

INTERNATIONAL SOCIETY FOR SOIL MECHANICS AND GEOTECHNICAL ENGINEERING



This paper was downloaded from the Online Library of the International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). The library is available here:

<https://www.issmge.org/publications/online-library>

This is an open-access database that archives thousands of papers published under the Auspices of the ISSMGE and maintained by the Innovation and Development Committee of ISSMGE.

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.

How Many Samples Are Required in Geotechnical Site Investigation?

Zheng Guan¹, Yu Wang¹, and Tengyuan Zhao¹

¹Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong SAR.

E-mail: zhengguan3-c@my.cityu.edu.hk
yuwang@cityu.edu.hk
tengyzhao2@cityu.edu.hk

Abstract: Site investigation is an important part of a geotechnical engineering project. In practice, only a small portion (e.g., 1/100,000 or less) of the total volume of soil is examined, resulting in the significant uncertainty in the interpreted ground conditions. The interpretation error is significantly affected by the number of measurement data (i.e., sample size). Generally speaking, a large sample size provides detail information about the ground conditions, leading to increasing reliability of the interpreted ground conditions and reduced risk to engineering projects. However, increasing sample size requires increasing investigation cost, time and human resources. Therefore, it is important to develop a rational method for determining the required sample size which can provide a satisfactory level of accuracy for the interpreted geotechnical properties profiles. Sampling theory is a well-known topic in statistics, and it can be found in many textbooks on statistics. However, traditional sampling theory only deals with independent data and has limited application in geotechnical site investigation, because geotechnical properties are usually spatially varying and correlated. There is no scientific or quantitative method available for sample size determination in geotechnical site investigation considering spatial variation and correlation of geotechnical properties in existing design codes around world. To address these issues, this paper presents a rational method and statistical charts for sample size determination with consideration of spatial variation and correlation using Bayesian compressive sensing or sampling. Real laboratory test data is used to illustrate application of the statistical chart, and the method is shown to perform well.

Keywords: Geotechnical site investigation; sample size; Bayesian method; compressive sensing; random field.

1 Introduction

Site investigation is an important part of a geotechnical engineering project. Geotechnical engineers use measurement data from laboratory and in-situ tests to infer ground conditions (e.g., variation of soil properties with depth) and design profiles of geotechnical properties. Without the reliable design parameters from proper site investigation and data interpretation, geotechnical construction projects might be subjected to significant risk. In practice, only a small portion (e.g., 1/100,000 or less) of the total volume of soil is examined, resulting in the significant uncertainty in the interpreted ground conditions (e.g., Baecher and Christian 2003; Fenton and Griffiths 2008; Zhao et al. 2018). The interpretation error is significantly affected by the number of measurement data (i.e., sample size). Generally speaking, a large sample size provides detail information about the ground conditions, leading to increasing reliability of the interpreted ground conditions and reduced risk to engineering projects. However, increasing sample size requires increasing investigation cost, time and human resources. Therefore, it is important to develop a rational method for determining the required sample size which can provide a satisfactory level of accuracy for the interpreted geotechnical property profiles.

Currently, there is no quantitative method available to determine the sample size in engineering practice. Existing geotechnical design codes around the world only provide conceptual principles on sample size determination. Some previous studies on sample size determination have been reported in geotechnical literature (e.g., Protodyakonov 1969; Yamaguchi 1970; Gill et al. 2005; Jaksa et al. 2005; Goldsworthy et al. 2007; Cui et al. 2017), but they either are based on the assumption of independence of samples or aimed for quality control (e.g., Fenton et al. 2015; Liza et al. 2016), rather than the spatial distribution of geotechnical properties which is the aim of site investigation. Recently, Wang et al. (2018b) developed a rational method to determine sample size for geotechnical site investigation using random field simulation.

This paper presents a quantitative method and statistical chart to determine sample size in geotechnical site investigation considering the spatial variation and correlation of geotechnical properties based on the work by Wang et al. (2018). The relationship between the sample size and the interpretation accuracy is established using an extensive parametric study. The statistical chart developed from the parametric results can be used to determine the required sample size for different target accuracy of interpretation results and site conditions. Real laboratory test data is used to illustrate application of the statistical chart.

Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR)

Editors: Jianye Ching, Dian-Qing Li and Jie Zhang

Copyright © ISGSR 2019 Editors. All rights reserved.

