Proc. of the 6th Intl. Symposium on Reliability Engineering and Risk Management (6ISRERM)
31 May – 1 June 2018, Singapore
Editor(s) Xudong Qian, Sze Dai Pang, Ghim Ping Raymond Ong, Kok-Kwang Phoon

SELECTION OF CHARACTERISTIC VALUE FROM SPATIALLY VARYING BUT SPARSELY MEASURED GEOTECHNICAL DATA USING BAYESIAN COMPRESSIVE SAMPLING

TENGYUAN ZHAO[1] and YU WANG[2]

[1] *Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China. E-mail:* tengyzhao2@cityu.edu.hk
[2] *Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China. E-mail:* yuwang@cityu.edu.hk

Characteristic value of geotechnical properties is essential for probability-based geotechnical design codes (e.g., Eurocode 7). In practice, it is generally selected by engineers based on sparse geotechnical data and assisted by engineering experience and judgment. Due to subjective nature of individual judgment, the characteristic values derived by different engineers may vary greatly, even from the same dataset. The problem becomes more challenging when the spatially varying but auto-correlated pattern of geotechnical properties is considered. To address this issue, a novel method based on Bayesian compressive sampling (BCS) is presented for facilitating characteristic values selection from sparse measurement data. The method is illustrated and validated by a series of numerical examples. The results show that the method performs reasonably well, and the spatially varying but auto-correlated pattern of geotechnical properties is explicitly considered in a rational manner.

Keywords: Compressive sensing (CS), reliability-based design (RBD), spatial data, limit state design

1. Introduction

Uncertainties, such as spatial variability of soil properties, greatly affect geotechnical design or analysis (e.g., Phoon and Kulhawy 1999; Baecher and Christian 2003). This then motivates development of many probability-based geotechnical design codes (e.g., Eurocode 7) in the past three decades. For these codes, such as Eurocode 7, characteristic value of soil properties is a key component to ensure that the intended design can achieve a pre-specified target reliability level. Characteristic values of soil properties are generally determined by engineers from sparse geotechnical data. As the number of measured data is usually limited, engineering experience and judgment are often used to assist in characteristic value selection (e.g. Orr, 2017). Due to the subjective nature of experience and judgment, the characteristic values derived by different engineers, however, may vary greatly, even from the same dataset (e.g., Bond and Harris 2008).

For example, Bond and Harris (2008) showed three case studies where about one hundred engineers were asked for characteristic values selection from the same set of data based on Eurocode 7. It is shown that the selected characteristic value varied greatly, with the maximum characteristic value obtained being about three to five times larger than the minimum one. Therefore, more guidance is needed to select characteristic values in an objective manner, as suggested by Orr (2017). One possible solution is to resort to statistical methods, which have been increasingly used in geotechnical engineering (e.g., Baecher and Christian 2003; Zhang et al. 2004; Ching and Phoon 2011; Luo et al. 2012; Wang and Cao 2013). However, most of the available statistical methods for determining the characteristic values of geotechnical parameters focus on point statistics, such as mean and coefficient of variation for a previously defined homogeneous soil layer, and the spatially auto-correlated pattern of soil properties is often ignored (e.g., Pohl 2011; Wang et al. 2015; Orr 2017). Note that when the spatially varying but auto-correlated patterns of geotechnical properties is ignored, characteristic values determined might be biased and unrealistic, which may further lead to undesirable designs.

This paper aims to address these two issues by presenting a statistical procedure for characteristic value selection of geotechnical parameters from sparse measurement data, in a rational manner, with explicit consideration of the spatially varying but auto-correlated pattern of soil properties. The statistical procedure uses Bayesian compressive sampling (BCS) to probabilistically reconstruct a soil property profile from sparsely measured data, and provide quantified statistical uncertainty in terms of confidence interval (CI) associated with the interpreted profile (Wang & Zhao, 2017). The CI obtained in the BCS method has a clear statistical meaning: the confidence level α for a CI from the BCS is the expected coverage proportion (CP), i.e. fraction, of the complete profile that falls within the CI, if all data points along depth can be measured to provide the complete profile. The interpretation of CI may be used to facilitate characteristic values selection in practice from a statistical point of view.

