The paper was published in the proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR 2019) and was edited by Jianye Ching, Dian-Qing Li and Jie Zhang. The conference was held in Taipei, Taiwan 11-13 December 2019.

Quantification of Statistical Uncertainties in Mean, Standard Deviation and Cross-Correlation of Geotechnical Properties Estimated from Sparse Measurements

Tengyuan Zhao[1] and Yu Wang[2]

[1]School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Shanxi Province, China
E-mail: tyzhao@xjtu.edu.cn
[2]Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong SAR.
E-mail: yuwang@cityu.edu.hk

Abstract: Statistics, such as mean, standard deviation (SD), cross-correlation coefficient of/between geotechnical properties are commonly used in geotechnical design and analysis, especially in the probability-based design or analysis. Accurate estimation of these statistics generally requires numerous measurements, which are usually not available in geotechnical engineering practice due to time and budget constraints. In such cases, the estimated statistics might not be accurate but contain significant statistical uncertainty. As the statistical uncertainty in estimated statistics would propagate to subsequent design or analysis and might lead to undesirable results, quantification of the uncertainty in estimated statistics is important. Quantification of the statistical uncertainty in the estimated mean, SD, cross-correlation is a well-known topic in statistics, and it is usually performed analytically under some assumptions, such as the assumption that measurement data are independent and identically distributed with a Gaussian distribution. It is worth noting that, however, the geotechnical properties exhibit auto-correlation and the independence assumption is often invalid. In addition, geotechnical properties at a specific site may not follow a Gaussian distribution due to the unique geological formation process that soils at the site have undergone. As a result, the methods based on the assumptions above may not be applicable to quantify the statistical uncertainty in the estimated mean, SD and cross-correlation. This paper aims to address this problem using a Bayesian compressive sampling (BCS)-based method. The proposed method not only considers the spatial auto-correlation for a geotechnical property but also simultaneously considers the cross-correlation among different properties. Cone penetration test (CPT) data from a real site are used to illustrate the proposed method. The results show that the proposed method performs reasonably well.

Keywords: Bootstrapping; Bayesian compressive sampling/sensing (BCS); uncertainty quantification; geotechnical characterization; reliability-based design or analysis

1 Introduction

Statistics, such as mean, standard deviation (SD), and cross-correlation of/between geotechnical properties are commonly used in geotechnical design and analysis, especially in probability-based design or analysis (e.g., Baecher and Christian 2003; Phoon 2008). Accurate estimation of these statistics generally requires numerous measurements. In geotechnical engineering practice, however, the number of measurements collected from a specific site is often sparse and limited, due to time and budget constraints (e.g., Phoon 2016; Wang and Zhao 2016&2017). In such cases, the estimated statistics might not be accurate but contain significant statistical uncertainty. As the statistical uncertainty in the estimated statistics would propagate to subsequent design or analysis and might lead to undesirable results, quantification of the uncertainty in these commonly used statistics is important (e.g., Luo et al. 2013).

Quantification of uncertainty in estimated mean, SD and cross-correlation is a well-known topic in statistics (e.g., Kenney and Keeping 1951), which is usually performed analytically under some assumptions, such as the assumption that measurement data are independent and identically distributed with a Gaussian distribution. It is worth noting that, however, geotechnical properties exhibit auto-correlation and independence assumption is often not valid (e.g., Baecher and Christian 2003). In addition, geotechnical properties at a specific site may not follow a Gaussian distribution due to the unique geological formation process that soils at the site have undergone (e.g., Wang et al. 2015; Phoon 2016). As a result, the methods based on independent or Gaussian assumptions may not be applicable to quantify statistical uncertainty in the estimated statistics.

In addition to analytical methods, other methods are also available to quantify the statistical uncertainty associated with an estimator (e.g., mean), such as the bootstrap method initially developed by Efron and Tibshirani (1993). The basic idea of the bootstrap method is to assign accuracy of an estimator based on a repeated sampling of the available measurements. This method has been used in geotechnical engineering to quantify the statistical uncertainty of mean and SD of some soil parameters, such as normalized undrained shear strength when analyzing serviceability for a braced excavation (e.g., Luo et al. 2013). Although the bootstrap method bypasses the assumption of distribution for measurements (e.g., Efron and Tibshirani 1993), the spatial

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)
Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Published by Research Publishing, Singapore.
ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0_IS15-11-cd

auto-correlation embedded in the measurements are not considered. Therefore, it leads to a problem of how to properly quantify the statistical uncertainty in the estimated statistics for geotechnical properties when estimated from sparse measurements, with the consideration of the spatial auto-correlation of geotechnical properties.

