Data-driven simulation of multivariate cross-correlated geotechnical random fields from sparse measurements

Zheng Guan[1#] and *Yu* Wang[2]

[1]*Research Assistant Professor, State Key Laboratory of Internet of Things for Smart City, Department of Civil and Environmental Engineering, University of Macau, Macao, China.*
[2]*Professor, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.*
[#]*Corresponding author: guanzheng@um.edu.mo*

ABSTRACT

It is widely acknowledged that many geotechnical properties are correlated over space and/or time. Consequently, cross-correlated random fields play a pivotal role in geotechnical reliability analysis for properly modeling both the auto- and cross-correlation structures of correlated geotechnical properties. Existing methods for simulating cross-correlated random fields typically require precise knowledge of random field parameters as input. However, in a typical site investigation program, engineering constraints such as limited time, budget, and space often lead to sparse measurements of geotechnical properties. Estimating reliable random field parameters, particularly the auto-correlation and cross-correlation structures of a two-dimensional (2D) cross-correlated random field, from such sparse data is a notorious challenge. To address this issue, this study introduces a 2D cross-correlated random field generator that can directly simulate 2D multivariate cross-correlated geotechnical random field samples (RFSs) from sparsely measured data points. This generator leverages the method developed by Guan and Wang (2023), which employs a joint sparse representation to simultaneously exploit auto- and cross-correlation structures of various spatial/temporal quantities directly from sparse measurements. The effectiveness of the proposed generator is demonstrated using real geotechnical properties data. The results demonstrate that RFSs generated using this method from sparse measurements accurately capture the spatial auto- and cross-correlation structures of different geotechnical properties.

Keywords: Random field; Cross-correlation; Sparse representation; Sparse data.

1. Introduction

It is widely recognized that many geotechnical properties display cross-correlation and spatial variability (e.g., Vanmarcke et al., 2010). Therefore, properly accounting for both autocorrelation and cross-correlation structures in correlated geotechnical properties is crucial for subsequent geotechnical reliability analysis. Cross-correlated random fields have emerged as indispensable tools for simultaneously incorporating both auto- and cross-correlation aspects of geotechnical properties in stochastic analysis.

Several cross-correlated random fields methods have been developed in the literature (e.g., Shinozuka and Deodatis, 1991; Robin et al., 1993; Vořechovský, 2008). Existing methods for simulating cross-correlated random fields typically necessitate explicit information about random field parameters (e.g., functional forms and parameters of auto- and cross-correlation structures) as inputs, which are often estimated using site-specific measurements. In engineering practice, geotechnical properties are frequently sparsely measured within a typical site due to constraints in time, budget, and space (e.g., Guan and Wang, 2020; Guan et al. 2023), and estimating reliable random field parameters, particularly the auto-correlation and cross-correlation structures of a two-dimensional (2D) cross-correlated geotechnical properties random field, from such data is a notorious challenge.

To tackle these challenges, this study introduces a novel 2D cross-correlated geotechnical random fields generator capable of producing 2D cross-correlated geotechnical random field samples (RFSs) directly from sparsely obtained measurements. This generator builds upon the method developed by Guan and Wang (2023), which utilizes joint sparse representation to concurrently exploit auto- and cross-correlation structures of various spatial/temporal quantities directly from sparse measurements in a data-driven manner. Explicit functional forms and parameters of auto- and cross-correlation structures are not required for the introduced random field generator. The efficacy of the proposed method is demonstrated using real geotechnical data from California, USA.

2. Data-driven cross-correlated geotechnical random fields generator

Geotechnical properties are usually spatially and correlated, lending themselves to a sparse representation in suitable bases, such as the discrete cosine transform (DCT) (e.g., Wang and Zhao, 2017; Guan and Wang, 2021). Mathematically, geotechnical properties data within a 2D space can be represented by an $N_{x_1} \times N_{x_2}$

data matrix $\mathbf{F}$, which can be expressed by a linear combination of $N = N_{x_1} \times N_{x_2}$ 2D basis functions:

$$\mathbf{F} = \sum_{k=1}^{N=N_{x_1}\times N_{x_2}} \mathbf{B}_k^{2D} \omega_k^{2D} \tag{1}$$

where $\mathbf{B}_k^{2D}$ and ω_k^{2D} are the k-th 2D basis function and corresponding weight coefficient, respectively. When $\mathbf{F}$ has a sparse representation, it implies that only a small number of coefficients are non-zero.

