Bayesian Approach for Soil Stratification Using Reversible Jump Markov Chain Monte Carlo

Cong Miao[1], Luyu Ju[2] and Shuo Zheng[3]

[1]*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, No.8 Donghu South Road, Wuhan, PR China, Email: miaocong@whu.edu.cn.*
[2]*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, No.8 Donghu South Road, Wuhan, PR China, Email: ju_luyu@whu.edu.cn*
.[3]*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, No.8 Donghu South Road, Wuhan, PR China, Email: zhengshuo@whu.edu.cn.*

Abstract: Bayesian method has been widely used for underground soil stratification based on the profile of soil behavior type index I_c calculated from raw CPT data. However, in the case of high-dimension, the considerably computational effort is the mainly burden and hamper for its application in practice. This study develops a novel Bayesian framework that not only contains the statistical characteristics of I_c quantified by random field model, but also implements the corresponding engineering judgments in the context of Roberson SBT chart. An effective Bayesian updating technique, namely Reversible Jump Markov Chain Monte Carlo (RJMCMC), is applied to solving the Bayesian equations. The proposed method is illustrated and verified using four real-world datasets. Cases studies show that using RJMCMC for Bayesian framework is capable of contributing to remarkably computational saving, meanwhile identifying the most probable soil stratigraphy (including the number and boundaries of soil layers). Finally, the impact of engineering judgments for soil stratification is discussed.

Keywords: Bayesian Framework, Soil Stratification, Soil Behaviour Type Index, Reversible Jump, Engineering Judgments

1. Introduction

In the past few decades, cone penetration test (CPT) has been gaining increasing use worldwide because it is a rapid, repeatable, economical technique that provides nearly continuous measurements over the depth (Roberson 2009). Despite the fact that no soil samples are recovered from CPT for visual inspection and laboratory testing, the measurements of mechanical responses of soil during penetration are obtained, such as cone tip resistance q_c, sleeve friction f_s, and pore water pressure u, etc. These measurements could be used for identification of soil stratigraphy. In general, it consists of two major steps: (a) determine the soil type at each testing depth (i.e., soil classification) based on CPT measurements, yielding a profile of the soil type along depth; and (b) identify the number N and boundaries (or thicknesses) $\underline{D}_N = [D_0, D_1, D_2, \ldots, D_{N-1}, D_N]$ of soil layers based on the profile of the soil type. Among various CPT-based soil classification systems, the use of "Soil Behaviour Type" (SBT) index, denoted by I_c, is extensive (Roberson and Wride 1998, Roberson 2009). However, there is a profound challenge in identifying soil stratigraphy (i.e., determining N and D_N) from a single profile of I_c due to the discrete nature of CPT limited data, the inherent spatial variability of I_c along depth and the boundary effect between adjacent soil layers.

Several approaches have been developed to delineate soil stratigraphy using CPT data in an objective and quantitative way, such as wavelet transform modulus maxima method (Ching et al. 2015), semi-supervised clustering-based approach (Wang et al. 2019), and Bayesian methods (Cao and Wang 2013; Wang et al. 2013; Cao et al. 2019). Among them, the Bayesian methods are able to provide both the "best" estimates of N and $\underline{D}_N$ with some prescribed criterion for soil stratification and the information on the uncertainty in their estimates. Nevertheless, one common challenge of these methods is computationally complexity and costs. This is because that the Bayesian method would require solving high-dimensional integrals and nonconvex optimization problem (Ching et al. 2015). On the other hand, engineering judgments are supposed to be considered as a part of prior knowledge in Bayesian inference to stratify soil layers based on the I_c profile. However, it is a subjective and unquantifiable manner due to engineers of different experience, expertise, and judgments (Cao et al. 2019).

