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

QUANTIFICATION OF UNCERTAINTY IN SOIL STRATIGRAPHY BASED ON CONE PENETRATION TESTS

SHUO ZHENG[1], ZI-JUN CAO[1], DIAN-QING LI[1], and KOK-KWANG PHOON[2]

[1] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan University, 8 Donghu South Road, Wuhan 430072, P. R. China.*
[2] *Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07-03, 1 Engineering Drive 2, Singapore 117576, Singapore.*
E-mail: zhengshuo@whu.edu.cn; zijuncao@whu.edu.cn (corresponding author); dianqing@whu.edu.cn; kkphoon@nus.edu.sg

The uncertainty in soil stratigraphy identified using cone penetration test (CPT) data is usually unknown. This paper develops a Bayesian framework to probabilistically identify underground stratigraphy based on soil behavior type index (I_c) data. The proposed approach not only identifies the most probable soil layer boundaries with the consideration of spatial variability of I_c, but also quantifies the uncertainties in soil stratigraphy. It is illustrated and verified using simulated I_c data. Then, effects of layer thicknesses and statistical differences in I_c profiles of adjacent soil layers on the uncertainty of the soil layer boundaries are discussed using the simulated I_c data. Results show that the proposed approach properly identifies soil stratigraphy based on I_c data and, rationally, quantifies the identification uncertainty in the soil layer based on CPT data. The identification uncertainty in a soil layer boundary is affected by soil layer thicknesses and the statistical difference in I_c data within adjacent soil layers.

Keywords: Soil stratification, cone penetration test, soil behavior type index, Bayesian approach.

1 Introduction

Cone penetration test (CPT) has been widely used to determine the soil stratigraphy during geotechnical site investigation because it is rapid, repeatable, and economical, and provides nearly continuous measurements over the depth (Roberson 2009). In general, this consists of two major steps: (i) determine the soil type at each testing depth (i.e., soil classification) based on CPT measurements; and (ii) identify the number N and thicknesses (or boundaries) $\underline{H}_N = [H_1, H_2, \ldots, H_N]$ of soil layers based on the profile of the soil type. Among various CPT-based soil classification systems (e.g., Roberson 2009), the soil behavior type (SBT) index I_c is widely used, which, at different depths, varies spatially even for the same SBT soils. The spatial variability of I_c poses a profound challenge in identifying soil stratigraphy (i.e., determining N and $\underline{H}_N$) from a single profile of I_c with certainty. Soil stratigraphy provided by different engineers based on the same I_c profile might be inconsistent due to their different experience, expertise, and judgments. Several approaches have been developed to delineate soil stratigraphy using CPT data in an objective and quantitative way, such as clustering method (Hegazy and Mayne 2002), statistical analysis using modified Bartlett statistics (Phoon et al. 2003, 2004), wavelet transform modulus maxima method (Ching et al. 2015), and Bayesian methods (Cao and Wang 2013, Wang et al. 2013). These approaches are able to provide the "*best*" estimates of N and $\underline{H}_N$ in terms of prescribed criterion for soil stratification, but they provide little information on the uncertainty in estimated N and $\underline{H}_N$.

This paper develops a Bayesian framework for probabilistic soil stratification based on the profile of I_c, the inherent spatial variability of I_c along the depth is explicitly modelled by random fields, and the uncertainty in N and $\underline{H}_N$ estimated from the I_c profile is quantified by their posterior distributions. With the proposed Bayesian framework, effects of layer thicknesses and the statistical difference in I_c profiles of adjacent soil layers on the uncertainty of the soil layer boundary are discussed.

2 Bayesian Framework for Probabilistic Soil Stratification

For a given profile of I_c (i.e., $\underline{\xi}$), identification of soil stratigraphy under the proposed Bayesian framework is divided into two steps: (i) compare the soil stratification models with different numbers (e.g., N) of soil layers based on their conditional probabilities $P(N|\ \underline{\xi})$ given $\underline{\xi}$, and determine the most probable number of soil layers N^* among a number of possible N values; and (ii) evaluate the posterior distribution $P(\underline{H}_N|\ \underline{\xi},\ N)$ of soil layer thicknesses to quantify the uncertainty in $\underline{H}_N$ based on $\underline{\xi}$ for a given soil stratification model with N (e.g., $N = N^*$) soil layers, and determine the most probable thicknesses $\underline{H}^*_{N*} = [H_1^*, H_2^*, ..., H^*_{N*}]$ and internal boundaries. These two steps are introduced in the following two subsections.