Generally speaking, geotechnical properties with strong positive cross-correlation tend to exhibit similar global spatial or temporal patterns, while each property may also possess unique spatial or temporal characteristics. Consequently, when different positively correlated geotechnical property datasets are linearly normalized to a common scale, the normalized data often share a common component. Moreover, each individual property dataset may display its own distinct component. Both the common and individual components can be sparsely represented and formulated within a joint sparse representation framework. Under this framework, the normalized 2D data of two positively correlated geotechnical properties, Q_1 and Q_2 denoted as $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$ with the same dimensions, can be jointly represented as (Guan and Wang, 2023):

$$\mathbf{F}'_{Q_1} = \sum_{k=1}^{N=N_{x_1}\times N_{x_2}} \mathbf{B}_k^{2D} \omega_k^{C} + \sum_{k=1}^{N=N_{x_1}\times N_{x_2}} \mathbf{B}_k^{2D} \omega_k^{U_1} \tag{2}$$

$$\mathbf{F}'_{Q_2} = \sum_{k=1}^{N=N_{x_1}\times N_{x_2}} \mathbf{B}_k^{2D} \omega_k^{C} + \sum_{k=1}^{N=N_{x_1}\times N_{x_2}} \mathbf{B}_k^{2D} \omega_k^{U_2} \tag{3}$$

where ω_k^C indicate the k-th weight coefficient of common component for $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$, while $\omega_k^{U_1}$ and $\omega_k^{U_2}$ indicate the k-th weight coefficients of individual component for $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$, respectively. Let column vectors $\boldsymbol{y}_1'$ and $\boldsymbol{y}_2'$ represent measurements of $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$, respectively, as expressed:

$$\boldsymbol{y}'_{Q_1} = \mathbf{A}_1 \boldsymbol{\omega}^C + \mathbf{A}_1 \boldsymbol{\omega}^{U_1} \tag{4}$$

$$\boldsymbol{y}'_{Q_2} = \mathbf{A}_2 \boldsymbol{\omega}^C + \mathbf{A}_2 \boldsymbol{\omega}^{U_2} \tag{5}$$

where measurement matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are developed according to the locations of measurements in $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$, respectively; $\boldsymbol{\omega}^C = \left[\omega_1^C, \omega_2^C, ..., \omega_N^C\right]^{\mathrm{T}}$ indicates the weight coefficient vector for common component; $\boldsymbol{\omega}^{U_1} = \left[\omega_1^{U_1}, \omega_2^{U_1}, ..., \omega_N^{U_1}\right]^{\mathrm{T}}$ and $\boldsymbol{\omega}^{U_2} = \left[\omega_1^{U_2}, \omega_2^{U_2}, ..., \omega_N^{U_2}\right]^{\mathrm{T}}$ are weight coefficient vectors of individual component for $\boldsymbol{y}'_{Q_1}$ and $\boldsymbol{y}'_{Q_2}$. Combining Eqs. (4) and (5) leads to:

$$\begin{bmatrix} \boldsymbol{y}'_{Q_1} \\ \boldsymbol{y}'_{Q_2} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\omega}^C \\ \boldsymbol{\omega}^{U_1} \\ \boldsymbol{\omega}^{U_2} \end{bmatrix} \tag{6}$$

Eq. (6) can be further re-written as:

$$\boldsymbol{y}^{\mathrm{e}} = \mathbf{A}^{\mathrm{e}} \boldsymbol{\omega}^{\mathrm{e}} \tag{7}$$

where $\boldsymbol{y}^{\mathrm{e}} = \begin{bmatrix} \boldsymbol{y}'_{Q_1} \\ \boldsymbol{y}'_{Q_2} \end{bmatrix}$; $\mathbf{A}^{\mathrm{e}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{A}_2 \end{bmatrix}$; $\boldsymbol{\omega}^{\mathrm{e}} = \begin{bmatrix} \boldsymbol{\omega}^C \\ \boldsymbol{\omega}^{U_1} \\ \boldsymbol{\omega}^{U_2} \end{bmatrix}$

.