This work develops a novel Bayesian framework for underground soil stratification based on the profile of I_c. Under the proposed Bayesian framework, the statistical characteristics of I_c are taken into account through random field model, and the engineering judgments are also explicitly incorporated in the Bayesian framework. Then, an effective Bayesian updating technique called Reversible Jump Markov Chain Monte Carlo (RJMCMC) is utilized for Bayesian analysis to conduct, automatically and fast, underground soil stratification (Green 1995, Dettmer et al. 2012). RJMCMC is an extension of the popular Metropolis–Hastings algorithm, designed to allow movement trans-dimensional inversion. This study starts with Bayesian framework for soil stratification based on N random field model, followed by illustrating RJMCMC algorithm and its key movements for computational difficulty in solving Bayesian equations. In the end, the proposed approach is demonstrated and verified through four sets of real-world data given by TC304/309 Student Contest.

2. Bayesian framework for soil stratification

For given profile of I_c and engineering judgments (i.e., Robertson SBT chart), soil stratigraphy (i.e., a combination of $\underline{D}_N$ and N) is quantified by their joint probability density function (PDF) $p(\underline{D}_N, N \mid \underline{\xi}, T)$ within the Bayesian framework, where $\underline{D}_N = [D_0, D_1, D_2, \ldots, D_{N-1}, D_N]$ is the soil boundaries; $\underline{\xi} = [\underline{\xi}_1, \underline{\xi}_2, \ldots, \underline{\xi}_{N-1}, \underline{\xi}_N]$ is a set of ln($\underline{I}_c$) data obtained within a depth-range of interest and $\underline{\xi}_n$ (n = 1, 2, …, N-1, N) represent the ln($\underline{I}_c$) values in the *n-th* soil layer; T is an indicator function of engineering judgments based on soil behaviour type. To explicitly incorporate the spatial variability of I_c into CPT based-soil stratification, Cao et al. (2019) adopted N independent random fields to model the I_c profile as shown in Fig. 1. Herein, the statistical characteristics (including mean value μ_n, standard deviation σ_n and scale of fluctuation λ_n) of I_c in the n-th random field are represented by $\underline{\theta}_n$ (i.e., $\underline{\theta}_n = [\mu_n, \sigma_n, \lambda_n]$). Nevertheless, it shall be pointed out that, for the purpose of soil stratification that focuses on identifying the number and boundaries of soil layers, only N and $\underline{D}_N$ are of intrinsic interest. The $\underline{\theta}_n$ are treated as nuisance parameters and dealt with through marginalization in this study (Wang et al. 2013, Cao et al. 2019).

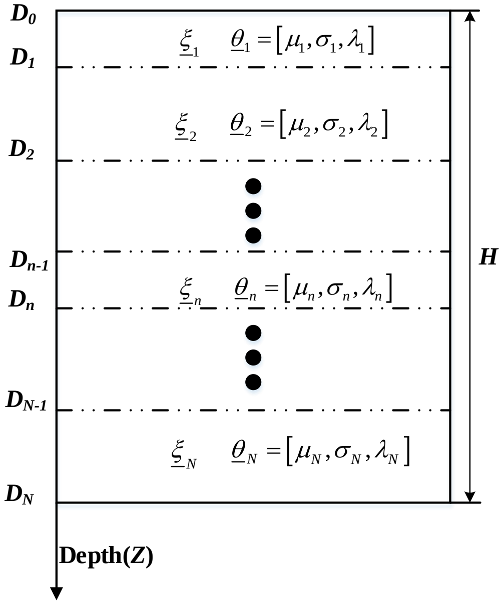

Figure 1. Illustration of Gaussian random field model

Alternatively, using the Bayes' Theorem, $p(\underline{D}_N, N \mid \underline{\xi}, T)$ is written as (Cao and Wang 2013; Cao et al. 2019):

$$p\left(\underline{D}_N, N \mid \underline{\xi}, T\right) = p\left(\underline{\xi}, T \mid \underline{D}_N, N\right) \times P\left(N\right) p\left(\underline{D}_N \mid N\right) / \, p\left(\underline{\xi}, T\right) \tag{1}$$

where $p(\underline{D}_N, N \mid \underline{\xi}, T)$ is the joint posterior distribution of $\underline{D}_N$ and N given T and $\underline{\xi}$, which is a dynamic dimension varying with N soil layers; $p(\underline{\xi}, T \mid \underline{D}_N, N)$ is the likelihood function that reflects the relationship between model parameters and observed data; $P(N)$ and $p(\underline{D}_N \mid N)$ are prior knowledge of soil layer quantity and boundaries in the absence of site information, respectively; $p(\underline{\xi}, T)$ is a normalizing constant and is united in different subspaces. The calculation of prior knowledge and likelihood function are introduced in the following two subsections, respectively.

