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.

A Fast Bayesian Approach for CPT-Based Soil Stratification with BUS

Shuo Zheng[1], Zi-Jun Cao[1], and Dian-Qing Li[1]

[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.
E-mail: zhengshuo@whu.edu.cn; zijuncao@whu.edu.cn (corresponding author); dianqing@whu.edu.cn

Abstract: Bayesian framework has been developed for probabilistic soil stratification based on the profile of soil behavior type index I_c calculated from cone penetration test (CPT) data. The unique feature of Bayesian soil stratification approach (BSSA), in comparison of with existing methods (e.g., clustering), is that it is able to provide not only the "best" estimate of soil stratigraphy but also its associated identification uncertainty that reflects the degrees-of-belief in identification results. To explicitly incorporate the spatial variability of I_c into soil stratification, random field theory is used to model the I_c profile. By this means, the knowledge on random field parameters (including mean value, standard deviation and scale of fluctuation) is needed for formulating the likelihood function, and they can be treated as nuisance parameters and are dealt with through marginalization. The marginalization involves a multidimensional integral to calculate the likelihood function, resulting in significant computational burden of BSSA and hampering its application in practice. This study attempts to tackle this computational difficulty using a fast computing strategy, which combines a simplified formula of the likelihood function with continuous nature of I_c data in each soil layer, with Bayesian Updating with Structural Reliability Methods (BUS) for identifying soil stratification model parameters (including the number and thicknesses of soil layers). The proposed fast computing strategy is illustrated and verified using real I_c data. Results of the soil stratification and the computational costs are compared with those obtained using brute-force direct numerical integration for marginalization.

Keywords: Soil stratigraphy; cone penetration test; soil behavior type index; Bayesian framework.

1 Introduction

Cone penetration test (CPT) is widely used to determine the soil stratigraphy during geotechnical site investigation (Roberson 2009). In general, CPT-based soil stratification 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; Wang et al. 2019), 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), which 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 $\underline{H}_N$. To quantify the uncertainty in N and $\underline{H}_N$, the Bayesian soil stratification approach (BSSA) based on the profile of I_c has been developed by Cao et al. (2018). 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.

Bayesian analysis is often criticized for its computational complexity and costs. With the BSSA, the knowledge on random field parameters (including mean value, standard deviation and scale of fluctuation) is needed for formulating the likelihood function, and they can be treated as nuisance parameters and are dealt with through marginalization. The marginalization involves a multidimensional integral to calculate the likelihood function, resulting in significant computational burden of BSSA and hampering its application in practice. This paper proposes a fast computing strategy to tackle this computational difficulty, taking advantage of the limited ranges of random field parameters of I_c and sequential nature of CPT data for identifying soil stratification model parameters (e.g., N and $\underline{H}_N$) using Bayesian Updating with Structural Reliability Methods (BUS). Moreover, the marginalization which involves a multidimensional integral to calculate the likelihood function was simplified as an algebra formula to avoid matrix manipulations under the assumption of a single exponential correlation function and equally spaced data. The proposed computing strategy are illustrated and verified using real I_c data from the clay site of the NGES at Texas A&M University. Results of the soil stratification and the computational

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_IS12-12-cd

costs are compared with those from the BSSA with repetitive brute-force numerical integration for marginalization.

2 Bayesian Soil Stratification Approach

For a given profile of I_c (i.e., $\underline{\xi}$), identification of soil stratigraphy under the proposed BSSA 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 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 $P(N)$ is the prior probability of N reflecting the prior knowledge on N in the absence of CPT data; $p(\underline{\xi})$ is a normalizing constant and is independent of N; $p(\underline{\xi}|\ N)$ is the 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. 2018):

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

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 $\underline{H}_N = [H_1, H_2, …, H_N]$, they can be considered uniformly distributed within a range from 0 to CPT sounding depth H, i.e., $0 < H_n < H$ for $n = 1, 2, …, N$, and all the possible combinations of $\underline{H}_N$ are uniformly distributed within a N-1 dimensional simplex $\Omega = \{\Sigma_{n=1}^{N} H_n = H, 0 < H_n < H\}$. Such a uniform distribution of $\underline{H}_N$ can be represented by a flat Dirichlet distribution, and is expressed as (Cao et al. 2018):

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

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 $\underline{I}_{cn}(Z)$, $n = 1, 2, …, N$, 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., $\underline{\xi} = [\underline{\xi}_1, \underline{\xi}_2, …, \underline{\xi}_N]$) obtained from the N soil layers are considered as a realization of the N random fields with model parameters $\underline{\theta}_n = [\mu_n, \sigma_n, \lambda_n]$, $n = 1, 2, …, N$. Then, $p(\underline{\xi}|\ \underline{H}_N, N)$ is expressed as (Cao and Wang 2013; Wang et al. 2019):

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

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 limited ranges reported by Cao et al. (2018). The model parameters $\underline{\theta}_n = [\mu_n, \sigma_n, \lambda_n]$ are treated as nuisance parameters and dealt with through marginalization in the likelihood function of each soil layer.

