_k|\xi_{-k})$, $\text{Var}(\xi_k|\xi_{-k})$, and $\text{Cov}(\xi_k, \xi_k)$ are sub-matrices of $\text{Var}(\xi_k|\xi_{-k})$. These sub-vectors and sub-matrices can be extracted from $E(\xi_k|\xi_{-k})$ and $\text{Var}(\xi_k|\xi_{-k})$ directly. The evaluation of $E(\xi_k|\xi_{-k})$ and $\text{Var}(\xi_k|\xi_{-k})$ in Eqs. (19) and (21) does not require inversion and storage of larger matrices. The simulation of $\xi_k$ sample with Eqs. (22) and (23) also does not need to make Cholesky decomposition of larger matrices.

### 4.1.2 Depth-wise MCS

Let $\xi_{zi} \in \mathbb{R}^{m \times n}$ contain all sounding data at the $zi$ depth. It is evident that $\xi_{zi}$ is a $m \times n$ column vector since the data at each depth at each sounding contain $m$ number of variables. Let $\xi_{zi}^o$ and $\xi_{zi}^u$ denote the observed and unobserved parts, respectively, in $\xi_{zi}$. The depth-wise MCS considers that $\xi_{zi}$ depends on $\xi_z$ only if $z$ is within the depth range of $[zi-\alpha \delta_v, zi+\alpha \delta_v]$, where $\delta_v$ denotes the vertical SOF and $\alpha$ 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 $\alpha$ 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 $\alpha$ 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 $\xi^u$ sample) Do the depth-wise sampling for depths $z_1, z_2, \ldots, \text{and } z_n$: 

a. Simulate $\xi_{z1}^u \sim f(\xi_{z1}^u | \xi_{z1}^o, \ldots, \xi_{z_{\max}}^o)$, where $z_{\max}$ denotes the maximum conditioning depth

($z_{\max}$ is either $z_1 + \alpha \delta v$ or $z_{nv}$; the smaller one is taken)

b. Successively simulate $\xi_{z2}^u, \ldots, \xi_{z_{nv}}^u$ in a way similar to that of $\xi_{z1}^u$. In other words, Simulate

$\xi_{z2}^u \sim f(\xi_{z2}^u | \xi_{z_{\min}}^o, \ldots, \xi_{z_{\max}}^o, \xi_{z_{\min}}^u, \ldots, \xi_{z_{1-1}}^u)$, where $z_{\min}$ denotes the minimum conditioning depth ($z_{\min}$ is either $z_1 - \alpha \delta v$ or $z_1$; the larger one is taken). Here, as shown in Figure 3, ($\xi_{z_{\min}}^u \ldots \xi_{z_{1-1}}^u$) take the values that were simulated in previous steps.

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

(2) Repeat step (1) for $t = 2, \ldots, N_{MC}$ to obtain $N_{MC}$ samples of $\xi^u$. 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 $\xi$ are normal variables, $f(\xi_{z1}^u | \xi_{z1}^o, \ldots, \xi_{z_{\max}}^o)$, $f(\xi_{z2}^u | \xi_{z_{\min}}^o, \ldots, \xi_{z_{\max}}^o, \xi_{z_{\min}}^u, \ldots, \xi_{z_{1-1}}^u)$ and $f(\xi_{z_{nv}}^u | \xi_{z_{\min}}^o, \ldots, \xi_{z_{nv}}^o, \xi_{z_{\min}}^u, \ldots, \xi_{z_{nv-1}}^u)$ also follow multivariate normal distributions. Their mean vectors and covariance matrices can be estimated using Eqs (12) and (13).

4.2. Step 2 – simulating $\xi'$ conditioning on $\xi$

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

$$E(\xi' \mid \xi) = \mu' + Cov(\xi', \xi) Var(\xi)^{-1} (\xi - \mu)$$  \hspace{1cm} (24.)

$$Var(\xi' \mid \xi) = Var(\xi') - Cov(\xi', \xi) Var(\xi)^{-1} Cov(\xi', \xi)^T$$  \hspace{1cm} (25.)