This paper aims to address this problem using a Bayesian compressive sampling (BCS)-based method (e.g., Zhao and Wang 2018). Comparing to Zhao and Wang (2018), where simulation of cross-correlated random field samples is of interest, the method in this paper aims to investigate its capability of quantifying statistical uncertainty in the mean and SD of geotechnical properties from sparse measurements, with explicit consideration of both the spatial auto-correlation for a geotechnical property and the consideration of the cross-correlation among different properties. In addition, the proposed method bypasses the assumption of probability distribution for the available measurements. After this introduction, the method is first developed based on BCS. Then, it is illustrated using a set of real cone penetration test (CPT) data.

2 Bayesian Compressive Sampling-Based Bootstrap Method

Bayesian compressive sampling/sensing (BCS) is a probabilistic version of compressive sampling/sensing (CS) to reconstruct a signal (e.g., soil property profile) $\boldsymbol{f}$, a column vector with a length of N, from its partial measurements (e.g., Wang and Zhao 2017; Zhao et al. 2018). The partial measurements are denoted as a column vector $\boldsymbol{y}$ with a length of M, and $M<<N$. In the context of BCS or CS, $\boldsymbol{f}$ can be concisely represented by a limited number of basis functions (e.g., wavelet Daubechies 16) in different orders. In math, $\boldsymbol{f} = \mathbf{B}\boldsymbol{\omega}$. $\mathbf{B}$ is an $N\times N$ orthonormal matrix, columns of which represent the basis functions (e.g., wavelet Daubechies 16) in different orders. $\boldsymbol{\omega}$ is a column vector with a length of N, representing the weights corresponding to columns of $\mathbf{B}$, and most elements of $\boldsymbol{\omega}$ are negligible except several ($S<<N$) non-trivial ones. Therefore, a signal of interest can be reconstructed if the non-trivial elements of $\boldsymbol{\omega}$ can be estimated using sparse measurements $\boldsymbol{y}$. The relation between $\boldsymbol{\omega}$ and $\boldsymbol{y}$ is formulated as $\boldsymbol{y} = \boldsymbol{\Psi f} = \boldsymbol{\Psi}\mathbf{B}\boldsymbol{\omega} = \mathbf{A}\boldsymbol{\omega}$. $\mathbf{A} = \boldsymbol{\Psi}\mathbf{B}$, where $\boldsymbol{\Psi}$ is a measurement matrix, reflecting the locations of $\boldsymbol{y}$ in $\boldsymbol{f}$. Given $\boldsymbol{y} = \mathbf{A}\boldsymbol{\omega}$, the non-trivial elements in $\boldsymbol{\omega}$ can be estimated by some deterministic or Bayesian methods (e.g., Wang and Zhao 2016&2017). The estimated weight vector is denoted as $\boldsymbol{\omega_s}$, which has the same length as $\boldsymbol{\omega}$. Note that all elements of $\boldsymbol{\omega_s}$ are set to zero except the S non-trivial ones. Once $\boldsymbol{\omega_s}$ is obtained, the reconstructed signal $\hat{\boldsymbol{f}}$ can be derived as $\hat{\boldsymbol{f}} = \mathbf{B}\boldsymbol{\omega_s}$.

Since $\boldsymbol{\omega_s}$ is estimated from sparse measurements $\boldsymbol{y}$, it may be inaccurate and contain significant statistical uncertainty. The uncertainty associated with $\boldsymbol{\omega_s}$ is quantified by its probability density function (PDF), which has been shown to follow a multivariate Student's t distribution with $2c_n$ degree of freedom, a mean vector of $\boldsymbol{\mu}_{\omega_s}$ and covariance matrix $\mathbf{COV}_{\omega_s}$. $\boldsymbol{\mu}_{\omega_s}$ and $\mathbf{COV}_{\omega_s}$ are expressed as (Wang and Zhao 2017):

$$\begin{aligned}\boldsymbol{\mu}_{\omega_s} &= \mathbf{H}\mathbf{A}^{\mathrm{T}}\boldsymbol{y} = (\mathbf{A}^{\mathrm{T}}\mathbf{A}+\mathbf{D})^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{y} \\ \mathbf{COV}_{\omega_s} &= \frac{d_n\mathbf{H}}{c_n-1} = \frac{d_n(\mathbf{A}^{\mathrm{T}}\mathbf{A}+\mathbf{D})^{-1}}{c_n-1}\end{aligned} \tag{1}$$