Leveraging on the joint sparse representation, non-trivial weight coefficients in $\boldsymbol{\omega}^{\mathrm{e}}$ can be jointly estimated from the measurement ensemble, $\boldsymbol{y}^{\mathrm{e}}$ under the Compressive sensing/sampling (CS) framework (e.g., Candès and Tao, 2006), and then the complete 2D geotechnical properties data, $\mathbf{F}'_{Q_1}$ and $\mathbf{F}'_{Q_2}$ can be simultaneously reconstructed using Eqs. (2)&(3). It is important to note that when the measured data points are limited, the reconstructed results may contain significant statistical uncertainty. To incorporate such uncertainty, CS can be combined with a Bayesian framework, known as Bayesian compressive sampling (BCS), for generating random field samples (e.g., Wang et al., 2019).

Under a Bayesian framework, an approximation of the joint weight coefficients vector for two correlated geotechnical properties, $\boldsymbol{\omega}^{\mathrm{e}}$, denoted as $\hat{\boldsymbol{\omega}}^{\mathrm{e}}$ can be probabilistically estimated from the measurement ensemble, $\boldsymbol{y}^{\mathrm{e}}$ (e.g., Ji et al., 2008):

$$p(\hat{\boldsymbol{\omega}}^{\mathrm{e}} \left| \boldsymbol{y}^{\mathrm{e}} \right.) = \frac{p(\boldsymbol{y}^{\mathrm{e}} \left| \hat{\boldsymbol{\omega}}^{\mathrm{e}} \right.) \times p(\hat{\boldsymbol{\omega}}^{\mathrm{e}})}{p(\boldsymbol{y}^{\mathrm{e}})} \tag{8}$$

where $p(\boldsymbol{y}^{\mathrm{e}} \left| \hat{\boldsymbol{\omega}}^{\mathrm{e}} \right.)$ is the Gaussian likelihood function reflecting the probability of observing the data ensemble, $\boldsymbol{y}^{\mathrm{e}}$ given $\hat{\boldsymbol{\omega}}^{\mathrm{e}}$; a three-level hierarchical prior, i.e., Gaussian-Inverse Gamma-Gamma prior, $p(\hat{\boldsymbol{\omega}}^{\mathrm{e}})$ is adopted to promote the sparsity of $\hat{\boldsymbol{\omega}}^{\mathrm{e}}$; $p(\boldsymbol{y}^{\mathrm{e}})$ is a normalizing constant. Based on these settings, $p(\hat{\boldsymbol{\omega}}^{\mathrm{e}} \left| \boldsymbol{y}^{\mathrm{e}} \right.)$ can be conveniently represented as a multivariate normal distribution with a mean, $\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{\mathrm{e}}}$ and a covariance, $\mathbf{COV}_{\hat{\boldsymbol{\omega}}^{\mathrm{e}}}$ (e.g., Wang et al., 2019):

$$\begin{aligned} \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{\mathrm{e}}} &= \mathbf{COV}_{\hat{\boldsymbol{\omega}}^{\mathrm{e}}} (\mathbf{A}^{\mathrm{e}})^{\mathrm{T}} \boldsymbol{y}^{\mathrm{e}} \tau \\ \mathbf{COV}_{\hat{\boldsymbol{\omega}}^{\mathrm{e}}} &= ((\mathbf{A}^{\mathrm{e}})^{\mathrm{T}} \mathbf{A}^{\mathrm{e}} \tau + \mathbf{D}^{\alpha})^{-1} \end{aligned} \tag{9}$$

where $\mathbf{D}^{\alpha}$ is an $3N \times 3N$ diagonal matrix with the elements of vector $\boldsymbol{\alpha} = \left[\alpha_1, \alpha_2, ..., \alpha_{3N}\right]$ on the main diagonal, where $\boldsymbol{\alpha}$ follows a generalized inverse Gaussian distribution:

$$p(\boldsymbol{\alpha}\,|\,\hat{\boldsymbol{\omega}}^{\mathrm{e}},\tau,\gamma,\boldsymbol{y}^{\mathrm{e}}) = \prod_{t=1}^{3N}\exp(-\frac{a_t\alpha_t+\gamma\alpha_t^{-1}}{2})(\alpha_t)^{p-1} \times \frac{(\alpha_t/\gamma)^{p/2}}{2K_p(\sqrt{\alpha_t\gamma})} \tag{10}$$