Published by Research Publishing, Singapore.

ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0-IS12-4-cd

2 Development of a Relationship between Sample Size and Interpretation Accuracy Considering Spatial Correlation of Geotechnical Properties

In order to determine the samples size required for achieving a level of accuracy on the interpreted geotechnical profiles (e.g., spatial variation of a soil properties along depth), a relationship between sample size and the target level of accuracy should be established first. An extensive parametric study is performed to develop such a relationship. The overall philosophy of the parametric study is to compare a completely known soil property profile with the soil property profile interpreted from sparse measurements frequently encountered during site investigation. As shown in Figure 1, the parametric study involved three stages: (1) the relationship for a given soil property profile, (2) the relationship for a given site, and (3) the relationship for a wide variety of sites.

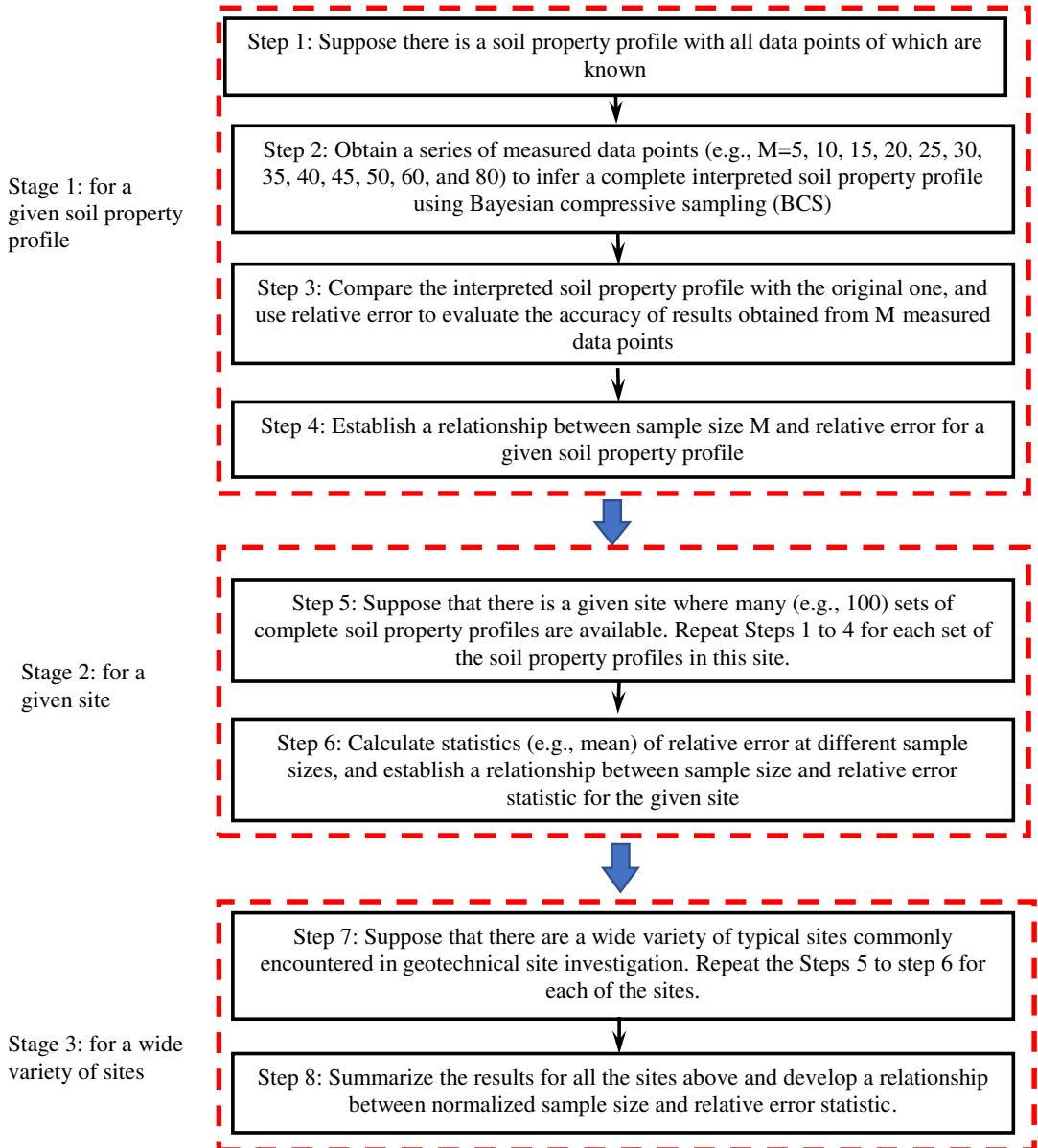


Figure 1. Framework for establishing a relationship between sample size and accuracy in geotechnical site investigation.