Note that $Cov(\xi', \xi) = [R_{h,\xi' \xi} \otimes R_{v,\xi' \xi} \otimes C]$, $Var(\xi) = [R_{h,\xi} \otimes R_{v,\xi} \otimes C]$, and $Var(\xi') = [R_{h,\xi'} \otimes R_{v,\xi'} \otimes C]$. Considering the matrix identities $(A \otimes B)^T = A^T \otimes B^T$, $(A \otimes B)^T = A^T \otimes B^T$, and $(A^T \otimes B) \cdot \text{vec}(C) = \text{vec}(BCA)$, Eqs. (24) and (25) can be further expressed as
\[
E(\xi'|\xi) = \mu' + \left( R_{h,\xi'} \otimes R_{v,\xi'} \otimes C \right) \left( R_{h,\xi} \otimes R_{v,\xi} \otimes C \right)^{-1} (\xi - \mu)
\]
\[
= \mu' + \left[ \left( R_{h,\xi} R_{h,\xi}^{-1} \right) \otimes \left( R_{v,\xi} R_{v,\xi}^{-1} \otimes I_s \right) \right] (\xi - \mu) = \text{vec} \left[ \left( R_{v,\xi} R_{v,\xi}^{-1} \otimes I_s \right) \text{mat}(\xi - \mu) R_{h,\xi}^{-1} R_{h,\xi}^T \right]
\]

(26.)

\[
\text{Var}(\xi'|\xi) = \left[ R_{h,\xi'} \otimes R_{v,\xi'} \otimes C - \left( R_{h,\xi} \otimes R_{h,\xi}^{-1} \otimes C \right) \left( R_{h,\xi} \otimes R_{v,\xi} \otimes C \right)^{-1} \left( R_{h,\xi} \otimes R_{v,\xi} \otimes C \right)^T \right]
\]
\[
= \left[ R_{h,\xi'} \otimes R_{v,\xi'} \otimes C - \left[ R_{h,\xi} R_{h,\xi}^{-1} R_{h,\xi}^T \otimes R_{v,\xi} R_{v,\xi}^{-1} R_{v,\xi}^T \right] \right] \otimes C
\]

(27.)

where \( \text{mat}(\xi) = [\xi_1 \xi_2 \ldots \xi_{nh}] \), an \((m \times n_v) \times nh\) matrix; \(\text{vec}[A]\) is a vector formed by stacking the columns in \(A\) vertically; \(I_s\) is a \(m \times m\) identity matrix. Ching et al. (2020) found that if the “densely covers” conditions hold, it has \(R_{v,\xi'} \approx R_{v,\xi} R_{v,\xi}^{-1} R_{v,\xi}^T\). The “densely covers” condition means that the sampling intervals for the sounding data \((z_1, z_2, \ldots, z_{nv})\) should significantly smaller than the vertical SOF and the sounding depth range \([z_1, z_{nv}]\) should cover the random field element depth range \([z_1', z_{nv}']\). Yang and Ching (2020) further pointed out that the “densely covers” conditions can be satisfied by inserting additional un-observed depths into the sample intervals for the sounding data. The sounding data at the additional un-observed depths can be simulated at Step 1 so that the sounding data is make-complete and satisfies the “densely covers” conditions. When the make-complete sounding data \(\xi\) are used for simulating \(\xi'\), Eq. (27) reduces to the following equation

\[
\text{Var}(\xi'|\xi) = \left( R_{h,\xi'} - R_{h,\xi} R_{h,\xi}^{-1} R_{h,\xi'} \otimes C \right)
\]

(28.)

Let \(L_{h,\xi'}\), \(L_{v,\xi'}\) and \(L_c\) be the lower triangular Cholesky decompositions of \([R_{h,\xi'} - R_{h,\xi} R_{h,\xi}^{-1} R_{h,\xi'}]\) and \(R_{v,\xi'}\), respectively. The lower triangular Cholesky decomposition of \([R_{h,\xi'} - R_{h,\xi} R_{h,\xi}^{-1} R_{h,\xi'}]\) \(\otimes\) \(R_{v,\xi'} \otimes C\) is equal to \(L_{h,\xi'} \otimes L_{v,\xi'} \otimes L_c\). With the Eqs. (26) and (28), a conditional sample of \(\xi'\) can then be simulated by the following equation:

\[
\xi' = \text{vec} \left[ \left( R_{v,\xi'} R_{v,\xi}^{-1} \otimes I_s \right) \text{mat}(\xi - \mu) R_{h,\xi}^{-1} R_{h,\xi'}^T \right] + \text{vec} \left[ \left( L_{h,\xi'} \otimes L_{v,\xi'} \otimes L_c \right) \text{vec}(Z) \right]
\]

\[
= \text{vec} \left[ \left( R_{v,\xi'} R_{v,\xi}^{-1} \otimes I_s \right) \text{mat}(\xi - \mu) R_{h,\xi}^{-1} R_{h,\xi'}^T \right] + \text{vec} \left[ \left( L_{v,\xi'} \otimes L_c \right) Z L_{h,\xi'}^T \right]
\]

(29.)

where \(\text{vec}(Z)\) is an \((m \times n_v') \times n_h'\) matrix containing independent standard normal variables.

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 \( N' = 2601 \times 151 \) = 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; \( \sigma_{v} \) is vertical total stress; and \( \sigma'_{v} \) is vertical
effective stress. As a result, \( \xi' \) containing the values at the centroids of the 392,751 RFEs is a column
vector with \( 392,751 \times 3 = 1,178,253 \) 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
\[
Y = \sinh \left[ \frac{X - b_X}{a_X} \right] a_Y + b_Y
\]

(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 \(\delta_h = 20\) m and \(\delta_v = 1\) m:

\[
\rho_h(\Delta x, \Delta y) = \exp \left( -2\sqrt{(\Delta x)^2 + (\Delta y)^2}/\delta_h \right)
\]

(31.)

\[
\rho_v(\Delta z) = \exp \left( -2\Delta z/\delta_v \right)
\]

(32.)

Suppose that there are five explored soundings \(n_b = 5\) at this virtual site. The locations of the five soundings are denoted as \(S_1\) - \(S_5\) 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 \(S_1\) - \(S_5\). 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 \(S_1\) - \(S_4\). There is no sampling at \(S_5\) for \(B_q\). The sampling depth at \(S_1\) and \(S_2\) is 15 m, whereas it is 10 m at \(S_3\) and \(S_4\). As a result, there is \(2 \times 16 + 2 \times 10 = 52\) number of observed \(B_q\) data at \(S_1\) - \(S_5\).
2. The vertical sampling interval of ln(Qt) 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(Qt) data at S1-S5.

3. The vertical sampling interval of ln(Qe) is 0.2 m at S1, S3, S4 and S5. There is no sampling at S2 for ln(Qe). 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(Qe) data at S1-S5.

In total, the virtual site contains 52+114+197 = 363 observed Bq, ln(Qt) and ln(Qe) data. It is noted that the CPTu data (i.e., Bq, ln(Qt) and ln(Qe)) 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 Bq, Qt and Qe at various soundings for better demonstration.

The adoption of various sampling intervals for Bq, Qt and Qe is to model the possible practical site investigation where other soil data such as cohesion c, friction angle $\phi$ and elastic modulus E measured at different depth intervals at S1-S5 are also of interest. The observed Bq, ln(Qt) and ln(Qe) data at the virtual site are then converted to the zero-mean normal random data $\xi^o$ by the inversion of Eq. (30). The purpose of this simulated example is to simulate the data $\xi'$ at the centroids of the 392,751 RFEs ($\xi' \in \mathbb{R}^{1,178,253 \times 1}$) by conditioning on the observed sounding data ($\xi^o \in \mathbb{R}^{363 \times 1}$). The simulated $\xi'$ values for the RFEs can be converted to Bq, ln(Qt) and ln(Qe) values by Eq. (30). The un-observed data ($\xi^u$) 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 Bq, ln(Qt) and ln(Qe) 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 $\xi^u$ conditioning on $\xi^o$