where "$(\cdot)^{\mathrm{T}}$" means transpose operation; $\mathbf{H} = (\mathbf{A}^{\mathrm{T}}\mathbf{A}+\mathbf{D})^{-1}$; $c_n = M/2+c$; $d_n = d+(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} - \boldsymbol{\mu}_{\omega_s}^{\mathrm{T}}\mathbf{H}^{-1}\boldsymbol{\mu}_{\omega_s})/2$; c and d are small non-negative constants (e.g., $c = d = 10^{-4}$); $\mathbf{D}$ = a diagonal matrix with $D_{i,i} = \alpha_i$ (i = 1, 2, …, N), which are unknown non-negative variables and can be obtained by maximizing the likelihood function of measurements $\boldsymbol{y}$ (e.g., Ji et al. 2009; Wang and Zhao 2017; Ching and Phoon 2017):

$$L = \ln[P(\boldsymbol{y})] = -\frac{1}{2}[(\mathrm{M}+2c)\ln(\boldsymbol{y}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{y}+2d)+\ln(\mathbf{C})]+const \tag{2}$$

where $\mathbf{C} = \mathbf{I}_{M\times M}+\mathbf{A}\mathbf{D}^{-1}\mathbf{A}^{\mathrm{T}}$. $\mathbf{I}_{M\times M}$ is an identity matrix with a dimension of $M\times M$. Once the $D_{i,i} = \alpha_i$ (i = 1, 2, …, N) are determined, $\boldsymbol{\mu}_{\omega_s}$ and $\mathbf{COV}_{\omega_s}$ in Eq. (1) can be calculated accordingly. Then, substituting Eq. (1) into $\hat{\boldsymbol{f}} = \mathbf{B}\boldsymbol{\omega_s}$ leads to statistics of $\hat{\boldsymbol{f}}$ as below with the definition of mean and covariance matrix:

$$\begin{aligned}\boldsymbol{\mu}_{\hat{f}} &= \mathrm{E}(\hat{\boldsymbol{f}}) = \mathrm{E}(\mathbf{B}\boldsymbol{\omega_s}) = \mathbf{B}\,\mathrm{E}(\boldsymbol{\omega_s}) = \mathbf{B}\boldsymbol{\mu}_{\omega_s} \\ \mathbf{COV}_{\hat{f}} &= \mathrm{E}[(\hat{\boldsymbol{f}}-\mathrm{E}(\hat{\boldsymbol{f}}))(\hat{\boldsymbol{f}}-\mathrm{E}(\hat{\boldsymbol{f}}))^{\mathrm{T}}] = \mathbf{B}\mathbf{COV}_{\omega_s}\mathbf{B}^{\mathrm{T}}\end{aligned} \tag{3}$$

where "E (·)" represents the expectation. $\boldsymbol{\mu}_{\hat{f}}$ represents the mean of $\hat{\boldsymbol{f}}$; whereas diagonal elements of $\mathbf{COV}_{\hat{f}}$ quantify the statistical uncertainty in estimating $\hat{\boldsymbol{f}}$. Eq. (3) suggests that once sparse measurements $\boldsymbol{y}$ of a geotechnical property along a spatial direction are available, the mean and statistical uncertainty associated with the profile of this property can be obtained. Note that the method described above is only applicable to one geotechnical property. In geotechnical engineering practice, however, variations of several geotechnical properties along a spatial direction may be required. In addition, these geotechnical properties are typically multivariate, and correlation exists among different properties (e.g., Ching et al. 2016), such as the cohesion and friction angle of soils, uniaxial compressive strength and Young's modulus of rocks, among others (e.g., Wang and Aladejare 2016; Wang and Akeju 2016). In such cases, different geotechnical properties should be reconstructed simultaneously, which can be done by slightly modifying Eq. (2) as (e.g., Ji et al. 2009):

$$L = \sum_{j=1}^{n} \ln[P(\boldsymbol{y}_{Q_j})] = -\frac{1}{2}\sum_{j=1}^{n}[(\mathrm{M}+2c)\ln(\boldsymbol{y}_{Q_j}{}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{y}_{Q_j}+2d)+\ln(\det(\mathbf{C}))]+const \tag{4}$$

"n" represents the number of different geotechnical properties and it is taken as $n = 2$ in this paper for derivation and illustration convenience. After maximizing Eq. (4), the most probable $D_{i,i} = \alpha_i$ (i = 1, 2, …, N) are determined. Then, using Eqs. (1) and (3), geotechnical property Q_1 and Q_2 profile can be obtained readily from measurements $\boldsymbol{y}_{Q_1}$ and $\boldsymbol{y}_{Q_2}$. Mean and covariance matrices of estimated Q_1 and Q_2 profile are expressed as $\boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_1}$, $\mathbf{COV}_{\hat{\boldsymbol{Q}}_1}$ and $\boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_2}$, $\mathbf{COV}_{\hat{\boldsymbol{Q}}_2}$, respectively.