In Stage 1, one complete known soil property profile is firstly considered. To mimic the data measurement and interpretation in geotechnical site investigation, different number of measurement data points (i.e., $M=5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60,$ and 80) are taken from the completely known soil property profile and used as measurement data points for interpreting the corresponding soil property profile using interpolation method, such as Bayesian compressive sampling (BCS) (e.g., Wang and Zhao 2016; Wang and Zhao 2017), leading to a relationship between sample size M and relative error for this soil property profile. Subsequently, the parametric study is extended from a given soil property profile to a given site, where many (e.g., 100) sets of completely known soil property profile are available. The procedures used in Stage 1 are repeatedly performed for each set of soil property profile in this site, leading to many sets of the relationship between relative error and M of measured data points. Then, statistical analysis is performed to obtain the mean relative error, μ_{RE} , at different sample size. Thus, a relationship between the accuracy of interpreted soil property profile (i.e., μ_{RE}) and sample size for a given site is established (i.e., Stage 2).

To illustrate Stages 1 and 2, 50 sets of real CPT tip resistance q_c data are taken from Jaksa et al. (1999) and used in this study. CPT No. CD 5 is firstly adopted to illustrate Stage 1. Different number of measured data points (i.e. sample size), $M=5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 80,$ are taken from the CPT No. CD 5 and used as input to the BCS method for interpreting the complete q_c data profile. Figure 2 shows the interpretation results for $M=10, 15, 20, 60$. In this figure, the RE decreases from 8.29% to 1.90% as the number M of measured data points increases from 10 to 60. Figure 3 shows the RE for different number of measurement data points. It is observed that the RE decreases from about 10% to about 1% as the number of measurement data increases from 5 to 80. Figure 3 establishes a relationship between the RE and sample size for q_c data profile of CD 5. To illustrate Stage 2 of the framework, the Steps 1 to 4 in Stage 1 are repeated for each of the 50 sets of CPT data in the Keswick clay layer. The relationships between RE and M for the 50 sets of CPT data are summarized in Figure 4. Then, the mean, μ_{RE} , of the 50 RE values at each d_s value is obtained for the 50 sets of CPT data, leading to a relationship between the interpretation accuracy (i.e., μ_{RE}) and sample size (i.e., sampling interval d_s) shown in Figure 5.

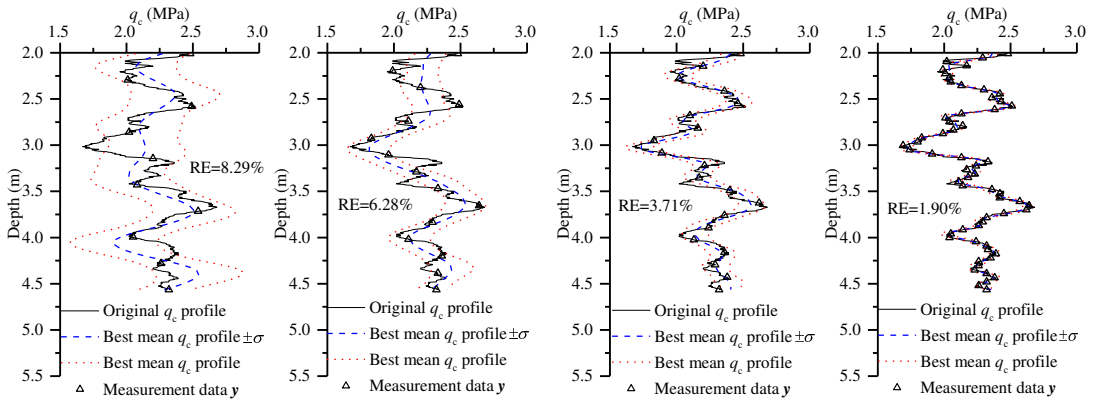


Figure 2. The interpreted q_c profiles from BCS method for different number M of measured data points: (a) $M=10$; (b) $M=15$; (c) $M=20$; (d) $M=60$ for CPT No. CD5 (After Wang et al. 2018).