2.1 Prior information

In the case of no prior knowledge on N and $\underline{D}_N$, they can be considered bounded, uniform distributions with the range of values chosen to represent physically reasonable and set wide enough that data predominantly determine posterior. In particular, the prior distribution of N is given by

$$P(N) = \frac{1}{N_{\max}} \qquad N = 1, 2, \cdots, N_{\max} \tag{2}$$

where $N_{\max}$ is the upper bound of the number of soil layers, and $P(N)$ is a constant in all of the state spaces. The prior distribution of $\underline{D}_N$ can be represented by a flat Dirichlet distribution, and is expressed as (Cao et al. 2019)

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

where $\Gamma(\cdot)$ is the Gamma function evaluated at N, and H is the maximum depth of cone penetration test. As indicated by Eq. (3), $p(\underline{D}_N \mid N)$ is a constant for a given N value and testing depth H.

2.2 Likelihood function

Using the Theorem of Conditional Probability (Ang and Tang, 2007), $p(\underline{\xi}, T \mid \underline{D}_N, N)$ is re-written as:

$$p(\underline{\xi}, T \mid \underline{D}_N, N) = p(T \mid \underline{D}_N, N, \underline{\xi})\, p(\underline{\xi} \mid \underline{D}_N, N) \tag{4}$$

where $p(T \mid \underline{D}_N, N, \underline{\xi})$ is the indicator likelihood function of engineering judgments based on soil behavior type and $p(\underline{\xi} \mid \underline{D}_N, N)$ is the likelihood function considering statistical characteristic of I_c.

The engineering judgments based on soil behaviour type approximately divide soils into six zones from 2 to 7 (i.e., Roberson SBT chart). Hence, the SBT of soil layer in the corresponding depth can be obtained by I_c profile on the basis of six soil zones. Then the probabilities of the soil layer belonging to each SBT, which are named SBT= X (i.e., X varies from 2 to 7) probability in this study, can be calculated by counting I_c data points in different SBT zones. Consider, for example, if a quarter of the I_c data points in a soil layer falls into the zone where SBT = 5, and the probability of the SBT = 5 of the soil layer is 25%. There is no asserting that a soil layer belongs to a SBT zone with perfect certainty, which causes the presence of noisy thin layers, and these thin layers may not be reasonable. Then, the existence probability of one boundary can be defined by the SBT probabilities of two adjacent soil layers on both sides of it. Without loss of generality, the threshold probability 0.5 is chosen to determine whether a boundary exists. Given a profile of I_c and a combination of N and $\underline{D}_N$, the existence probability of each boundary can be calculated and compared with the threshold 0.5. Herein, as long as any one of existence probabilities of boundaries is less than 0.5, and the combination of soil layer boundaries is unreasonable. Using the criterion, the indicator likelihood function based on Roberson SBT chart can be calculated as:

$$p(T \mid \underline{D}_N, N, \underline{\xi}) = \begin{cases} 1 & P(SBT_{n-1} \neq SBT_n) > 0.5 \text{ for } \forall\, n \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where SBT_n (n = 1, 2, …, N-1, N) represent the soil behaviour type values in the n-th soil layer. Further, through the aforementioned N random fields model, $p(\underline{\xi} \mid \underline{D}_N, N)$ is written as (Cao et al. 2019):

$$p(\underline{\xi} \mid \underline{D}_N, N) = \prod_{n=1}^{N} p(\underline{\xi}_n \mid \underline{D}_N, N) \tag{6}$$

where $p(\underline{\xi}_n \mid \underline{D}_N, N)$ is the likelihood function with statistical characteristic of I_c in the n-th soil layer. Using the Theorem of Total Probability, $p(\underline{\xi}_n \mid \underline{D}_N, N)$ is expressed as