In addition, the high-dimensional integral is involved in the evidence and the posterior distribution (see Eqs. (1) and (2), respectively), particularly as N is relatively large. 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, details of the algorithm and implementing procedures of the BUS with SuS are not provided herein. Interested readers can refer to Straub and Papaioannou (2015), DiazDelaO et al. (2017), and Cao et al. (2018) for more details. A large number of samples $\underline{H}_N$ drawn from the BUS with SuS are needed to calculate the integral in Eq. (5) for the likelihood function. Inevitably, the soil layers of different samples might contain the same data points $\underline{\xi}_n$. Computing Eq. (5) by brute-force numerical integration is computationally expensive, and the computational cost increases as the number of CPT data points increases. The $p(\underline{\xi}_n| \underline{\theta}_n, \underline{H}_N, N)$ is further simplified to achieve fast computation of Eq. (5) in the next subsection.

2.3 Fast computing strategy of likelihood function

For a given set of $\underline{\theta}_n$, the natural logarithm of joint Gaussian PDF $\ln p(\underline{\xi}_n| \underline{\theta}_n, \underline{H}_N, N)$ of $\underline{\xi}_n$ is given by:

$$\ln p\left(\underline{\xi}_n \mid \underline{\theta}_n, \underline{H}_N, N\right) = -\frac{k_n}{2}\ln(2\pi) - k_n \ln(\sigma_{\ln,n}) - \frac{1}{2}\ln\left|\det \underline{R}_n\right| - \frac{1}{2\sigma_{\ln,n}^2} \times \left(\underline{\xi}_n - \mu_{\ln,n}\underline{l}_n\right)\underline{R}_n^{-1}\left(\underline{\xi}_n - \mu_{\ln,n}\underline{l}_n\right)^T \tag{6}$$

where $\mu_{\ln,n}$ and $\sigma_{\ln,n}$ are the mean and standard deviation of $\ln I_c$ in the n-th soil layer; $\underline{R}_n$ is the correlation matrix of $\underline{\xi}_n$, and its (i, j)-th entry represents the correlation coefficient ρ_n of $\ln I_c$ values at respective depths Z_i and Z_j; and $\underline{l}_n$ is a column vector with k_n components that are all equal to one. The size of the correlation matrix $\underline{R}_n$ depends on the number of data points in the n-th soil layer. Computational costs for calculating $|\det\underline{R}_n|$ and $\underline{R}_n^{-1}$ increase with the increase of the number of data points contained in $\underline{\xi}_n$.

The correlation coefficient ρ_n can be calculated from the single exponential correlation function:

$$\rho_n[\ln I_c(Z_i), \ln I_c(Z_j)] = \exp\left(-2\left|d_{i,j}\right| / \lambda_n\right) \tag{7}$$

where $d_{i,j} = |Z_i - Z_j|$ $(i, j=1, 2, ..., k_n)$ is a distance between depths Z_i and Z_j within the n-th layer; λ_n is a scale of fluctuation of $\ln I_c$ in the n-th soil layer. When data are equispaced, the correlation matrix has a simple closed form determinant and a tridiagonal inverse (Fenton 1999):

$$\left|\det \underline{R}_n\right| = (1-q^2)^{k_n-1} \tag{8}$$

$$\underline{R}_n^{-1} = \left(\frac{1}{1-q^2}\right)\begin{bmatrix} 1 & -q & 0 & 0 & \cdots & 0 & 0 \\ -q & 1+q^2 & -q & 0 & \cdots & 0 & 0 \\ 0 & -q & 1+q^2 & -q & \cdots & 0 & 0 \\ 0 & 0 & -q & 1+q^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1+q^2 & -q \\ 0 & 0 & 0 & 0 & \cdots & -q & 1 \end{bmatrix} \tag{9}$$

where $q = \exp(-2\Delta Z/\lambda_n)$ for I_c data spaced at ΔZ apart for $k_n > 1$. When there is only one data point in the soil layer, e.g., $k_n = 1$, q equals to zero. The inverse of the correlation matrix $\underline{R}_n$ given by Eq. (9) is a sparse symmetric matrix. Therefore, Eq. (6) can be simplified by combining Eqs. (8) and (9) as:

$$\ln p\left(\underline{\xi}_n \mid \underline{\theta}_n, \underline{H}_N, N\right) = -\frac{k_n}{2}\ln(2\pi) - k_n \ln(\sigma_{\ln,n}) - \frac{(k_n-1)}{2}\ln(1-q^2) - \frac{1}{2\sigma_{\ln,n}^2(1-q^2)} \times (A - 2B + C) \tag{10}$$

in which $$A = \sum_{i=1}^{k_n} \xi_i^2 + \sum_{i=2}^{k_n-1} \xi_i^2 q^2 - 2\sum_{i=1}^{k_n-1} \xi_i \xi_{i+1} q \quad (11)$$