It is clear that the observed data $\xi^o$ do not satisfy the dense-cover condition: (a) the sampling interval is not always significantly smaller than $\delta_v$ (e.g., the sampling interval for $B_q$ at S1-S4 is 1 m, whereas $\delta_v$ is also 1 m), and (b) the sounding depth range $[z_1, z_{nv}]$ does not cover the RFE’s depth range $[z'_1, z'_nv]$ (e.g., the maximum sounding depth for $\ln(Q_h)$ at S4 is 4 m, far less than the maximum RFE’s depth 15 m). It is also clear that the observed data $\xi^o$ do not have a 3D lattice structure: all the five soundings do not share the same observed depths. The proposed method requires $\xi$ to fulfill the above condition. This is achieved by augmenting the observed depths for $\xi^o$ by the un-observed depths for $\xi^u$. 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 $\xi_i$ denote the augmented sounding data at the $i$-th sounding. It is clear that $\xi_1, \xi_2, \xi_3, \xi_4$ and $\xi_5$ are all $(151 \times 3) \times 1$ 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 $\xi$ is a $(5 \times 151 \times 3) \times 1$ column vector. $\xi$ contain 363 observed data (i.e., $\xi^o$) and 1902 unobserved data (i.e., $\xi^u$).

The sounding-wise GS first simulate $T = 4 \times 10^4 + 1000$ samples of $\xi^u$. The initial 1000 samples are discarded because they cannot represent the target distribution $f(\xi^u|\xi^o)$. To mitigate the correlation issues between the adjacent samples, the remaining $4 \times 10^4$ samples are re-sampled with an interval of 4 to obtained $N_{GS} = 1 \times 10^4$ samples of $\xi^u$. 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 $N_{MCS} = 1 \times 10^4$ samples of $\xi^u$ directly. When the $\xi^u$ samples are simulated, they are converted to $B_q$, $\ln(Q_h)$ and $\ln(Q_e)$ samples by Eq. (30). Based on the $1 \times 10^4$ simulated samples of $B_q$, $\ln(Q_h)$ 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 $\xi'$ replaced by $\xi^u$. 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 $B_4$ 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 $\xi^u$ value at the first un-observed depth at a sounding and the $\xi^u$ value at the remaining un-observed depths at this sounding. The correlations can be estimated by the $1 \times 10^4$ simulated samples of $\xi^u$ 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 $1 \times 10^4$ simulated samples of $\xi^u$, 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 $\xi'$ conditioning on $\xi^o$

It is indicated in Step 1 that both the sounding-wise GS and the depth-wise MCS can accurately simulate the $\xi^u$ conditioning on $\xi^o$. Therefore, this section only presents the Step 2 results based on the $1 \times 10^4$ samples of $\xi^u$ simulated by the sounding-wise GS. In Step 2, each $\xi^u$ sample simulated by the sounding-wise GS is used to simulate a $\xi'$ sample. As a result, Step 2 finally obtains $1 \times 10^4$ samples of $\xi'$. Similar to Step 1, $\xi'$ 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 $1 \times 10^4$ 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
interesting to observe how the width of the 95% CI depends on the distance to a nearby sounding. The 95% CI in Figure 11 for $B_q$ and $\ln(Q_l)$ at T1 and T5 is relatively narrow compared with that at T2-T4. This is because T1 and T5 are very close to the explored soundings S1 and S2, respectively. The uncertainty at T1 and T5 is effectively reduced by the nearby observed data. In contrast, such uncertainty reduction decreases as the un-explored locations are farther away from the explored soundings. This can be further observed from Figure 12. Figure 12 shows the mean and standard deviations of the simulated $B_q$, $\ln(Q_l)$ and $\ln(Q_e)$ at A-A section (see Figure 4). The mean values of the simulated data at the observed locations are well constrained to the observed data. The uncertainty of the simulated data reduces significantly at the adjacent locations of the observed locations. This indicates that the extended method can properly simulate the multivariate 3D cross-correlated conditional random fields.