Then, cross-correlated Q_1 and Q_2 samples may be generated as follows (Zhao and Wang 2018)

$$\begin{bmatrix}\hat{\boldsymbol{Q}}_1\\ \hat{\boldsymbol{Q}}_2\end{bmatrix} \approx \begin{bmatrix}\boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_1}\\ \boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_2}\end{bmatrix} + \hat{\mathbf{D}}_{\hat{Q}_{12}}\mathbf{t}^{S}_{\hat{Q}_{12}}\sqrt{\boldsymbol{\lambda}^{S}_{\hat{Q}_{12}}}\boldsymbol{\xi}^{S}_{\hat{Q}_{12}} = \begin{bmatrix}\boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_1}\\ \boldsymbol{\mu}_{\hat{\boldsymbol{Q}}_2}\end{bmatrix} + \hat{\mathbf{D}}_{\hat{Q}_{12}}\sum_{i=1}^{2S}\boldsymbol{t}^{S}_{\hat{Q}_{12_i}}\sqrt{\lambda^{S}_{\hat{Q}_{12_i}}}\xi^{S}_{\hat{Q}_{12_i}} \tag{5}$$

where $\hat{\mathbf{D}}_{\hat{Q}_{12}} = \begin{bmatrix}\mathbf{D}_{\hat{\boldsymbol{Q}}_1} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}_{\hat{\boldsymbol{Q}}_2}\end{bmatrix}$; $\mathbf{D}_{\hat{\boldsymbol{Q}}_1}$ and $\mathbf{D}_{\hat{\boldsymbol{Q}}_2}$ are the diagonal matrices representing the square root of diagonal elements of $\mathbf{COV}_{\hat{\boldsymbol{Q}}_1}$ and $\mathbf{COV}_{\hat{\boldsymbol{Q}}_2}$, respectively; $\mathbf{t}^{S}_{\hat{Q}_{12}}$ and $\boldsymbol{\lambda}^{S}_{\hat{Q}_{12}}$ are the eigenvector and eigenvalue matrices of the correlation matrix $\mathbf{C_R} = \begin{bmatrix}\mathbf{C}_{\mathbf{R}_1} & \rho_{12}\mathbf{C}_{\mathbf{R}_1}\\ \rho_{12}\mathbf{C}_{\mathbf{R}_1} & \mathbf{C}_{\mathbf{R}_2}\end{bmatrix}$, where $\mathbf{C}_{\mathbf{R}_1}$ and $\mathbf{C}_{\mathbf{R}_2}$ are the correlation matrix corresponding to $\mathbf{COV}_{\hat{\boldsymbol{Q}}_1}$ and $\mathbf{COV}_{\hat{\boldsymbol{Q}}_2}$, respectively. In this study, $\mathbf{C}_{\mathbf{R}_1} = \mathbf{C}_{\mathbf{R}_2}$ due to the usage of Eq. (4) (e.g., Zhao and Wang 2018). This suggests that the autocorrelation structure for different soil properties in this study is taken as the same. This may be justified by noting that the autocorrelation of soil properties is the result of the spatially varying constitutive nature of the soil over space (Fenton and Griffiths 2003). $\boldsymbol{\xi}^{S}_{\hat{Q}_{12}}$ represent a series of uncorrelated random variables with zero mean and unit variances, such as standard Gaussian random variables. ρ_{12} is the cross-correlation between Q_1 and Q_2, which may be estimated as (e.g., Ang and Tang 2007)

$$\hat{\rho}_{12} = \frac{\sum_{i=1}^{\mathrm{M}}(y_{Q_{1,i}}-\mu_{Q_1})(y_{Q_{2,i}}-\mu_{Q_2})}{\sqrt{\sum_{i=1}^{\mathrm{M}}(y_{Q_{1,i}}-\mu_{Q_1})}\sqrt{\sum_{i=1}^{\mathrm{M}}(y_{Q_{2,i}}-\mu_{Q_2})}} \tag{6}$$

where μ_{Q_1} and μ_{Q_2} represent the mean of $\boldsymbol{y}_{Q_1}$ and $\boldsymbol{y}_{Q_2}$, respectively. $y_{Q_{1,i}}$ and $y_{Q_{2,i}}$ are the i-th element of $\boldsymbol{y}_{Q_1}$ and $\boldsymbol{y}_{Q_2}$, respectively.