2.1 *The most probable number of soil layers*

The number of soil layers contained in a profile of I_c is considered varying from 1 to a maximum value of N_{max}. Then, N is defined as a discrete random variable ranging from 1 to N_{max}. Using the Bayes' Theorem, $P(N|\ \underline{\xi})$ is written as (Cao and Wang 2013; Wang et al. 2013):

$$P\left(N|\underline{\xi}\right) = P\left(\underline{\xi}\mid N\right)P\left(N\right)/P\left(\underline{\xi}\right) \tag{1}$$

where $P(N)$ = prior probability of N reflecting the prior knowledge on N in the absence of CPT data; $P(\underline{\xi})$ is a normalizing constant and independent of N; $P(\underline{\xi}|\ N)$ = conditional probability of $\underline{\xi}$ given the soil stratification model with N layers, and it is frequently referred to as the "*evidence*" for the soil stratification model with N layers provided by $\underline{\xi}$. In the case of no prevailing prior knowledge on N, the N_{max} possible values (i.e., 1, 2, …, N_{max}) of N are considered having the same prior probability, i.e., $P(N) = 1/N_{max}$. Then, based on Eq. (1), $P(N|\ \underline{\xi})$ is proportional to the evidence $P(\underline{\xi}|\ N)$, which means that maximizing $P(\underline{\xi}|\ N)$ with respect to N leads to the maximum value of $P(N|\ \underline{\xi})$ and, hence, N^*.

2.2 *Uncertainty in soil layer thicknesses*

In this subsection, the number N of soil layers is a fixed value and is used as a condition for inferring $\underline{H}_N$ from $\underline{\xi}$ according to $P(\underline{H}_N|\ \underline{\xi},\ N)$. Within a Bayesian framework, $P(\underline{H}_N|\ \underline{\xi},\ N)$ is referred to as the posterior distribution of $\underline{H}_N$ based on $\underline{\xi}$, and it is expressed as (Cao et al. 2017):

$$P\left(\underline{H}_N\mid\underline{\xi},N\right) = P\left(\underline{\xi}\mid\underline{H}_N,N\right)P\left(\underline{H}_N\mid N\right)\Big/P\left(\underline{\xi}\mid N\right) \tag{2}$$

The $P(\underline{H}_N|\ \underline{\xi},\ N)$ in Eq. (2) quantifies the uncertainty in layer thicknesses $\underline{H}_N$ (or, equivalently, layer boundaries $\underline{D}_N$) of the soil stratification model with N layers based on both CPT data and prior knowledge. It involves the likelihood function $P(\underline{\xi}|\ \underline{H}_N,\ N)$, the prior distribution $P(\underline{H}_N|\ N)$, and a normalizing constant $P(\underline{\xi}|\ N)$ independent of $\underline{H}_N$ for a given N value.

In the case of no prevailing prior knowledge on soil layer thicknesses $\underline{H}_N = [H_1, H_2, \ldots, H_N]$, they can be considered uniformly distributed within a range from 0 to CPT sounding depth H, i.e., $0 < H_n < H$ for $n = 1, 2, \ldots, N$, and all the possible combinations of $\underline{H}_N$ are uniformly distributed within a N-1 dimensional simplex $\Omega = \{ \sum_{n=1}^{N} H_n = H, 0 < H_n < H\}$. Such a uniform distribution of $\underline{H}_N$ can be represented by a flat Dirichlet distribution, and is expressed as (Cao et al. 2017):

$$P\left(\underline{H}_N \mid N\right) = \Gamma(N) / H^{N-1} \tag{3}$$

where $\Gamma(\cdot)$ is the Gamma function evaluated at N; and H^{N-1} serves as a normalizing constant. As indicated by Eq. (3), $P(\underline{H}_N| N)$ is a constant for a given N value and testing depth H.