where K_p represents a modified Bessel function of the second kind with parameter $p=-1/2$ and $a_t=(\hat{\omega}_t^{\mathrm{e}})^2$. $p(\tau\,|\,\hat{\boldsymbol{\omega}}^{\mathrm{e}},\gamma,\boldsymbol{\alpha},\boldsymbol{y}^{\mathrm{e}})$ and $p(\gamma\,|\,\hat{\boldsymbol{\omega}}^{\mathrm{e}},\tau,\boldsymbol{\alpha},\boldsymbol{y}^{\mathrm{e}})$ follow two Gamma distributions:

$$p(\tau\,|\,\hat{\boldsymbol{\omega}}^{\mathrm{e}},\gamma,\boldsymbol{\alpha},\boldsymbol{y}^{\mathrm{e}}) = Gamma\left(c_n,\ d_n\right) \tag{11}$$

$$p(\gamma\,|\,\hat{\boldsymbol{\omega}}^{\mathrm{e}},\tau,\boldsymbol{\alpha},\boldsymbol{y}^{\mathrm{e}}) = Gamma\left(3N+a_0,\ b_0+\sum_{t=1}^{3N}\alpha^{-1}{}_t\right) \tag{12}$$

where $c_n = M/2+1$; M represents the total number of measurements; $d_n = d_0 + 1/2[(\boldsymbol{y}^{\mathrm{e}})^{\mathrm{T}}\boldsymbol{y}^{\mathrm{e}} - 2(\hat{\boldsymbol{\omega}}^{\mathrm{e}})^{\mathrm{T}}(\mathbf{A}^{\mathrm{e}})^{\mathrm{T}}\boldsymbol{y}^{\mathrm{e}} + (\hat{\boldsymbol{\omega}}^{\mathrm{e}})^{\mathrm{T}}(\mathbf{A}^{\mathrm{e}})^{\mathrm{T}}\mathbf{A}^{\mathrm{e}}\hat{\boldsymbol{\omega}}^{\mathrm{e}}]$; a_0, b_0 and d_0 may be taken as a small value to achieve the non-informative prior information for $\hat{\boldsymbol{\omega}}^{\mathrm{e}}$ (e.g., $a_0 = b_0 = d_0 = 10^{-4}$). Note that the distribution types of $\boldsymbol{\alpha}$, τ and γ are derived based on the Gaussian likelihood function and three-level hierarchical prior. The above equations can be efficiently solved using Markov chain Monte Carlo (MCMC), and many (e.g., N_b = 200) 2D cross-correlated random field samples of Q_1 and Q_2 can be generated using MCMC simulation. The detailed formulation for the generation of random field samples using MCMC can be referred to Guan and Wang (2023).

3. Illustration example

It is widely recognized that cone penetration test (CPT) tip resistance, q_c and sleeve friction, f_s data exhibit positive correlations (e.g., Robertson and Campanella 1983). In this section, the proposed cross-correlated random fields method is employed to generate cross-correlated random field samples of q_c and f_s within a cross-section from sparse measurements at the Wildlife Liquefaction Array (WLA) site situated in the Imperial Valley, California, USA. The surface soil layer at the WLA site primarily comprises a 2.5 to 3.3m thick layer of silty clay to clayey silt, underlain by a 3.5m to 4m thick layer of silty sand and sandy silt. The water table at the site is approximately 1.2m below the ground surface.

A representative geotechnical cross-section, denoted as A-A', is delineated in Fig. 1 by a red solid line. The width of cross-section A-A' spans approximately 43m, while its depth ranges from about 3.4m to around 7.0m, containing the layer of silty sand and sandy silt. Five CPT soundings (i.e., 1Cg, 3Cg, 4Cg, 5Cg, and 7Cg) are conducted within this cross-section, as illustrated in Fig. 1 by solid circles. The corresponding CPT q_c and f_s data are obtained from the Network for Earthquake Engineering Simulation Research (NEES) Database (URL: http://nees.ucsb.edu/facilities/wla), as depicted in Fig. 2. The calculated cross-correlation coefficient for these measurements is determined to be 0.4, indicating a medium positive correlation. Subsequently, these measurement data serve as the input for the proposed method to generate N_b = 200 cross-correlated RFSs of q_c and f_s.