Given Eq. (5), a large number N_B of cross-correlated Q_1 and Q_2 samples can be generated readily by realizing N_B sets of $\boldsymbol{\xi}^{S}_{\hat{Q}_{12}}$, i.e., $\xi^{S}_{\hat{Q}_{12_i}}$ (i = 1, 2, …, 2S) (Zhao and Wang 2018). The Q_1 and Q_2 samples are denoted as $\hat{\boldsymbol{f}}^{(1)}_{\hat{Q}_1}, \hat{\boldsymbol{f}}^{(2)}_{\hat{Q}_1}, \cdots, \hat{\boldsymbol{f}}^{(N_B)}_{\hat{Q}_1}$ and $\hat{\boldsymbol{f}}^{(1)}_{\hat{Q}_2}, \hat{\boldsymbol{f}}^{(2)}_{\hat{Q}_2}, \cdots, \hat{\boldsymbol{f}}^{(N_B)}_{\hat{Q}_2}$, respectively. Although sparse measurements $\boldsymbol{y}_{Q_1}$ and $\boldsymbol{y}_{Q_2}$ only have a length of M, each Q_1 and Q_2 samples, i.e., $\hat{\boldsymbol{f}}^{(i)}_{\hat{Q}_1}$ and $\hat{\boldsymbol{f}}^{(i)}_{\hat{Q}_2}$ have a length of N. Using the N_B sample, statistical uncertainty of mean and SD can be calculated. Consider, for example, the Q_1 case. Given a Q_1 sample, e.g., $\hat{\boldsymbol{f}}^{(1)}_{\hat{Q}_1}$, its mean and SD can be obtained as $(\hat{\mu}^{*}_{Q_1})^{(1)} = \sum_{i=1}^{N}\hat{f}^{(1)}_{Q_{1,i}}/N$ and $(\hat{\sigma}^{*}_{Q_1})^{(1)} = \sqrt{\sum_{i=1}^{N}(\hat{f}^{(1)}_{Q_{1,i}} - \hat{\mu}^{(1)}_{Q_1})^2/(N-1)}$. N_B Q_1 sample leads to N_B $\hat{\mu}^{*}_{Q_1}$ and $\hat{\sigma}^{*}_{Q_1}$, i.e., $(\hat{\mu}^{*}_{Q_1})^{(i)}$ and $(\hat{\sigma}^{*}_{Q_1})^{(i)}$ (i = 1, 2, …, N_B). Using the N_B $\hat{\mu}^{*}_{Q_1}$ and $\hat{\sigma}^{*}_{Q_1}$, the sampling distributions and statistics (e.g., 95% confidence interval, CI) can be obtained, which reflect the statistical uncertainty in the estimated mean and SD of Q_1 from $\boldsymbol{y}_{Q_1}$. A similar procedure can be performed to quantify the statistical uncertainty in the estimated mean and SD of Q_2 and cross-correlation between Q_1 and Q_2. Note that the method proposed in this study is essentially a variant of bootstrap method, and it automatically preserves the spatial auto-correlation of Q_1 and Q_2 due to the reconstruction by BCS and explicitly considers the cross-correlation between Q_1 and Q_2. In addition, there is no assumption of probability distribution made on the measurements $\boldsymbol{y}_{Q_1}$ and $\boldsymbol{y}_{Q_2}$. Therefore, all issues mentioned in Section 1 are reasonably addressed by the proposed method.

3 Illustrative Example

In this section, the proposed method is illustrated using a set of non-stationary cone penetration test (CPT) data, i.e., tip resistance q_c data, and sleeve friction f_s data. The CPT was performed in a lower clay layer at a site in Texas, USA (e.g., Stuedlein et al. 2012), and the data was downloaded from the website of TC304 databases (http://140.112.12.21/issmge/tc304.htm?=6). The lower clay layer varies from a depth of around 4.5m to a depth of around 15.3m. Both the q_c and f_s data are recorded at an interval of 0.02m. In this example, 512 (q_c, f_s) data pairs, i.e., the solid lines in Figure 1a&1b, which vary from a depth of 5m to a depth of 15.22m, are adopted for comparison and validation. M = 11 measurement data pairs extracted from the solid lines are taken as input for the proposed method, and they are represented by open circles in Figure 1a&1b. It is emphasized that CPT data are used in this study because CPT data are recorded with a high resolution and are almost continuous. In addition, q_c and f_s are somehow cross-correlated. Therefore, CPT q_c and f_s data can be used to explore the efficiency of the proposed method.