2. Review of Bayesian Compressive Sampling (BCS)

Bayesian compressive sampling/sensing (BCS) is a coupling of Bayesian methods and compressive sampling/sensing (CS) to reconstruct a signal ($\boldsymbol{f}$) from sparse measurements on $\boldsymbol{f}$, i.e., $\boldsymbol{y}$ (e.g., Ji et al. 2008). A signal $\boldsymbol{f}$ is defined as variation of a quantity (e.g., soil property) with time or space (e.g., depth). BCS and CS utilize the compressibility in many real-world signals (e.g., Candès & Wakin 2008). The term "compressibility" means that $\boldsymbol{f}$ can be concisely represented as a weighted summation of a proper basis functions (e.g., Daubechies 16 wavelet function). In math, $\boldsymbol{f}=\mathbf{B}\boldsymbol{\omega}$. $\boldsymbol{f}$ is an N-length real-valued column vector; $\mathbf{B}$ is an N×N orthonormal matrix composed of columns of pre-specified basis functions; and $\boldsymbol{\omega}$ is the N-length weight coefficient vector corresponding to columns of $\mathbf{B}$. Due to the compressibility of signals, most entries in $\boldsymbol{\omega}$ are very near to zero. Therefore, $\boldsymbol{f}$ can be properly reconstructed if the non-trivial entries in $\boldsymbol{\omega}$ are identified and estimated from $\boldsymbol{y}$, a column vector with a length of M (M<<N) through $\boldsymbol{y}=\boldsymbol{\Psi f}=\boldsymbol{\Psi}\mathbf{B}\boldsymbol{\omega}=\mathbf{A}\boldsymbol{\omega}$. $\mathbf{A}=\boldsymbol{\Psi}\mathbf{B}$ and $\boldsymbol{\Psi}$ are both M×N matrices, where $\boldsymbol{\Psi}$ represents the locations of components $\boldsymbol{y}$ in $\boldsymbol{f}$ (e.g., Wang and Zhao 2016). The non-trivial coefficients $\boldsymbol{\omega}$ is obtained by solving $\boldsymbol{y}=\mathbf{A}\boldsymbol{\omega}$, which leads to an approximation of $\boldsymbol{\omega}$, i.e., $\boldsymbol{\omega}_s$. All components of $\boldsymbol{\omega}_s$ are zero except of the several non-trivial ones. Then, signal $\boldsymbol{f}$ is approximated as $\hat{\boldsymbol{f}}=\mathbf{B}\boldsymbol{\omega}_s$.

When $\boldsymbol{y}$ is sparse and limited, $\boldsymbol{\omega}_s$ estimated from $\boldsymbol{y}$ might not be accurate and contain significant statistical uncertainty, which may be quantified in a Bayesian framework (e.g., Wang and Zhao 2017). The quantified uncertainty in $\boldsymbol{\omega}_s$ would lead to quantified uncertainty associated with $\hat{\boldsymbol{f}}$. It has been shown that $\boldsymbol{\omega}_s$ follows a multivariate Students' t distribution, with a mean of $\boldsymbol{\mu}_{\omega_s}$, degree of freedom of $2c_n$, and a covariance matrix of $\mathbf{COV}_{\omega_s}$ (e.g., Wang and Zhao 2017). As $\hat{\boldsymbol{f}}=\mathbf{B}\boldsymbol{\omega}_s$, $\hat{\boldsymbol{f}}$ also follows a multivariate Student's t distribution with a degree of freedom of $2c_n$ (e.g., Fenton and Griffiths 2008). The mean $\boldsymbol{\mu}_{\hat{f}}$ and covariance $\mathbf{COV}_{\hat{f}}$ of $\hat{\boldsymbol{f}}$ are expressed as (e.g., Wang and Zhao 2017)

$$\begin{aligned}\boldsymbol{\mu}_{\hat{f}} &= \mathrm{E}(\hat{\boldsymbol{f}}) = \mathrm{E}(\mathbf{B}\boldsymbol{\omega}_s) = \mathbf{B}\,\mathrm{E}(\boldsymbol{\omega}_s) = \mathbf{B}\boldsymbol{\mu}_{\omega_s} \\ \mathbf{COV}_{\hat{f}} &= \mathrm{E}[(\hat{\boldsymbol{f}}-\boldsymbol{\mu}_{\hat{f}})(\hat{\boldsymbol{f}}-\boldsymbol{\mu}_{\hat{f}})^{\mathrm{T}}] = \mathbf{B}\mathbf{COV}_{\omega_s}\mathbf{B}^{\mathrm{T}}\end{aligned} \tag{1}$$