5.4. Comparison of computational efficiency among different methods

Table 1 shows the comparison results of the proposed method and the brute-force method. Compared with the brute-force method, both sounding-wise GS and depth-wise MCS can reduce the size of the matrices that need to be inverted or decomposed with Choleskey. For instance, the brute-force method requires the inversion of the covariance matrix $\text{Var}(\xi^o)$ (with a size of $363 \times 363$ for the investigated example) and Cholesky decomposition of the covariance matrix $\text{Var}(\xi^i|\xi^o)$ (with a size up to $1177890 \times 1177890$). Nevertheless, the sounding-wise GS only requires five evaluations of the inversions for $R_{h,\xi-1}$, $R_{h,\xi-2}$, ..., $R_{h,\xi-nh}$ (with a size of $(n_h-1) \times (n_h-1)=4 \times 4$) and $\text{Var}(\xi^i|\xi^o)$, $\text{Var}(\xi^1|\xi^o)$, ..., $\text{Var}(\xi^n|\xi^o)$ (with a size up to $[m \times n_v] \times [m \times n_v] = 123 \times 123$). In addition, five Cholesky decompositions need to be performed for covariance matrix $\text{Var}(\xi^i|\xi^o, \xi^k)$ (with a size up to $[m \times n_v] \times [m \times n_v] = 409 \times 409$). The depth-wise MCS requires 151 evaluations for the inversion of the covariance matrix $\text{Var}(\xi, ..., \xi^{max}, \xi^{min}, ..., \xi^{zi-1})$ (with a size up to $369 \times 369$ for the investigated example) and Cholesky decomposition of the covariance matrix $\text{Var}(\xi^{hi}|\xi^{min}, ...,\xi^{max}, \xi^{min}, ..., \xi^{zi-1})$ (with a size up to $15 \times 15$). Step 2 of the proposed method requires one
evaluation of the inversion of the covariance matrix $R_{v,\xi}$ (with a size up to $151 \times 151$) and Cholesky decomposition of the covariance matrix $R_{v,\xi'}$ (with a size up to $2601 \times 2601$). Table 1 also compared the computational time required by the proposed method and the brute-force method. The computation time is based on the MATLAB platform in a desktop computer with 64 GB RAM and Intel Core i9-10900K CPU clocked at 3.70GHz. To simulate $1 \times 10^4$ number of $\xi^u$ samples, the sounding-wise GS takes 1.87 minutes, which is favorably comparable with that of 0.98 minutes taken by the depth-wise MCS. Step 2 takes 10.99 minutes. In total, the proposed method takes about 12 minutes to generate $1 \times 10^4$ realizations of cross-correlated conditional random fields of $B_q$, ln($Q_t$) and ln($Q_e$). The brute-force method is not able to simulate the trivariate 3D cross-correlated conditional random fields because the storage space needed to operate the matrix $\text{Var}(\xi'|\xi)$ (with a size up to $1177890 \times 1177890$) for the brute-force method exceeds the memory of the used desk computer.

It is noted that, for the depth-wise MCS, $(\xi_{z\text{min}}^o, \ldots, \xi_{z\text{max}}^o, \xi_{z\text{min}}^u, \ldots, \xi_{z_{i-1}}^u)$ collectively does not have a lattice structure in general, because $\xi^u$ at depths of $z_{i+1}, \ldots, z_{\text{max}}$ has not yet been simulated when simulating the $\xi_{zi}^u$, hence it is in general not feasible to represent the covariance matrix $\text{Var}(\xi_{z\text{min}}^o, \ldots, \xi_{z\text{max}}^o, \xi_{z\text{min}}^u, \ldots, \xi_{z_{i-1}}^u)$ as a Kronecker product. The size of this covariance matrix is $(m \times n_h \times n_{\text{cd}}) \times (m \times n_h \times n_{\text{cd}})$, where $n_{\text{cd}}$ denotes the number of conditioning depths, i.e., the number of sounding depths within the range $[z_{i-\alpha \delta v}, z_{i+\alpha \delta v}]$. $n_{\text{cd}}$ is roughly equal to $2 \alpha \delta v/(\text{sounding depth interval})$. When $\delta v$ is large or $\alpha$ takes a very large value, the range $[z_{i-\alpha \delta v}, z_{i+\alpha \delta v}]$ might cover the whole depth range $[z_1, z_{\text{nv}}]$, the depth-wise MCS’s computational efficiency might be reduced since the size of the covariance matrix to be inverted becomes large. For instance, if $\alpha$ takes a value of 80, simulating the data at the $i$-th depth needs to condition on the data at all the depths. As a result, the matrices need to be inverted or make Cholesky decomposition larger. The advantage of depth-wise MCS is that it can generate independent samples. On the contrary, the sounding-wise GS method requires to use the Gibbs sampler to simulate the un-observed data at each sounding. When the 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 $\xi^{u}$ to make the sounding data complete and Step 2 simulates the conditional random fields based on the make-complete sounding data $\xi$. 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
Appendix A. Proof of reduction of cross-correlations in conditional random fields