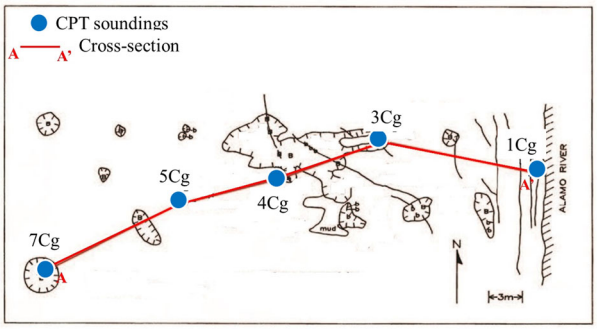


Figure 1. Locations of CPT used in this study at the Wildlife Liquefaction Array (after Holzer and Youd 2007)

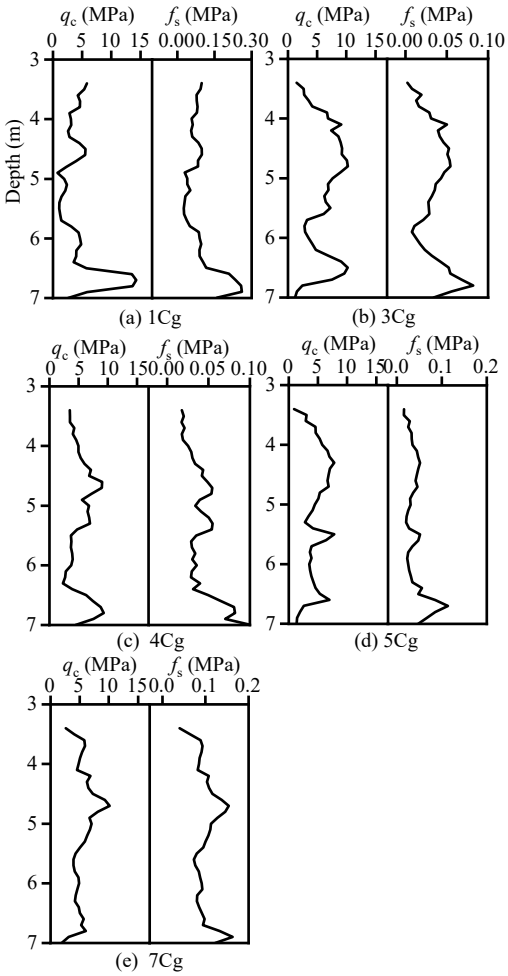
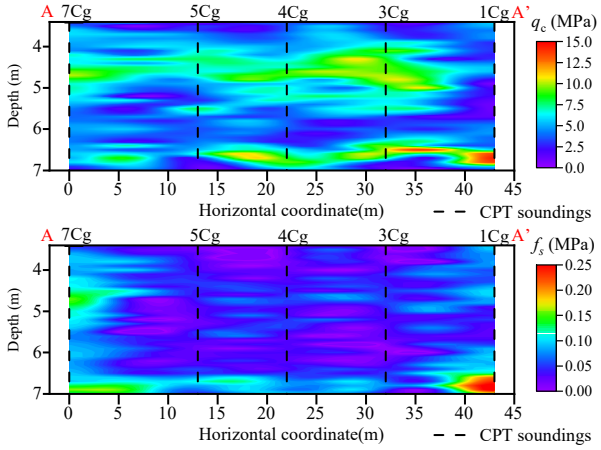


Figure 2. Five CPT tip resistance q_c and sleeve friction, f_s data profiles at the Wildlife Liquefaction Array

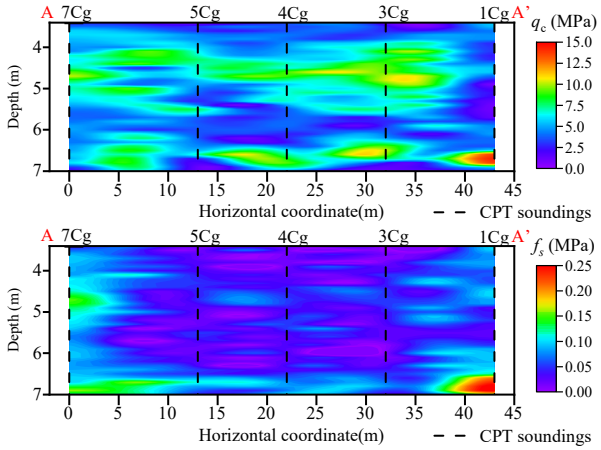
Two typical cross-correlated RFS pairs of q_c and f_s generated from the sparse measurements are depicted in Fig. 3. For each generated RFS pair of q_c and f_s, a Pearson cross-correlation coefficient, ρ, between q_c and f_s data can