The likelihood function $P(\underline{\xi}| \underline{H}_N, N)$ quantifies information on $\underline{H}_N$ of the soil stratification model with N soil layers provided by $\underline{\xi}$. This study models the I_c profile by N mutually independent lognormal random fields $\underline{I_{cn}}(Z)$, $n = 1, 2, \ldots, N$, where I_c at different depths are spatially correlated lognormal random variables with a mean μ_n and standard deviation σ_n. Here, the correlation structure of lnI_c is taken as a single exponential correlation function with a scale of fluctuation of λ_n, which is frequently used to analyze CPT data (Phoon et al. 2003, 2004). Correspondingly, the profile of lnI_c (i.e., $\underline{\xi} = [\underline{\xi}_1, \underline{\xi}_2, \ldots, \underline{\xi}_N]$) obtained from the N soil layers are considered as a realization of the N random fields with model parameters $\underline{\theta}_n = [\mu_n, \sigma_n, \lambda_n]$, $n = 1, 2, \ldots, N$. Then, $P(\underline{\xi}| \underline{H}_N, N)$ is expressed as (Cao and Wang 2013; Wang et al. 2017):

$$P\left(\underline{\xi} \mid \underline{H}_N, N\right) = \prod_{n=1}^{N} P\left(\underline{\xi}_n \mid \underline{H}_N, N\right) \tag{4}$$

where $P(\underline{\xi}_n| \underline{H}_N, N)$, $n = 1, 2, ..., N$ is the likelihood function for the n-th soil layer. Using the Theorem of Total Probability, $P(\underline{\xi}_n| \underline{H}_N, N)$ is written as:

$$P\left(\underline{\xi}_n \mid \underline{H}_N, N\right) = \int P\left(\underline{\xi}_n \mid \underline{\theta}_n, \underline{H}_N, N\right) P\left(\underline{\theta}_n \mid \underline{H}_N, N\right) d\underline{\theta}_n \tag{5}$$

where $P(\underline{\xi}_n| \underline{\theta}_n, \underline{H}_N, N)$ is a joint Gaussian PDF of $\underline{\xi}_n$ for a given set of $\underline{\theta}_n$, $\underline{H}_N$ and N; and $P(\underline{\theta}_n| \underline{H}_N, N)$ is the prior distribution of $\underline{\theta}_n$ in the n-th soil layer for a given N soil layers with layer thicknesses equal to $\underline{H}_N$ and is simply taken as a joint uniform prior distribution of $\underline{\theta}_n$ defined by their typical ranges reported by Cao et al. (2017).

Solving the $P(\underline{\xi}| N)$ and $P(\underline{H}_N| \underline{\xi}, N)$ is a key step to determine the soil stratigraphy and its associated uncertainty. This is a non-trivial task in this study because of the discontinuity of the likelihood function with respect to $\underline{H}_N$, constraint relationship among soil layer thicknesses, and high-dimensional integral involved in the evidence and the posterior distribution, particularly as N and the number of data are relatively large, such as $N \geq 3$. The Bayesian Updating with Structural Reliability Method (BUS) using Subset Simulation (SuS) (Straub and Papaioannou, 2015; DiazDelaO et al. 2017) is adopted to, simultaneously, obtain $P(\underline{\xi}| N)$ and $P(\underline{H}_N| \underline{\xi}, N)$. For the sake of conciseness, detailed algorithm and implementing procedures of the BUS with SuS are not provided herein. Interested readers are referred to DiazDelaO et al. (2017) for details.

3 Illustrative example

3.1 *Baseline case*

The proposed approach has been applied to a real site of the NGES at Texas A&M University Cao et al. (2017). However, soil stratigraphy at a real site, where the actual soil layer boundaries are unknown, can only be inferred from site investigation data and prior knowledge. For illustration and validation, the proposed approach is applied to a virtual site in this section,

Table 1. Summary of simulated I_c profiles and standard deviations of the boundary depth