$$p(\underline{\xi}_n \mid D_n, N) = \int_{\Omega} p(\underline{\xi}_n \mid \underline{\theta}_n, D_n, N)\, p(\underline{\theta}_n \mid D_n, N)\, d\underline{\theta}_n \tag{7}$$

where $p(\underline{\xi}_n \mid \theta_n, \underline{D}_N, N)$ is the joint Gaussian PDF of $\underline{\xi}_n$, and $p(\theta_n \mid \underline{D}_N, N)$ is the prior information of $\underline{\theta}_n$. For the sake of conciseness, detailed calculation of the joint Gaussian PDF is not provided herein. Interested readers are referred to Cao et al. (2019) for details.

Substituting Eqs. (2) -(7) into Eq. (1) gives the $p(\underline{D}_N, N \mid \underline{\xi}, T)$ for conducting the underground soil stratification. However, due to the discontinuity of the likelihood function with respect to $\underline{D}_N$, constraint relationship among soil boundaries (i.e., $D_n > D_{n-1}$), and high-dimensional integral involved in posterior, most of available inference techniques for Bayesian updating are computational difficulty. The next section introduces a high-efficiency Bayesian updating technique named RJMCMC (Green 1995) to, simultaneously and efficiently, obtain the most probable number of soil layers N^* and their most probable boundaries $\underline{D}^*_{N^*}$ according to $p(\underline{D}_N, N | \xi, T)$.

3. RJMCMC for Bayesian Updating

RJMCMC (Green 1995) generalizes the Metropolis-Hastings (M-H) algorithm to cases where the proposal distribution not only moves within the current state (or model) space, but also between state spaces, of different dimension or type. In the process of undergoing dimension changes (i.e., jump), the requirement of detailed balance and dimension matching should not be violated. The number of soil layer N in Eq. (1) can be treated as a variable dimension parameter, resulting in the joint posterior distribution $p(\underline{D}_N, N | \xi, T)$ that spans multiple subspaces of different dimensions. The scheme of three move types: *Update*, *Birth* and *Death*, are applied here, and the latter two implements dimensional change in the RJMCMC algorithm. A key advantage of this approach is that the nature of RJMCMC renders a well-suited approach to choose an optimal candidate model rather than comparing model evidence of all the models, so it is a fast computing strategy to obtain the most probable number N^* of soil layers and their most probable boundaries $\underline{D}^*_{N^*}$ (Green 1995).

At each run RJMCMC, an independent random choice is made among three move types (i.e., *Birth* move, *Death* move and *Update* move) that have probabilities b_i, d_i, and u_i, respectively, depending only on the current number of soil layer i, and satisfying $b_i + d_i + u_i = 1$ ($i = 2, \ldots, N_{\max}-1$). Generally speaking, $d_1 = 0$ and $b_{N\max} = 0$ impose the preassigned bound limit on the number of soil layer. Apart from these constraints, these probabilities are chosen as:

$$b_i = d_i = u_i = \frac{1}{3} \tag{8}$$

On the other hand, to further reduce computing costs and improve the acceptance in the Markov Chain, perturbing one boundary in a particular state space only changes the adjacent two layers and maintain the rest of the boundary invariant (Dettmer et al. 2012). Therefore, there is no need to compute the likelihood function of entire soil layers for each sample. As has been noted above, one sweep of RJMCMC algorithm for $(N, \underline{D}_N)$ consists of three move types that have the equal probability except the special situations (i.e., $i = 1$ and $i = N_{max}$). The details of these types are introduced in the following subsections.