where "E(·)" represents expectation. Note that the mean $\boldsymbol{\mu}_{\hat{f}}$ represents the best estimate of $\hat{\boldsymbol{f}}$; while diagonal elements of $\mathbf{COV}_{\hat{f}}$ represent variance of elements of $\hat{\boldsymbol{f}}$ and they quantify the statistical uncertainty. Using Eq. (1), the confidence interval (CI) at a specified confidence level

α can be constructed and the meaning of CI is presented and demonstrated in the next section, which can be used to facilitate the characteristic value selection.

3. Coverage Proportion of the BCS Confidence Interval Profiles

3.1 Construction of CI Profiles from BCS

As $\hat{\boldsymbol{f}}$ follows a multivariate Student's t distribute, CI of the *i*-th element of $\hat{f}_i$ is expressed as (e.g., Zhao et al. 2018):

$$\mathrm{CI}_{\alpha}^{i} = \mu_{\hat{f}_i} \pm t_{(1-\alpha)/2,2c_n} \sqrt{(2c_n - 2)/2c_n} \sqrt{\mathrm{COV}_{\hat{f}_{i,i}}} \tag{2}$$

where $\mu_{\hat{f}_i}$ represents *i*-th element of $\boldsymbol{\mu}_{\hat{\boldsymbol{f}}}$; $\mathrm{COV}_{\hat{f}_{i,i}}$ represents the *i*-th diagonal element of $\mathbf{COV}_{\hat{\boldsymbol{f}}}$. Note that $\hat{f}_i$ varies spatially such as with depth, therefore CI_{α}^{i} also varies with depth. CI_{α}^{i} (i = 1, 2, …, N) constitutes two column vectors, denoted as $\mathbf{CI}_{\alpha}$. Zhao et al. (2018) have shown that the $\mathbf{CI}_{\alpha}$ has a clear statistical meaning: the corresponding confidence level α for a CI from the BCS is the expected coverage proportion (CP), i.e. fraction, of the complete profile that falls within the $\mathbf{CI}_{\alpha}$, if all data points along depth can be measured to provide the complete profile. The CP at level α, denoted as CP_{α}, is expressed as:

$$\mathrm{CP}_{\alpha} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} [\mathrm{I}(\hat{f}_i \in \mathrm{CI}_{\alpha}^{i})] \tag{3}$$

where "I(·)" is the indicator function. I(·) =1 if $\hat{f}_i$ (i = 1, 2, …, N) is within the upper and lower bounds of CI_{α}^{i}, and otherwise, I(·) = 0. For example, given the measurement data shown in Figure 1a to Figure 1c by open squares as input, the best estimate and 95% CI profiles are obtained from the BCS and shown in Figure 1a to 1c by a dashed line and a pair of dotted lines, respectively. Then, using Eq. (3), the $\mathrm{CP}_{95\%}$ is evaluated as 94.5%, 97.9% and 95.7%, respectively, for the cases shown in Figure 1a to 1c. All $\mathrm{CP}_{95\%}$ is very close to α = 0.95, with differences less than around 3%.

3.2 Probability Distribution of CP_{α} for BCS CI_{α} Profiles

Note that the soil property X profile shown in Figure 1a to 1c are only three realizations of a stationary Gaussian random field with a constant mean of μ_X = 30 and standard deviation σ_X=2. The auto-correlation of X at different depths is quantified by an exponential function $\rho_{i,j} = \exp(-\delta/\lambda_c)$, where $\delta = |d_{xi}-d_{xj}|$ represent the distance between X at depths d_{xi} and d_{xj}. λ_c represents the correlation length and it is taken 2m in this example. To further evaluate the meaning of CI from the BCS method, 1000 random field samples (RFSs) in total are generated from which 20 measurement data are extracted as sparsely measured data. This leads to 1000 sets of sparse data, and 1000 sets of $\mathbf{CI}_{\alpha}$ and CP_{α} for a given confidence level α using BCS method. Subsequently, using the 1000 CP_{α}, the distribution of CP_{α} in terms of box-and-whiskers plot, are constructed as shown in Figure 2a. The box represents the inter-quartile range (IQR) calculated as IQR = 75%-25% percentiles; while the whiskers represent the maxima and minima values within 1.5IQR. The maximum and minimum values of 1000 CP_{α} for a given α are shown by crosses in Figure 2a. Figure 2a also includes the mean of 1000 CP_{α} as open diamonds and a 1:1 dashed line. It is evident that all mean CP_{α} values at different α levels are close to the 1:1 line. This shows that the $\mathbf{CI}_{\alpha}$ obtained from the BCS method may be statistically interpreted as the