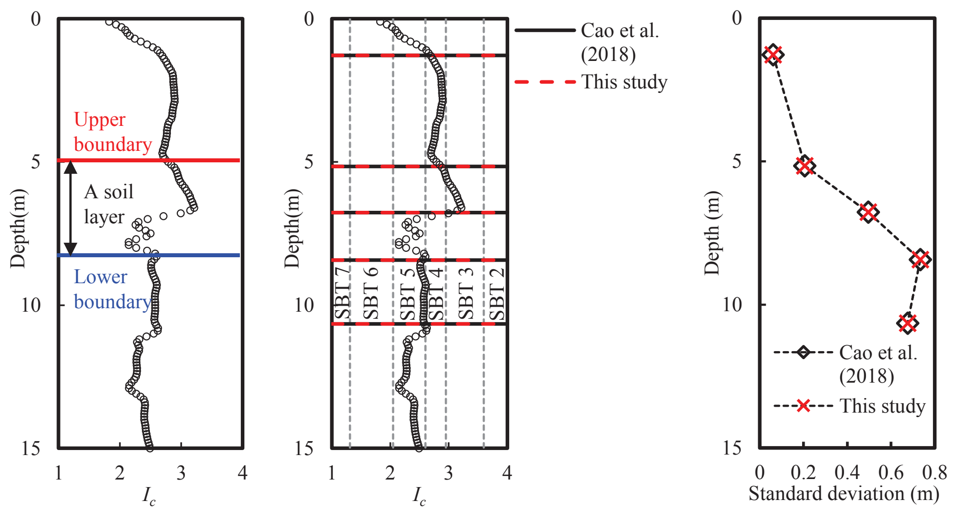
$$B = \mu_{\ln,n}(1-q)\times\left(\sum_{i=1}^{k_n}\xi_i - q\sum_{i=2}^{k_n-1}\xi_i\right) \quad (12)$$

$$C = \mu_{\ln,n}^2 \times \left[k_n + (2-2k_n)q + (k_n-2)q^2\right] \quad (13)$$

Simplification in Eq. (10) and Eqs. (11)–(13) allow calculating $\ln p(\underline{\xi}_n | \underline{\theta}_n, \underline{H}_N, N)$ without needs of matrix manipulations for the soil layer with $k_n>1$, the computational cost of which is hence significantly reduced and becomes irrelevant to the number of the I_c data points in each soil layer. Making use of the simplified form of Eq. (6) and the limited ranges of $\underline{\theta}_n = [\mu_n, \sigma_n, \lambda_n]$ for I_c reported by Cao et al. (2018) allow fast numerical integration in the likelihood function for each soil layer.

Moreover, the likelihood function $p(\underline{\xi} | \underline{H}_N, N)$ of each sample $\underline{H}_N$ generated from BUS with SuS needs to be calculated. Careful examination of the likelihood function does provide more useful insights to facilitate the calculation in this problem. As shown in Figure 1(a), CPT provides nearly continuous and sequential data points, and each soil layer has to contain a part of the CPT sounding between its upper and lower boundaries, which leads to a limited number of combinations of the locations of upper and lower boundaries of a soil layer.

These limited combinations of upper and lower boundaries of a soil layer can be determined beforehand for a given set of CPT data, allowing that the number of data points in a soil layer ranges from one to all the data points in the I_c profile. Then, the corresponding likelihood function of each possible soil layer defined by the combination of upper and lower boundaries is calculated using Eq. (10), which is used to determine the likelihood function of each sample of soil layer thicknesses generated by BUS with SuS. This avoids repeatedly calculating likelihood functions of soil layers with the same I_c data points for different random samples, and leads to significant computational saving for populating the posterior distribution of soil layer thicknesses given by Eq. (2) using BUS with SuS. The proposed approach is illustrated and validated using real-life CPT data in the next section.



(a) I_c profile at the NGES clay site

(b) Boundaries for the most probable model N*=6

(c) Standard deviations of layer boundary depths

Figure 1. Comparison of soil stratification results with those in Cao et al. (2018).