be computed, resulting in 200 cross-correlation coefficients. The histogram illustrating these 200 calculated cross-correlation coefficients ranges from approximately 0.30 to about 0.47, with a mean of 0.4 and a standard deviation of 0.03, as presented in Fig. 4. It is noteworthy that the cross-correlation coefficients of the generated RFS pairs of q_c and f_s generally align with those of the measured data points.

Furthermore, the mean and standard deviation (SD) of the 200 generated RFSs of q_c and f_s are depicted in Figs. 5&6. Figs. 5(b)&6(b) reveals that the statistical uncertainties at locations proximal to the measured data points are notably smaller than those at locations distant from the measurements. These observations imply that the cross-correlation between q_c and f_s data can be properly characterized using 200 RFSs generated using the proposed method.



(a) Cross-correlated RFS pair #1 (cross-correlation coefficient, ρ = 0.41)



(b) Cross-correlated RFS pair #2 (cross-correlation coefficient, ρ = 0.38)

Figure 3. Two cross-correlated random field sample (RFS) pairs of cone penetration test (CPT) tip resistance, q_c and sleeve friction, f_s data generated from the proposed method

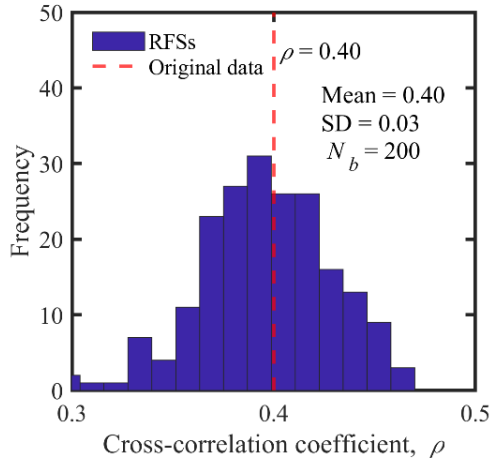
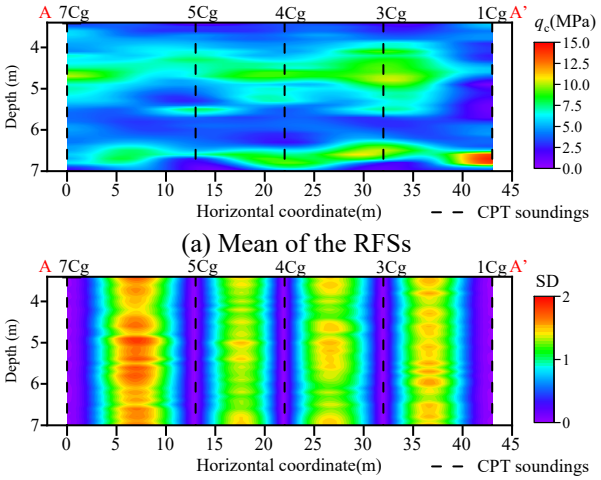


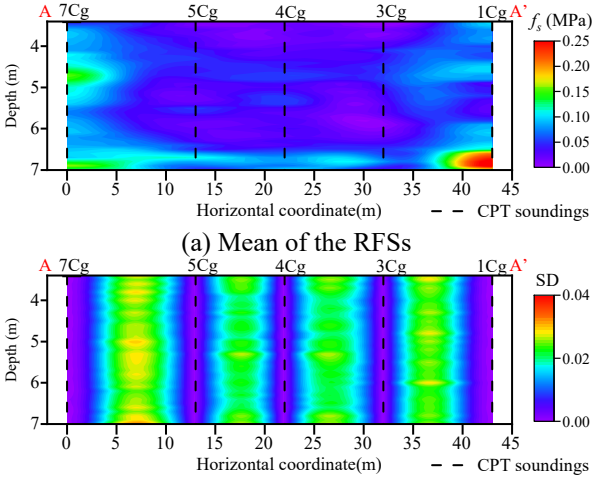
Figure 4. Histogram of the cross-correlation coefficients of N_b = 200 pairs of generated random field samples