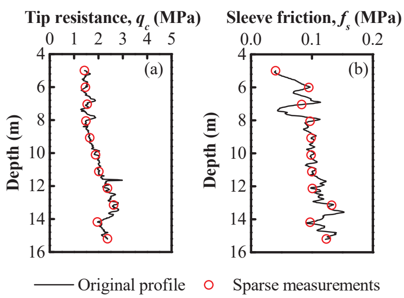

Figure 1. Original q_c/f_s profile and sparse data.

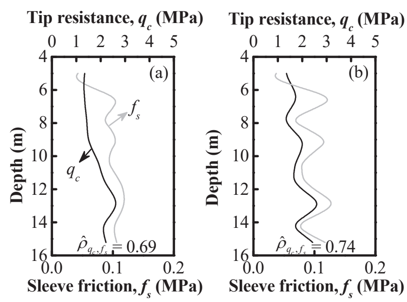

Figure 2. BCS-based samples of q_c/f_s profiles.

Table 1. 95% CI of mean and SD of q_c and f_s from the proposed method using M = 11 measurements.

	Mean			SD		
	True	95% confidence interval (CI)		True	95% confidence interval (CI)	
		The proposed method	Bootstrapping		The proposed method	Bootstrapping
q_c	1.94	[1.74, 2.01] (0.27)	[1.67, 2.13] (0.46)	0.41	[0.30, 0.55] (0.25)	[0.24, 0.51] (0.27)
f_s	0.10	[0.088, 0.105] (0.017)	[0.084, 0.108] (0.024)	0.02	[0.015, 0.029] (0.014)	[0.007, 0.034] (0.027)

 # Note that the value in the bracket (·) represents the length of the corresponding 95% CI.

3.1 Mean and SD

In this example, q_c is taken as Q_1 while f_s is taken as Q_2. Then, given the sparse measurements on q_c and f_s, i.e., the open circles in Figure 1a&1b, a large number (e.g., N_B = 1000) of q_c and f_s profile samples can be generated, two of which are shown in Figure 2a&b. Subsequently, following the procedure discussed in the last section, the N_B = 1000 mean and SD of q_c and f_s can be calculated accordingly. Using the N_B = 1000 mean and SD, the histograms can be constructed. For example, Figure 3a plots the histogram of the mean of q_c (i.e., μ_{qc}) and summarizes its mean and SD. For comparison, Figure 3a also includes the true mean of the original q_c profile. Figure 3a shows that the mean of the N_B μ_{qc} is 1.87, which only has a relative difference of 3.6% when compared with that of the original q_c profile. In addition, using the N_B = 1000 values of μ_{qc}, its 95% CI can be obtained. The 95% CI of μ_{qc} is constructed as the range between the 2.5th and 97.5th percentile of the N_B = 1000 μ_{qc} values, which is sorted in ascending order. For clear presentation, the 95% CI of μ_{qc} is summarized in Table 1, which shows that the 95% CI of μ_{qc} covers the true mean of the original q_c profile. The consistency between the mean of μ_{qc} obtained from M = 11 measurements using the proposed method and the true mean of the original q_c profile demonstrates that the proposed method works reasonably well. This point is further supported by comparing the statistics of SD of q_c and the true SD of the original q_c profile, which is summarized in Figure 3b and Table 1. Observations similar to the q_c case are obtained in the f_s case (See Figure 4 and Table 1), which once again demonstrates that the proposed method is efficient in quantifying the statistical uncertainty associated with the estimated mean and SD of geotechnical properties even only sparse measurements are available.

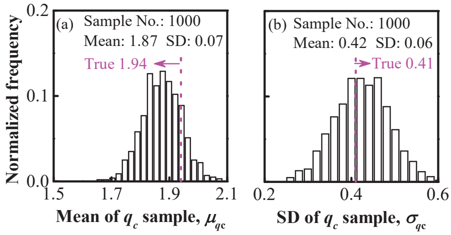

Figure 3. Histogram of mean and SD of q_c, i.e., μ_{qc} and σ_{qc}.