3.1 Update move

The *Update* move is consistent with ordinary Markov Chain Monte Carlo in the fixed dimension case. Herein, the M-H algorithm can be implemented by calculating the prevailing M-H acceptance ratio of *Update* move for a step from the current state $(N, \underline{D}_N)$ to the next state $(N', \underline{D}'_{N'})$

$$\alpha_{u_i} = \min\left\{1, \left(\textit{posterior ratio}\right) \times \left(\textit{proposal ratio}\right)\right\} \tag{9}$$

where, here and later, *posterior ratio* = $p(N', \underline{D}'_{N'} | \xi, T) / p(\underline{D}_N, N | \xi, T)$; *proposal ratio* = $Q(\underline{D}_N) / Q(\underline{D}'_{N'})$ and $Q(\cdot)$ is the proposal function. The *proposal ratio* is united due to the symmetry of the proposal function at current state and next state in the *Update* move of RJMCMC algorithm. Referring the zone of influence may be about 0.2-0.5 m for a CPT (Idriss et al. 2008), the proposal function in this study can be conservatively defined as:

$$Q\left(D'_{n'}\right) = \frac{1}{D_{n+1} - D_{n-1} - 1} \tag{10}$$

where $D'_{n'}$ is the n-th soil boundary generated from the uniform distribution that is limited by the interface at 0.5m of adjacent soil layer boundaries. In other words, the thickness of each soil layer should be more than 0.5m.

3.2 Birth move and Death move

In this section, the *Birth* and *Death* moves are illustrated for the trans-dimensional posterior in Eq. (1) while not violating the requirement of detailed balance and dimension matching. In the process of dimension changes, the proposal distribution in Eq. (10) is replaced with a two-step procedure (Green 1995). Firstly,

$$Q(u) = \frac{1}{H} \tag{11}$$

where u is a uniform distribution from surface (i.e., 0) to a maximum depth of interest (i.e., H). Then calculate the next state (N', $\underline{D}'_{N'}$), using u and the current state (N, $\underline{D}_N$). An one to one invertible transformation function, $f(\cdot)$, is chosen

$$\left(N', D'_{N'}\right) = f\left(\left(N, D_N\right), u\right) \tag{12}$$

Note that the only restriction on the transformation function is that it is a diffeomorphism (i.e., both $f(\cdot)$ and its inverse are differentiable) for calculation of Jacobian determinant. To preserve positivity and maintain simplicity in the acceptance ratio calculations, u and the position of soil boundaries of the current state are sorted by ascending for the next state, which is $[D'_{0'}, D'_{1'}, D'_{2'}, \ldots, D'_{N'}, D'_{N+1'}] = [D_0, D_1, D_2, \ldots, D_n, u, D_{n+1}, \ldots, D_{N-1}, D_N]$. Noted that the two vectors on both sides of above equation are of size m' and m, respectively, and $m' = m$ (i.e., dimension matching). Before calculating the acceptance ratio of *Birth* move, the constraint of 0.5m (see 3.1 section) switch should be conducted to judge the rationality of the new soil boundary. Green (1995) shows that a trans-dimensional move of RJMCMC algorithm can be implemented similar to the M-H algorithm by generalizing the M-H acceptance to apply to a step from the current state (N, $\underline{D}_N$) to the next state (N', $\underline{D}'_{N'}$) as

$$\alpha_{b_i} = \min\left\{1, \textit{posterior ratio} \times \frac{d_{i+1}}{b_i Q(u)}|J|\right\} \tag{13}$$

where b_i, d_{i+1} are the probability of choosing *Birth* move and *Death* move, respectively. Their subscript represents the dimension of state space (i.e., the number of soil layer); $|J|$ is the Jacobian determinant that accounts for a diffeomorphism in multiple subspace of different dimensionalities and is united by pre-establishing the diffeomorphism for computational efficiency. As the inverted procedure of *Birth* move, alternatively, *Death* move randomly picks and deletes a soil boundary, reducing the current high dimension state by 1. In deriving an expression for the acceptance probability of the *Death* move, it is helpful to re-write Eq. (13). Then, the acceptance probability for the corresponding *Death* move has the same form with the appropriate change of labelling of the variables, and the ratio terms inverted as

$$\alpha_{d_i} = \min\left\{1, \textit{posterior ratio} \times \frac{b_{i-1} Q(u)}{d_i |J|}\right\} \tag{14}$$

The samples of joint posterior of N and $\underline{D}_N$ are generated through random walk of *Update, Birth and Death* moves for quantifying underground soil stratification. Interested reader can obtain the detailed algorithm and implementing procedures of RJMCMC formulation from Green (1995).