No. of Case	Thickness H_n(m)		μ_n		σ_n		λ_n		Standard deviation of D (m)
	Layer 1	Layer 2	Layer 1	Layer 2	Layer 1	Layer 2	Layer 1	Layer 2	
Baseline case	6	6	3.2	2.7	0.4	0.2	0.6	0.3	0.21
Case I	6	1	3.2	2.7	0.4	0.2	0.6	0.3	1.13
Case II	6	2	3.2	2.7	0.4	0.2	0.6	0.3	0.72
Case III	6	4	3.2	2.7	0.4	0.2	0.6	0.3	0.28
Case IV	6	8	3.2	2.7	0.4	0.2	0.6	0.3	0.20
Case V	6	10	3.2	2.7	0.4	0.2	0.6	0.3	0.16
Case VI	6	6	3.2	3.2	0.4	0.2	0.6	0.3	2.19
Case VII	6	6	3.2	2.7	0.4	0.4	0.6	0.3	0.33
Case VIII	6	6	3.2	2.7	0.4	0.2	0.3	0.3	0.32

where the actual soil layer boundaries are known and can be used to simulate I_c profile. As the baseline case shown in Table 1, the virtual site is comprised of two soil layers with the same thickness of 6m and the data are simulated at the interval of 0.05m. Correspondingly, there is an internal boundary located at depths of 6m (see the horizontal solid line in Figure 1(a)). The I_c profile in the two soil layers is represented by two one-dimensional and mutually independent lognormal random fields, and their random field parameters of I_c are shown in the Table 1. Soils in top layer mainly belong to SBT 3 (Clays: silty clay to clay), and soils in the second layer are of SBT 4 (Silt mixtures: clayey silt to silty clay) according to the soil classification system based on I_c (Robertson and Wride 1998). The boundaries between different SBTs based on the I_c are shown by vertical dashed lines in Figure 1(a).

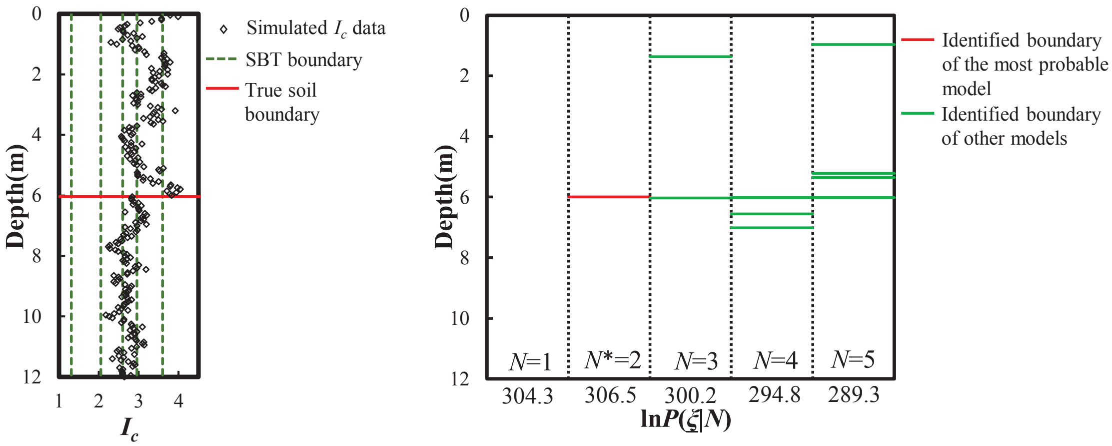
Consider, for example, that there are, to the maximum, 5 soil layers within the testing depth, i.e., $N_{\max} = 5$. When N equals to 1 and 2, Bayesian equations can be solved by direct numerical analysis integration with relative ease. As $N \geq 3$, the BUS with SuS is used due to the large number of possible combinations of $\underline{H}_N$. Figure 1(b) shows the soil stratification results obtained from the proposed approach for different N values at the virtual site. The maximum value (i.e., 306.5) of $\ln P(\underline{\xi}|\, N)$ occurs at $N^* = 2$, which is identical to the true number of soil layers at the site (see Figure 1). Figure 1(b) shows an evolution of layer identification as N increases from 1 to 5. For $N^* = 2$, the internal boundary identified from the proposed approach is close to the true boundary located at 6m, as shown by horizontal red solid line in column 2 of Figure 1(b). Effects of factors (including random field parameters and soil thicknesses) on probabilistic soil stratification are discussed in next subsection.