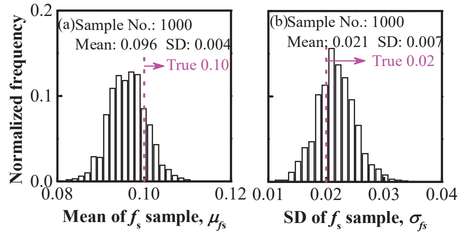

Figure 4. Histogram of mean and SD of f_s, i.e., μ_{fs} and σ_{fs}.

3.2 Auto-correlation and cross-correlation

Note that the proposed method is able to preserve the auto-correlation embedded in the geotechnical properties, with explicit consideration of the cross-correlation amongst different properties, which is investigated in this subsection. The auto-correlation of a geotechnical property profile may be quantified by the semi-variogram (SV) of the profile (e.g., Baecher and Christian 2003). For example, Figure 5a shows the SV of the original q_c profile by a bold solid line. Similarly, the SV for each of the N_B = 1000 q_c sample is shown by a gray line, ten of which are shown in Figure 5a. This leads to 1000 gray lines, using which the average and 95% CI of SV at each lag distance are calculated and shown in Figure 5a by a dashed line and two dotted lines, respectively. Figure 5a shows that the dashed line is consistent with the bold solid line, and the bold solid line falls within the 95% CI defined by the dotted lines. A similar observation can be obtained in Figure 5b, where the SV for f_s samples is explored. These observations suggest that the auto-correlation in the original q_c and f_s profile is statistically preserved. In addition, the cross-correlation coefficient $\rho_{qc,fs}$ between each q_c and f_s pair is calculated, leading to N_B = 1000 $\rho_{qc,fs}$. Figure 5c plots the histogram of $\rho_{qc,fs}$, and it also includes the $\rho_{qc,fs}$ estimated directly from sparse measurements on q_c, f_s using Eq. (6). Figure 5c shows that the mean of N_B $\rho_{qc,fs}$ is identical to the one estimated using Eq. (6), which indicates that the cross-correlation between q_c and f_s is preserved. In addition, the proposed method is able to quantify the statistical uncertainty associated the estimated $\rho_{qc,fs}$, which is 0.11 in terms of SD.

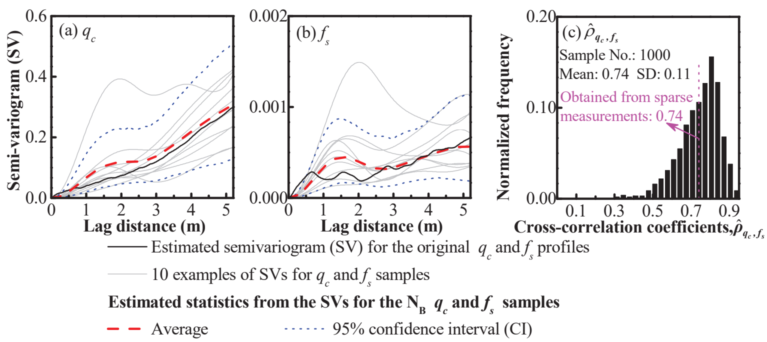

Figure 5. (a) and (b) for semivariogram of N_B = 1000 q_c and f_s samples, respectively, and (c) for a histogram of cross-correlation coefficient between q_c and f_s samples

3.3 Comparison with commonly used bootstrap methods

In this subsection, the performance of the proposed method is compared with the commonly used bootstrap method (e.g., Efron and Tibshirani 1993). Consider, for example, the same set of measurements in Figure 1. Then, the statistics of mean and SD of q_c and f_s can be obtained, which are summarized in Table 1. Table 1 shows that the 95% CI obtained from the bootstrap method also covers the true mean and SD of the original q_c and f_s profile, which is similar to that shown in the proposed method. However, the length of the 95% CI in the bootstrap method is significantly larger than that in the proposed method, as shown in Table. Consider, for example, the length of the 95% CI of μ_{qc}. The length in the bootstrap method is calculated 2.13-1.67 = 0.46, which is 70% larger than the one in the proposed method (i.e., 2.01-1.74 = 0.27). Similar observations can also be obtained for other estimators, such as the SD of q_c, mean and SD of f_s. This implies that the statistical uncertainty quantified by the proposed method is smaller than that by the commonly used bootstrap method (e.g., Efron and Tibshirani 1993). This may be attributed to the fact that auto-correlation of geotechnical properties are preserved in the method proposed in this paper.