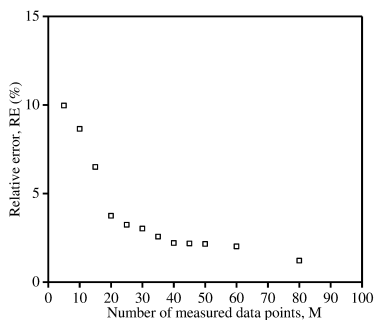


Figure 3. The relationship between number, M, of measured data points and relative error for CPT No. CD5 (After Wang et al. 2018).

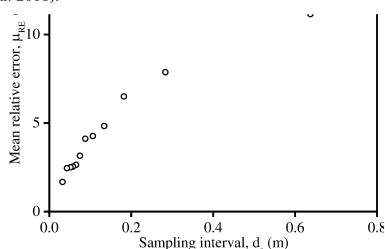


Figure 5. Relationship between mean relative error and sampling interval (After Wang et al. 2018).

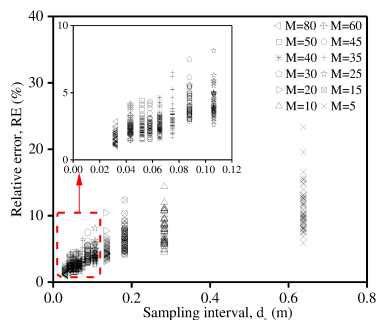


Figure 4. Variation of relative error with sampling interval from 50 sets of real CPT data (After Wang et al. 2018).

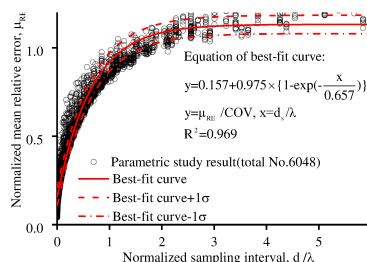


Figure 6. Dimensionless plot of normalized mean relative error versus normalized sampling interval (After Wang et al. 2018).

Finally, the parametric study is further extended from a typical site to a wide variety of sites commonly encountered in geotechnical site investigations (i.e., Stage 3). Random fields are adopted to simulate these completely known sites with the typical ranges of the mean, coefficient of variation, COV, and correlation length λ for a variety of soil properties. Table 1 summarizes the typical ranges of parameters used in parametric study. The procedures in Stage 2 are repeatedly performed for these simulated site conditions, resulting in many sets of the relationship between mean relative error μ_{RE} and sample size M for different site conditions. The parametric study results for μ_{RE} and M (i.e., sample interval d_s) are shown in Figure 6 with two dimensionless parameters, normalized mean relative error μ_{RE}/COV and normalized sampling interval d_s/λ . The parametric results provide a connection between the sample size and the interpretation accuracy. It should be noted that the presented method only deals with the sampling along depth, and this paper focuses on the one-dimensional cases.

Table 1. Summary of parameters used in parametric study (After Wang et al. 2018).

Parameter	Variation
Mean, μ	0.5, 20, 40, 50, 60, 80, 100, and 1000
COV (%)	1, 5, 10, 20, 30, 40, 50, and 100
Correlation length, λ (m)	0.5, 1, 2, 3, 5, 7, 9, 11, 13, 15, 18, and 20
Number, M, of measurement data points	5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, and 80
Thickness (m)	2.0, 5.0, 10.2, 20.4, 30.7, 40.9, and 51.1