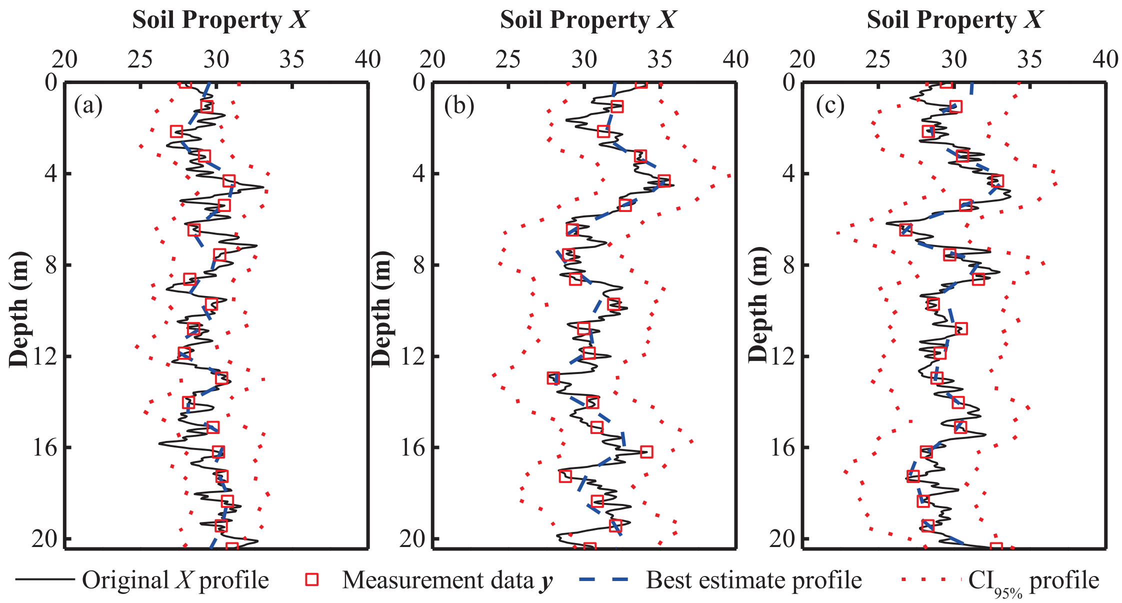

Figure 1 Three simulated X profiles and those reconstructed from BCS using M = 20 measurement data

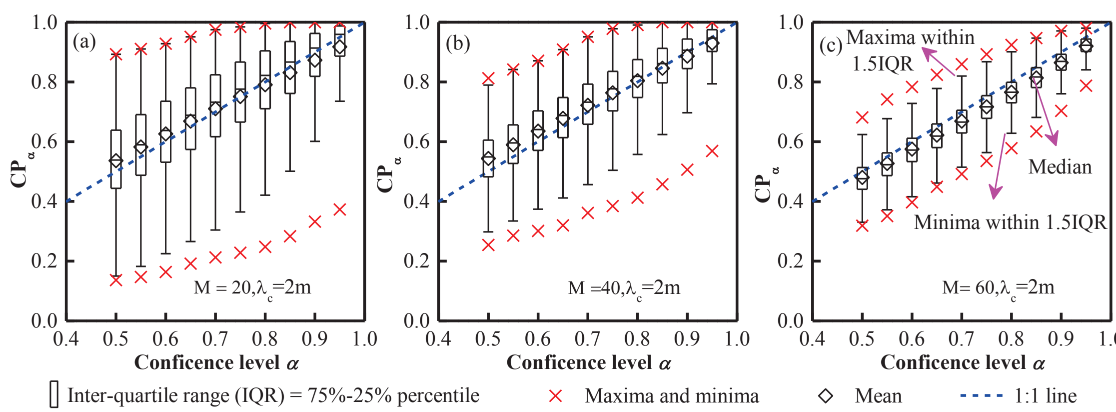

Figure 2 Box-and-whiskers plot for the coverage proportion (CP_α) for different α values given different M (after Zhao et al. 2018)