Consider the data at the i-th depth and j-th for example. It is assumed that each depth contains two soil property parameters: a and b. Suppose that cross-covariance matrix between a and b is $\begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}$ and the auto-correlation matrix between the i-th depth and j-th depth is $[1 r_{ij}; r_{ji} 1]$.

The data at the two depths follow the following normal distribution:

$$f(x_i, x_h, x_i, x_h) = N \left( \begin{bmatrix} \mu_x, \mu_h, \mu_x, \mu_h \end{bmatrix} \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix} \right)$$ \hspace{1cm} (A1)

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

$$\text{Cov}(x_i, x_j | x_i, x_h) = \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \sigma_a^2 & \rho \sigma_a \sigma_b & r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b \\ \rho \sigma_a \sigma_b & \sigma_b^2 & r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a \\ r_i \sigma_a^2 & r_j \rho \sigma_a \sigma_b & \sigma_a^2 & r_i \rho \sigma_a \sigma_b \\ r_j \rho \sigma_b \sigma_a & r_i \rho \sigma_b \sigma_a & r_i \rho \sigma_a \sigma_b & \sigma_b^2 \end{bmatrix}^{-1}$$

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


34
methods. Structural safety, 30(4), 337-363.


Zhao, T., and Wang, Y. (2020). Non-parametric simulation of non-stationary non-gaussian 3D random field samples directly from sparse measurements using signal decomposition and Markov Chain


Table 1  Summary of the computational cost

<table>
<thead>
<tr>
<th>Proposed method</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Brute-force method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sounding-wise GS</td>
<td>Depth-wise MCS</td>
<td></td>
</tr>
<tr>
<td>Max. size for matrix inversions (# of evaluations)</td>
<td>123×123 (5)</td>
<td>369×369 (151)</td>
<td>151×151 (1)</td>
</tr>
<tr>
<td>Max. size for Cholesky decomposition (# of evaluations)</td>
<td>409×409 (5)</td>
<td>15×15 (151)</td>
<td>2601×2601 (1)</td>
</tr>
<tr>
<td>Max. size of stored matrices</td>
<td>453×453</td>
<td>369×369</td>
<td>2601×2601</td>
</tr>
<tr>
<td>Computational time</td>
<td>1.87a mins</td>
<td>0.98b mins</td>
<td>10.99c mins</td>
</tr>
</tbody>
</table>

Notes:

a Time for simulating $4 \times 10^4 + 1000$ GS samples of $\xi_u$

b Time for simulating $1 \times 10^4$ MCS samples of $\xi_u$

c Time for simulating $1 \times 10^6$ samples of $\xi'_u$

d Result is not available due to the large computational cost
Figure 1  Schematic diagram for sounding data at a typical site (6 soundings in total) and RFEs (5, 5 and 10 elements along the x-, y- and z-directions) generated for simulating spatial variability of soil properties $X_1$ and $X_2$ of this site.

Figure 2  Schematic diagram of sounding-wise GS for simulating the un-observed data at the k-th sounding (i.e., $\xi_k^u$).
Figure 3  Schematic diagram of depth-wise MCS for simulating the un-observed data at the i-th depth (i.e., $\xi_{ziu}$).

Figure 4  Top view for A-A cross section, soundings S1-S5, and selected un-explored location T1-T5.
Figure 5  Observed (filled marker) and un-observed (open marker) data at S1-S5

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

Figure 7  Estimation error for the auto-correlations of $X_1$ at S1-S2
Figure 8  Estimation results for the cross-correlations among X1, X2 and X3 at S1-S5

Figure 9  One typical realization of multivariate 3D cross-correlated conditional random fields of Bq, ln(Qt) and ln(Qe)

Figure 10  Estimated median and 95% CI profiles of Bq and ln(Qt) 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