3 Development of a Statistical Chart for Sample Size Determination

A statistical chart developed from the parametric results is shown in Figure 7. Similar to the Figure 6, Figure 7 is plotted with two dimensionless parameters, a target normalized relative error (i.e., RE_T/COV) in the horizontal axis and normalized sampling interval (i.e., d_s/λ) in vertical axis. Note that the target relative error, RE_T , is the desired interpretation accuracy that a user would like to have. When replacing the μ_{RE} by RE_T and switching the

vertical and horizontal axis of Figure 6, the statistical chart is generated. To use the statistical chart, the user shall firstly determine the target relative error which is based on the importance and risk of the engineering project. Then, parameters used for charactering the spatial variation and correlation of soil properties (i.e., coefficient of variation, COV and vertical correlation length, λ) shall be estimated based on existing knowledge of the site and soil properties concerned. Such existing knowledge can be obtained from the local engineering experience, geological maps and the published literatures (e.g., Phoon and Kulhawy 1999; Cao et al. 2016). When the RE_T/COV in the horizontal axis of Figure 7 is determined, the corresponding d_s/λ in the vertical axis is calculated. Finally, the d_s value is calculated using the d_s/λ obtained from the Figure 7 and the estimated λ . The required samples size M is determined from the d_s value and thickness of soil layer.

4 Application Example

To demonstrate application of the statistical chart, laboratory test result (i.e., natural water content) is used as illustrative example in this section. Consider, for example, determination of sample size for quantifying spatial variation of the natural water content, ω_n , at St James’s Park, London within a mainly sandy clayey silt layer from a depth of about 10m to 20m (Hight et al. 2002).

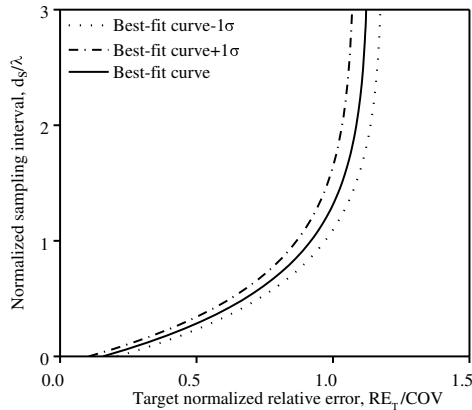


Figure 7. Statistical chart for sampling interval determination.

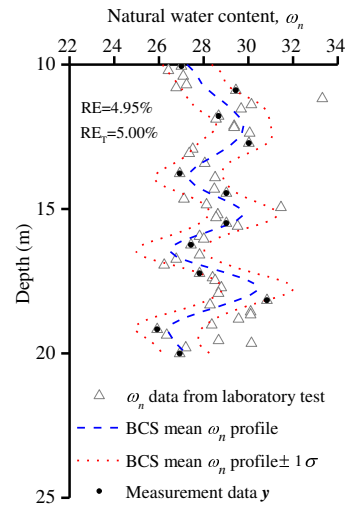


Figure 8. Comparison between the original and the profile interpreted from BCS method.

Suppose that the target relative error is 5%. The COV and vertical λ of natural water content in this sandy clayey silt are estimated as 14% and 6.0m, respectively. based on local experience (e.g., nearby site ω_n data). These ranges are also within the typical ranges of 7~46% for COV and 1.6~12.7m for vertical λ , respectively, for the sandy clayey silt natural water content data reported in literatures (e.g., Phoon and Kulhawy 1999). Then, the $RE_T/COV = 0.36$ is calculated. Using the best-fit curve in the statistical chart, $d_s/\lambda=0.15$ is calculated. Because $\lambda=6.0m$, the sampling interval $d_s=0.90m$ is estimated. The thickness of silty clay layer is 10.0m, and the number of measurement data points is estimated as $M=10/0.90+1\approx 12$. Subsequently, 12 samples within the sandy clayey silt layer (i.e., from a depth of about 10m to 20m) shall be obtained from a borehole with an equal sampling interval for carrying out natural water content tests in laboratory and quantifying spatial variation of the natural water content along depth.

Hight et al. (2002) happened to report that a total number of 52 water content value extents from a depth of 10.0m to 20.0m with a measurement interval of about 0.2m. Figure 8 shows the 52 ω_n values measured along depth by open triangles. For validating the performance of the statistical chart in this illustrative example, 12 ω_n values are taken from the 52 ω_n values with an approximately equal sampling interval, and they are shown in Figure 8 by solid circles. The 12 ω_n values obtained are used as the measurement data, i.e., input to BCS for interpolating a high-resolution natural water content profile with 256 data points. Figure 8 also shows the mean natural water content profile interpreted from BCS by a dashed line and the mean plus or minor one standard deviation profiles by two dotted lines, respectively. It is observed from Figure 8 that the dashed line follows a trend quite similar to that of the open triangles (i.e., all 52 ω_n values), and that most of the 52 ω_n data points (i.e.,

the open triangles) fall within the dotted lines. The relative error RE between the 52 original natural water content values and mean natural water content values at the corresponding depths obtained from BCS is calculated as 4.95%, which is very consistent with the target relative error of 5.00%. The consistency suggests that the sample size determined by the statistical chart (i.e., Figure 7) achieves the target level of accuracy for the results interpreted from the given number of measurements. In other words, the statistical chart performs well in the illustrative example for laboratory test data.