3.2 *Effect of layer thickness*

To explore the effect of layer thickness, five cases (i.e., cases I to V shown in Table 1) are considered in addition to the baseline case. In the five cases, the thicknesses of the 2-nd layer are 1, 2, 4, 8 and 10 m, respectively. As shown in Figures 2(a)-(e), simulated I_c data in the baseline case and cases I-V are represented with black solid lines and green dashed lines, respectively. The identified boundaries based on simulated I_c data in cases I-V for $N = 2$ are also shown by horizontal yellow dashed lines in Figures 2(a)-(e). The standard deviation of the boundary depth D representing the uncertainty of soil stratigraphy gradually decreases as the thickness of the 2-nd layer increases, as shown in the Table 1. As the CPT sounding depth in the 2-nd layer increases, more information is incorporated into the proposed Bayesian framework to determine the soil layer boundary. This leads to reduction of identification uncertainty in the layer boundary, yielding a more reliable estimation of the boundary location.

3.3 *Effect of the statistical difference of I_c profiles in adjacent soil layers*

To explore effects of the statistical difference of I_c profiles in adjacent soil layers, three



(a) Simulated I_c profile

(b) Boundaries and the logarithm of evidence for different N values

Figure 1. Soil stratification results in the baseline case

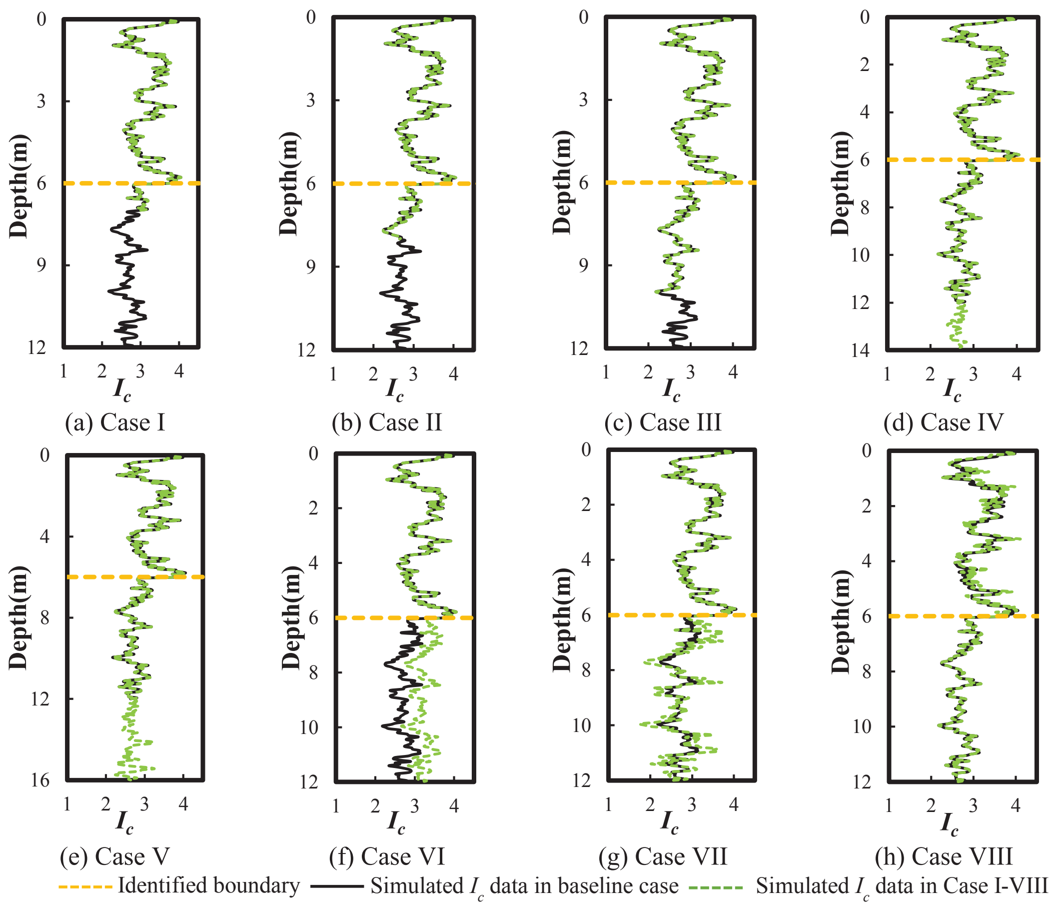


Figure 2. Simulated I_c profiles for cases I-VIII

additional cases are considered, i.e., cases VI-VIII in Table 1. As shown in Table 1, thicknesses of the two layers and the true boundary depth of 6m remain unchanged in cases VI-VIII, but the ratios μ_1/μ_2, σ_1/σ_2, and λ_1/λ_2 of random field parameters of the two layers are taken as 1, respectively. As shown in Figures 2(f)-(h), simulated I_c profiles in cases VI-VIII are represented by green dashed lines, and the identified boundaries based on them for $N = 2$ are also shown by horizontal yellow dashed lines. Comparing with the baseline case, the standard deviation of D increases in cases VI-VIII. This indicates that the identification uncertainty in soil layer boundary increases as the statistical difference of I_c profiles in the two layers decreases. As shown in Table 1, the effect of the mean value is more significant than standard deviation and scale of fluctuation of I_c.