3 Illustrative Example

For illustration, the proposed computing strategy for BSSA is applied to identifying the soil stratigraphy at the clay site of the NGES at Texas A&M University based on a set of CPT data obtained from the site, which is shown in Figure 1(a). This set of CPT data was used to illustrate the BSSA framework by Cao et al. (2018), where the numerical integration needed in Eq. (5) was carried out in a brute-force manner for each sample generated by BUS with SuS. Consider, for example, that the number of soil layers N varies from 1 to 10. In addition, the conditional probability p_0 and the number, N_{cs}, of samples per SuS simulation level are set as 0.1

and 100,000, respectively, to perform SuS for BUS. The logarithms (i.e., $\ln p(\xi| N)$) of the evidence and the most probable depths of soil layer boundaries for each possible value of N obtained from the proposed approach at the NGES site are summarized in Table 1. The most probable number of soil layers at the clay site is six, i.e., $N^*= 6$, since the maximum of the evidence occurs at $N = 6$, i.e., $\ln p(\xi| N=6)=352.93$.

Figure 1(b) shows the boundaries between different SBTs based on the I_c by vertical gray dashed lines. It also shows the probable soil layer boundaries identified using the proposed algorithm for the $N^*= 6$ by red dash lines, which are identical with those (see black solid lines) reported by Cao et al. (2018). More importantly, using the posterior samples of depths of the five internal layer boundaries, their standard deviations are obtained through conventional statistical analyses. The standard deviation of boundary depths quantitatively reflects the uncertainty in the location of soil layer boundaries. As shown in Figure 1(c), the standard deviations of boundary depths obtained from this study are in good agreement with those reported by Cao et al. (2018). In summary, Figures 1(b) and 1(c) show that soil stratification results obtained from the proposed computation strategy are identical with those reported by Cao et al. (2018), where the brute-force numerical integration was implemented for calculating the likelihood function with matrix manipulations for each sample generated by SuS.

Table 1. Soil stratification results from the proposed approach at the NGES clay site.

N	$\ln p(\xi\|N)$	Most probable depths of soil layer lower boundaries D_n^*, n = 1, 2, …, N										Computation time[a]	
		D_1^*	D_2^*	D_3^*	D_4^*	D_5^*	D_6^*	D_7^*	D_8^*	D_9^*	D_{10}^*	Brute-force numerical integration	This study
.	304.99	15.00										0.04s	0.01s
2	317.91	8.15	15.00									2.7h	34.4s
3	329.43	1.26	6.82	15.00								4.5h	1.5min
4	344.79	1.23	6.77	8.12	15.00							6.8h	3min
5	350.35	1.24	6.78	8.41	10.64	15.00						8.0h	4.4min
6	**352.93**	**1.28**	**5.16**	**6.77**	**8.43**	**10.65**	**15.00**					10.6h	6min
7	352.70	1.20	5.18	6.78	8.48	10.68	13.31	15.00				15.1h	8min
8	346.66	1.33	5.41	6.75	8.47	10.95	11.19	13.21	15.00			19.5h	11.3min
9	342.49	0.38	1.21	5.36	6.77	8.36	10.92	11.11	13.28	15.00		17.9h	12.8min
10	337.11	0.42	1.29	5.56	6.73	6.83	8.89	10.92	11.11	13.27	15.00	18.9h	14.2min

Note a: On a desktop computer with 8 GB RAM and one Intel Core i5 CPU clocked at 2.7 GHz, and N_{cs} =100,000

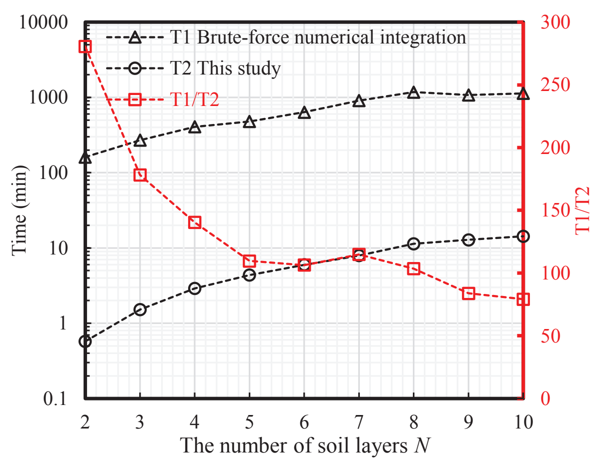


Figure 2. Computation time comparison.