upper and lower bounds of an interval within which a spatially varying soil property X profile falls with an expected CP (or fraction) of α. Detailed exploration on the meaning of $\mathbf{CI}_\alpha$ is carried out by Zhao et al. (2018), and the effect of number of measurement data M and correlation length are summarized below.

3.3 Effect of M and Correlation Length on CP_α

In this subsection, the BCS method is repeated with different M scenarios, such as M = 10 to 60 with an increment of 10 points, to further explore the effect of M on the interpretation of $\mathbf{CI}_\alpha$. As M increases, more local variations of the original profiles are captured by the best estimate profile from the BCS method (Wang and Zhao 2017). Moreover, the region defined by the bounds of CI_α become generally narrower with the increase of M. This shows that the statistical uncertainty involved in the interpretation of soil property profile is effectively reduced with the increase of number of measurement data. The narrower region defined by the bounds of CI_α with larger M, however, would not greatly affect the interpretation of $\mathbf{CI}_\alpha$, as shown as follows. For each M scenario, 1000 sets of $\mathbf{CI}_\alpha$ and CP_α for a given confidence level α are obtained following

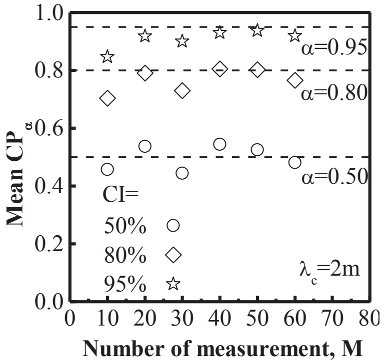

Figure 3 Effect of M (after Zhao et al. 2018)

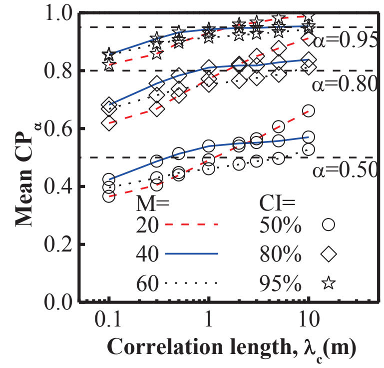

Figure 4 Effect of the λ_c (after Zhao et al. 2018)

the procedure presented in previous subsection. In such a case, box-and-whiskers plots for each M scenario can be constructed. For example, Figure 2a to 2c shows the M = 20, 40 and 60 scenarios, respectively. It shows that as M increases, the variability of CP_α values decreases, implying the increase of reliability of the interpretation of $\mathbf{CI}_\alpha$. The mean CP_α values, however, are quite close to the 1:1 line for all M scenarios, with relative difference between the average CP_α and α less than 15% in all the cases tested. For further evaluation, the mean CP_α values for α = 0.5, 0.80 and 0.95 under all M scenarios tested are summarized in Figure 3. It shows that the mean CP_α fluctuates around the α value, and is not greatly affected by M.