4 Conclusions

This paper proposed a Bayesian framework for probabilistic soil stratification based on I_c, which was illustrated and verified using I_c data simulated at a virtual site. Effects of the layer thickness and the statistical difference in I_c profiles of adjacent soil layers on the identification uncertainty in the soil layer boundaries were discussed. Results showed that the proposed approach properly identifies the most probable soil stratigraphy and rationally quantifies the uncertainty in soil stratigraphy. Increasing the CPT sounding depth in a layer can provide information for reducing the identification uncertainty in the estimated soil layer boundary and improving its reliability using the proposed approach. The identification uncertainty provided by the proposed approach rationally reflects the statistical difference in I_c profiles of adjacent soil layers.

Acknowledgments

This work was supported by the National Key R&D Program of China (Project No. SQ2016YFSF060351), the National Natural Science Foundation of China (Project Nos. 51679174, 51409196, 51579190). The financial support is gratefully acknowledged.

References

Cao, Z. J. and Wang, Y., Bayesian Approach for Probabilistic Site Characterization Using Cone Penetration Tests, *J. Geotech. Geoenviron. Eng.*, 139(2), 267-276, Feb, 2013.

Cao, Z. J., Zheng, S., Li, D. Q. and Phoon, K.K. 2017. Bayesian Identification of Soil Stratigraphy based on Soil Behavior Type Index, Submitted to *Can. Geotech. J.*

Ching, J., Wang, J. S., Juang, C. H. and Ku, C. S., Cone Penetration Test (CPT)-Based Stratigraphic Profiling Using the Wavelet Transform Modulus Maxima Method, *Can. Geotech. J.*, 52(12), 1993-2007, Dec, 2015.

Diazdelao, F. A., Garbuno-Inigo, A., Au, S. K. and Yoshida, I., Bayesian Updating and Model Class Selection with Subset Simulation, *Comput. Meth. Appl. Mech. Eng.*, 317 (Supplement C), 1102-1121, Apr, 2017.

Hegazy, Y. A. and Mayne, P. W., Objective Site Characterization Using Clustering of Piezocone Data, *J. Geotech. Geoenviron. Eng.*, 128(12), 986-996, Dec, 2002.

Phoon, K. K., Quek, S. T. and An, P., Identification of Statistically Homogeneous Soil Layers Using Modified Bartlett Statistics, *J. Geotech. Geoenviron. Eng.*, 129(7), 649-659, Jul, 2003.

Phoon, K. K., Quek, S. T. and An, P., Geostatistical Analysis of Cone Penetration Test (CPT) Sounding Using, *Can. Geotech. J.*, 41(41), 356-365, Apr, 2004.

Robertson, P. K., Interpretation of Cone Penetration Tests - A Unified Approach, *Can. Geotech. J.*, 46(46), 1337-1355, Nov, 2009.

Robertson, P. K. and Wride, C. E., Evaluating Cyclic Liquefaction Potential Using the Cone Penetration Test, *Can. Geotech. J.*, 37(3), 270-271, Jun, 1998.

Straub, D., and Papaioannou, I., Bayesian Updating with Structural Reliability Methods. *J. Eng. Mech.*, 141(3), 04014134, Mar, 2015.

Wang, Y., Huang, K. and Cao, Z. J., Probabilistic Identification of Underground Soil Stratification Using Cone Penetration Tests, *Can. Geotech. J.*, 50(7), 766-776, Jul, 2013.

Wang, Y., Fu, C., and Huang, K., Probabilistic Assessment of Liquefiable Soil Thickness Considering Spatial Variability and Model and Parameter Uncertainties, *Géotechnique*, 67(3), 228-241, Mar, 2017.