(a) Mean of the RFSs

(b) Standard deviation (SD) of the RFSs

Figure 5. Mean and standard deviation of 200 generated random field samples (RFSs) of cone penetration test (CPT) tip resistance, q_c



(a) Mean of the RFSs

(b) Standard deviation (SD) of the RFSs

Figure 6. Mean and standard deviation of 200 generated random field samples (RFSs) of cone penetration test (CPT) sleeve friction, f_s

4. Conclusions

This paper introduces a novel data-driven approach for generating 2D cross-correlated random field samples (RFSs) of geotechnical properties from sparse measurement data. By simultaneously exploiting auto- and cross-correlation structures of different geotechnical properties through joint sparse representation, this method eliminates the need for explicit prior information about random field parameters. Leveraging Bayesian Compressive Sampling (BCS) and Markov Chain Monte Carlo (MCMC) sampling techniques, the method enables the generation of cross-correlated RFSs for correlated geotechnical properties, incorporating statistical uncertainty arising from sparse data interpretation. The presented method was demonstrated using real CPT data from the Wildlife Liquefaction Array (WLA), USA. The illustration example indicated that the presented method can properly generate cross-correlated q_c and f_s cross-sections from sparse data.

Acknowledgements

The work described in this paper was supported by a grant from the Research Grant Council of Hong Kong Special Administrative Region (Project No: CityU 11203322), a grant from The Science and Technology Development Fund, Macao SAR (Project no.: 001/2024/SKL), a grant from the National Natural Science Foundation of China (grant No: 42307215) and a grant from Shenzhen Science and Technology Innovation Commission (Shenzhen-Hong Kong-Macau Science and Technology Project (Category C) No: SGDX20210823104002020), China. The financial support is gratefully acknowledged.

References

Candes, E. J., and Tao, T. “Near-optimal signal recovery from random projections: Universal encoding strategies?”, IEEE Trans. Inf. Theory, 52(12), pp. 5406-5425, 2006.

Guan, Z., and Wang, Y. “Data-driven simulation of two-dimensional cross-correlated random fields from limited measurements using joint sparse representation”, Reliab. Eng. Syst. Saf., 238, pp. 109408, 2023.

Guan, Z., and Wang, Y. “Non-parametric construction of site-specific non-Gaussian multivariate joint probability distribution from sparse measurements”, Struct. Saf., 91, pp. 102077, 2021.

Guan, Z., and Wang, Y. “Statistical charts for determining sample size at various levels of accuracy and confidence in geotechnical site investigation”, Geotechnique, 70(12), pp. 1145-1159, 2020.

Guan, Z., Wang, Y., and Phoon, K. K. “Fusion of sparse non-co-located measurements from multiple sources for geotechnical site investigation”, Can. Geotech. J., https://doi.org/10.1139/cgj-2023-028. 2023.

Holzer, T. L., and Youd, T. L. “Liquefaction, ground oscillation, and soil deformation at the Wildlife Array, California”, Bull. Seismol. Soc. Am., 97(3), pp. 961-976, 2007.

Robertson, P. K., and Campanella, R. G. “Interpretation of cone penetration tests: Part I: Sand”, Can. Geotech. J., 20(4), pp. 718-733, 1983.

Robin, M. J. L., Gutjahr, A. L., Sudicky, E. A., and Wilson, J. L. “Cross-correlated random field generation with the direct Fourier transform method”, Water Resour. Res., 29(7), pp. 2385-2397, 1993.

Shinozuka, M., and Deodatis, G. “Simulation of stochastic processes by spectral representation”. 1991.

Vanmarcke, E. H. “Random fields: analysis and synthesis”, World Scientific, London. 2010.

Vořechovský, M. “Simulation of simply cross correlated random fields by series expansion methods”, Struct. Saf., 30(4), pp. 337-363, 2008.

Wang, Y., and Zhao, T. “Statistical interpretation of soil property profiles from sparse data using Bayesian Compressive Sampling”, Geotechnique, 67(6), pp. 523-536, 2017.

Wang, Y., Zhao, T., Hu, Y., and Phoon, K. K. “Simulation of random fields with trend from sparse measurements without detrending”, J. Eng. Mech., 145(2), pp. 04018130, 2019.