To illustrate the efficiency of the proposed fast computing strategy, the last two columns of Table 1 compare the computation time of the proposed computing strategy with that for calculating the likelihood function through brute-force numerical integration (Cao et al. 2018). In this study, the calculation is performed on a desktop computer with 8 GB RAM and one Intel Core i5 CPU clocked at 2.7 GHz for each soil stratification model considered in the NGES example. The total computation time of 10 models for calculating the likelihood function by brute-force numerical integration is about 104 hours, while it only takes one hour in this study. Figure 2 further compares the computation time (denoted by T1 and T2, respectively) for marginalizing the likelihood function through brute-force numerical integration and the proposed computing strategy of likelihood function, which are shown by dash lines with triangles and circles, respectively. For a given set of CPT data, the computation time increases with the increase of N. Figure 2 also shows the ratio of T1 over T2. The proposed approach takes the time that is one to two orders of magnitudes less than that required for calculating the likelihood function by brute-force numerical integration, providing significant computational saving.

4 Conclusions

This paper developed a fast computing strategy for Bayesian soil stratification approach based on I_c. The proposed computing strategy consists of two parts: (i) calculating the likelihood function, which involves a multidimensional integral for marginalization (see Eq.(5)) of random field parameters, was simplified as an algebra formula (i.e., Eq. (10)) to avoid matrix manipulations under the assumption of a single exponential correlation function and equispaced data; (ii) making uses of the, nearly continuous and sequential, nature of CPT data points in each soil layer to avoid repeated calculations of the likelihood function of the same soil layer for different random samples generated for solving the posterior distribution of soil layer thicknesses. The proposed computing strategy for BSSA was illustrated using real CPT data. Results showed that using the proposed computing strategy for BSSA leads to significant computational saving in comparison with calculating the likelihood function through brute-force numerical integration for each sample, meanwhile providing the same soil stratification results for a given set of CPT data.

Acknowledgments

This work was supported by the National Key R&D Program of China (Project No. 2017YFC1501300), the National Natural Science Foundation of China (Project Nos. 51679174, 51779189, and 51579190), and Young Elite Scientists Sponsorship Program by CAST (Project No. 2017QNRC001). The financial support is gratefully acknowledged.

References

Cao, Z.J. and Wang, Y. (2013). Bayesian approach for probabilistic site characterization using cone penetration tests. *Journal of Geotechnical and Geoenvironmental Engineering*, 139(2), 267-276.

Cao, Z.J., Zheng, S., Li, D.Q., and Phoon, K.K. (2018). Bayesian identification of soil stratigraphy based on soil behavior type index. *Canadian Geotechnical Journal*, https://doi.org/10.1139/cgj-2017-0714.

Ching, J., Wang, J.S., Juang, C.H., and Ku, C.S. (2015). Cone penetration test (CPT)-based stratigraphic profiling using the wavelet transform modulus maxima method. *Canadian Geotechnical Journal*, 52(12), 1993-2007.

Diazdelao, F.A., Garbuno-Inigo, A., Au, S.K., and Yoshida, I. (2017). Bayesian updating and model class selection with subset simulation. *Computer Methods in Applied Mechanics and Engineering*, 317 (Supplement C), 1102-1121.

Fenton, G.A. (1999). Estimation for stochastic soil models. *Journal of Geotechnical and Geoenvironmental Engineering*, 125, 470-485.

Hegazy, Y.A. and Mayne, P.W. (2002). Objective site characterization using clustering of piezocone data. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(12), 986-996.

Phoon, K.K., Quek, S.T., and An, P. (2003). Identification of statistically homogeneous soil layers using modified Bartlett statistics. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(7), 649-659.

Phoon, K.K., Quek, S.T., and An, P. (2004). Geostatistical analysis of cone penetration test (CPT) sounding using. *Canadian Geotechnical Journal*, 41(41), 356-365.

Robertson, P.K. (2009). Interpretation of cone penetration tests - a unified approach. *Canadian Geotechnical Journal*, 46(46), 1337-1355.

Robertson, P.K. and Wride, C.E. (1998). Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian Geotechnical Journal*, 37(3), 270-271.

Straub, D. and Papaioannou, I. (2015). Bayesian updating with structural reliability methods. *Journal of Engineering Mechanics*, 141(3), 04014134.

Wang, Y., Huang, K., and Cao, Z.J. (2013). Probabilistic identification of underground soil stratification using cone penetration tests. *Canadian Geotechnical Journal*, 50(7), 766-776.

Wang, X.R., Wang, H., Liang, R.Y., and Liu, Y. (2019). A semi-supervised clustering-based approach for stratification identification using borehole and cone penetration test data. *Engineering Geology*, 248, 102-116.