Different λ_c values with λ_c = 0.1, 0.3, 0.5, 1, 2, 3, 5 and 10m are further used to investigate the effect of correlation length λ_c on interpretation of $\mathbf{CI}_\alpha$. For each λ_c scenario, 1000 RFSs are generated from which 1000 sets of sparse measurement data are extracted. Then, following the procedure for the λ_c = 2m scenario, 1000 CP_α values are obtained for each λ_c scenario. Subsequently, box-and-whiskers plot for combinations of α and M under different λ_c scenario are constructed. The mean CP_α values for α = 0.50, 0.80 and 0.95 with M = 20, 40 and 60 under eight λ_c scenarios are summarized in Figure 4. It is observed that when λ_c is large, the average CP_α tends to be greater than α. Namely, a relatively large proportion of many RFSs tested falls inside the corresponding CI_α profiles. In contrast, the average CP_α tends to be smaller than α, when λ_c is relatively small. For λ_c ranging from 0.5m and 2m, which are common values for soil properties (e.g., Phoon and Kulhawy 1999), the relative difference between the mean CP_α and α is less than 15%. This relatively small differences suggest that the BCS method is robust and the statistical interpretation of BCS CI is reasonable: the confidence level α for a CI from the BCS is the expected coverage proportion (CP), i.e. fraction, of the complete profile that falls within the $\mathbf{CI}_\alpha$, if all data points along depth can be measured to provide the complete profile. As such, the bounds of $\mathbf{CI}_\alpha$ (e.g., lower bound of $\mathbf{CI}_\alpha$ for α = 90%) obtained from the BCS method might be used as characteristic values profile of soil properties when characteristic values are defined from a statistical point of view (e.g., the lower 5% percentile of a probability distribution). Note that this procedure, involving a limited number of measurement data, explicitly considers the spatially varying but auto-correlated pattern of soil properties. It has been used to determine characteristic values of friction angle profile, and details are provided by Zhao et al. (2018).

4. Summary and Conclusion

In this paper, a statistical procedure based on Bayesian compressive sampling (BCS) was presented to facilitate objective selection of characteristic values of soil properties from sparse measurement on soil properties. The BCS method is able to provide best estimate of soil properties profiles and quantified statistical uncertainty in terms of confidence interval (CI). The CI at a confidence level was interpreted as the expected coverage proportion (i.e. fraction) of the complete profile that falls within the CI, if all data points over the depth can be measured to provide the complete profile. This interpretation was illustrated and evaluated using a series of simulated data. The results show that this interpretation is statistically meaningful and may facilitate an objective selection of soil property characteristic values from sparse data from a purely statistical point of view.

Acknowledgments

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

Baecher, G. B., & Christian, J. T., *Reliability and Statistics in Geotechnical Engineering*. Wiley, New Jersey, 2003.

Bond, A., & Harris, A., *Decoding Eurocode 7.* Taylor & Francis, London and New York, 2008.

Candès, E. J., & Wakin, M. B. An Introduction To Compressive Sampling. *IEEE Signal Process. Mag.*, *25*(2), 21–30, 2008.

Ching, J., & Phoon, K. K. A quantile-based approach for calibrating reliability-based partial factors. *Struct. Saf.*, 33(4), 275-285, 2011.

Fenton, G. A., & Griffiths, D. V. *Risk assessment in geotechnical engineering*. Wiley, New Jersey, 2008.

Ji, S., Xue, Y., & Carin, L. Bayesian compressive sensing. *IEEE Trans. Signal Process.*, 56(6), 2346–2356, 2008.

Luo, Z., Atamturktur, S., & Juang, C. H. Bootstrapping for characterizing the effect of uncertainty in sample statistics for braced excavations. *J. Geotech. Geoenviron. Eng.*, 139(1), 13-23, 2012.

Orr, T. L. L., Defining and selecting characteristic values of geotechnical parameters for designs to Eurocode 7. *Georisk*, 11(1), 103–115, 2017.

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

Pohl, C. Determination of characteristic soil values by statistical methods. *ISGR Geotechnical Safety and Risk*, 427-434, 2011.

Wang, Y. and Cao, Z. J. Probabilistic characterization of Young's modulus of soil using equivalent samples. *Eng. Geol.*, 159, 106-118, 2013.

Wang Y, Zhao T, Cao Z., Determination of soil property characteristic values from standard penetration tests, ICASP12, Vancouver, Canada, 2015.

Wang, Y., & Zhao, T. Interpretation of soil property profile from limited measurement data : a compressive sampling perspective. *Can. Geotech. J.*, 53(9), 1547-1559, 2016.

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

Wang, Y., Zhao, T., and Phoon, K. K. Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation. *Can. Geotech. J.*, https://doi.org/10.1139/cgj-2017-0254, 2017.

Zhang, L., Tang, W. H., Zhang, L., & Zheng, J. Reducing uncertainty of prediction from empirical correlations. *J. Geotech. Geoenviron. Eng.*, 130(5), 526-534, 2004.

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