4 Summary and Conclusions

In this paper, a Bayesian compressive sampling-based method was proposed to quantify the statistical uncertainty associated with mean, standard deviation and cross-correlation of geotechnical properties from sparse measurements. The proposed method automatically preserves the spatial auto-correlation of geotechnical properties, with explicit consideration of cross-correlation among different geotechnical properties. Furthermore, the assumption of probability distribution for measurements, which are often required in analytical methods is not needed in the proposed method. In this paper, equations were developed and the method was illustrated using a set of real CPT data. The results showed that the quantified statistical uncertainty associated with the mean, SD and cross-correlation is reasonable. In addition, the proposed method was compared to commonly used bootstrap methods, and the results showed that the statistical uncertainty quantified by the proposed method is smaller than that in the bootstrap method, since the auto-correlation is explicitly considered in the proposed method.

Acknowledgments

The authors would like to thank the members of the TC304 Committee on Engineering Practice of Risk Assessment & Management of the International Society of Soil Mechanics and Geotechnical Engineering for developing the database 304dB used in this study and making it available for scientific inquiry. We also wish to thank Dr. Armin W. Stuedlein for contributing this database to the TC304 compendium of databases. The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 9042331 (CityU 11225216) and Project No. 8779012 (T22-603/15N)). The financial support is gratefully acknowledged.

References

Ang, A. and Tang, W.H. (2007). *Probability Concepts in Engineering: Emphasis on Applications to Civil & Environmental Engineering 2nd ed*, Wiley, New York, 311-313.

Baecher, G. B. and Christian J. T. (2003). *Reliability and Statistics in Geotechnical Engineering*, John Wiley & Sons, Chichester, U.K.

Ching, J., Li, D. Q., and Phoon, K. K. (2016). Chapter 4: Statistical characterization of multivariate geotechnical data. *Reliability of Geotechnical Structures in ISO2394*, CRC Press/Balkema, 89-126.

Ching, J. and Phoon, K. K. (2017). Characterizing uncertain site-specific trend function by sparse Bayesian learning. *Journal of Engineering Mechanics*, 143(7), 04017028.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Fenton, G.A. and Griffiths, D.V. (2003). Bearing-capacity prediction of spatially random c - ϕ soils. C*anadian Geotechnical Journal*, 40, 54-65.

Ji, S., Dunson, D., and Carin, L (2009). Multitask compressive sensing. *IEEE Trans. on Signal Processing*, 57(1), 92-106.

Kenney, J. F. and Keeping, E. S (1951). *Mathematics of Statistics*, Part 2, 2nd ed. Van Nostrand, Princeton, New Jersey.

Luo, Z., Atamturktur, S., and Juang, C.H. (2013). Bootstrapping for characterizing the effect of uncertainty in sample statistics for braced excavations. *Journal of Geotechnical and Geoenvironmental Engineering*, 139(1), 13-23.

Phoon, K.K. (2008). *Reliability-Based Design in Geotechnical Engineering*: C*omputations and Applications*, Taylor and Francis, UK.

Phoon, K.K. (2016). Role of reliability calculations in geotechnical design. *Georisk*, 11(1), 4-21.

Stuedlein, A.W., Kramer, S.L., Arduino, P., and Holtz, R. D. (2012). Geotechnical characterization and random field modeling of desiccated clay. *Journal of Geotechnical and Geoenvironmental Engineering*, 138(11), 1301-1313.

Wang, Y. and Akeju, O. V. (2016). Quantifying the cross-correlation between effective cohesion and friction angle of soil from limited site-specific data. *Soils and Foundations*, 56(6), 1055-1070.

Wang, Y. and Aladejare A. E. (2016). Bayesian characterization of correlation between uniaxial compressive strength and Young's modulus of rock. *International Journal of Rock Mechanics and Mining Sciences*, 85, 10-19.

Wang, Y., Zhao, T., and Cao, Z. (2015). Site-specific probability distribution of geotechnical properties. *Computers and Geotechnics*, 70, 159-168.

Wang, Y. and Zhao, T. (2016). Interpretation of soil property profile from limited measurement data: a compressive sampling perspective. *Canadian Geotechnical Journal*, 53(9), 1547-1559.

Wang, Y. and Zhao, T. (2017). Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. *Géotechnique*, 67(6), 523-536.

Zhao, T. and Wang, T. (2018). Simulation of cross-correlated random field samples from sparse measurements using Bayesian compressive sensing. *Mechanical Systems and Signal Processing*, 112, 384-400.

Zhao, T., Montoya-Noguera, S., Phoon, K.K., and Wang, Y. (2018). Interpolating spatially varying soil property values from sparse data for facilitating characteristic value selection. *Canadian Geotechnical Journal*, 55(2), 171-181.