5 Conclusion

A quantitative method and statistical chart for sample size determination in geotechnical site investigation considering spatial variation and correlation of soil properties are presented in this paper. An extensive parametric study is performed to establish a quantitative relationship between the sample size and the level of accuracy of interpretation result. A statistical chart for sample size determination is subsequently developed from the parametric study result. Real water content data at St James's Park, London was used to illustrate application of the statistical chart. The statistical chart was shown to perform well in the illustrative example for laboratory test data.

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. 9042516 (CityU 11213117) and Project No. 8779012 (T22-603/15N)). The financial support is gratefully acknowledged.

References

- Baecher, G.B. and Christian, J.T. (2003). *Reliability and Statistics in Geotechnical Engineering*, John Wiley & Sons, Hoboken, New Jersey.
- Cao, Z., Wang, Y., and Li, D. (2016). Quantification of prior knowledge in geotechnical site characterization. *Engineering Geology*, 203, 107-116.
- Cui, J., Jiang, Q., Li, S., Feng, X., Zhang, M., and Yang, B. (2017). Estimation of the number of specimens required for acquiring reliable rock mechanical parameters in laboratory uniaxial compression tests. *Engineering Geology*, 222, 186-200.
- Fenton, G.A. and Griffiths, D.V. (2008). *Risk Assessment in Geotechnical Engineering*, John Wiley & Sons, New York, USA.
- Fenton, G.A., Liza, R., Lake, C.B., Menzies, W.T., and Griffiths, D.V. (2015). Statistical sample size for quality control programs of cement-based "solidification/stabilization". *Canadian Geotechnical Journal*, 52(10), 1620-1628.
- Gill, D.E., Corthésy, R., and Leite, M.H. (2005). Determining the minimal number of specimens for laboratory testing of rock properties. *Engineering Geology*, 78(1-2), 29-51.
- Goldsworthy, J.S., Jaksa, M.B., Fenton, G.A., Kaggwa, W.S., Griffiths, V., and Poulos, H.G. (2007). Effect of sample location on the reliability based design of pad foundations. *Georisk*, 1(3), 155-166.
- Hight, D.W., McMillan, F., Powell, J.J.M., Jardine, R.J., and Allenou, C.P. (2002). Some characteristics of London Clay. *Characterisation and Engineering Properties of Natural Soils (eds T. S. Tan, K. K. Phoon, D. W. Hight and S. Leroueil)*, pp. 851-908. Rotterdam, the Netherlands: Balkema.
- Jaksa, M., Goldsworthy, J., Fenton, G., Kaggwa, W., Griffiths, D., Kuo, Y., and Poulos, H. (2005). Towards reliable and effective site investigations. *Géotechnique*, 55(2), 109-121.
- Liza, R., Fenton, G.A., Lake, C.B., and Griffiths, D.V. (2016). An analytical approach to assess quality control sample sizes of cement-based "solidification/stabilization". *Canadian Geotechnical Journal*, 54(3), 419-427.
- Protodyakonov, M. (1969). Method of determining the strength of rocks under uniaxial compression. *Mechanical properties of rocks*, 1-8.
- Phoon, K.-K. and Kulhawy, F. H. (1999). Characterization of geotechnical variability. *Canadian Geotechnical Journal* 36(4), 612-624.
- 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.
- Wang, Y., Guan, Z., and Zhao, T. (2018). Sample size determination in geotechnical site investigation considering spatial variation and correlation. *Canadian Geotechnical Journal*, <https://doi.org/10.1139/cgj-2018-0474>.
- Yamaguchi, U. (1970). The number of test-pieces required to determine the strength of rock. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, Elsevier, 7, 209-227.
- 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.