4. Case studies

For illustration, the proposed approach is applied to conducting underground stratification based on four real-world I_c profiles. Table 1 summarizes the basic information of these I_c profiles including one training dataset and three testing datasets. The second row in the Table 1 is their maximum depth (i.e., H). In addition, the most probable number of soil layers N^* and corresponding computational cost are of in two last rows of below Table 1. To assure the posterior predominantly determined by site information, the maximum number of soil layer is prescribed as 50 (i.e., N_{max} = 50). The RJMCMC simulation is run for 200,000 updates using a MATLAB R2015b on a desktop computer with the Inter Core i7 CPU. All of the computational time of four datasets is around 200 seconds, which indicates that proposed computation is independent of the amount of data points and has high computational efficiency.

Consider, for example, Fig. 2 only shows the frequency for the number of soil layers of the training dataset, indicating significant probability for soil layers with 17 to 20, and the maximum probability 36.62% occurs at 18 soil layers. Hence, the most probable number of soil layers for the training datasets is eighteen (i.e., N^*=18). The maximum number of soil layers as shown in Fig. 2 is 24, which is much less than the prescribed the maximum number of soil layer (i.e., N_{max} = 50). It clearly demonstrates that the range of prior distribution is wide enough. The results of the most probable N^* for the rest of testing datasets occur at N = 5, 6, 14, respectively (see Table 1).

Table 1. Summary of datasets

	Training Data	Testing Data (a)	Testing Data (b)	Testing Data (c)
Interval (m)	0.05	0.05	0.05	0.05
Depth (m)	40.40	14.90	18.35	28.25
N^*	18	5	6	14
Time (s)[a]	227	193	216	228

Note [a]: On the MATLAB R2015b of a desktop computer with the Intel Core i7 CPU, and run 200,000 times

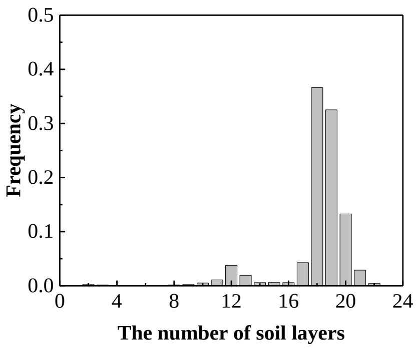

Figure 2. The frequency of soil stratification models at the training dataset (N_{max} = 50)

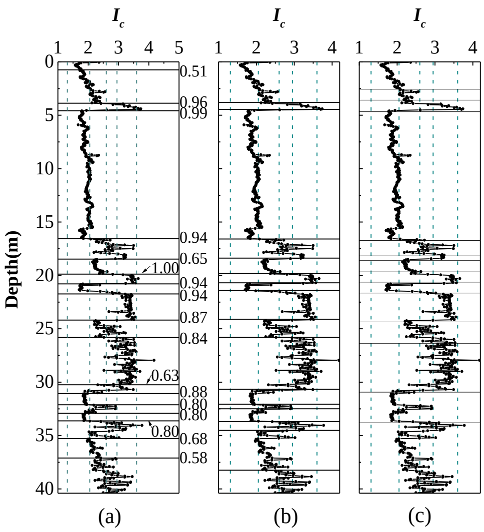

Figure 3. Soil identification results of different algorithms at the training dataset: (a) this study, (b) Ching et al. (2015) and (c) Wang et al. (2013)

The most probable depth of soil boundaries (i.e., $\underline{D}^*_{N*}$) for the training dataset using the proposed approach is shown in Fig. 3(a). The boundaries of different SBTs based on I_c are shown by vertical green dashed lines and each the most probable soil boundary identified by the proposed approach is shown by horizontal black solid line. The existence probability of each the most probable boundary defined in this study is also labelled beside the boundary in Fig. 3(a). Obviously, all the existence probabilities are greater than 0.5, which is affected by indicator likelihood function based on Roberson SBT chart. It is also noted that the existence probability of the first soil boundary is close to the prescribed threshold 0.5, reflecting lower degrees-of-belief based on Roberson SBT chart.

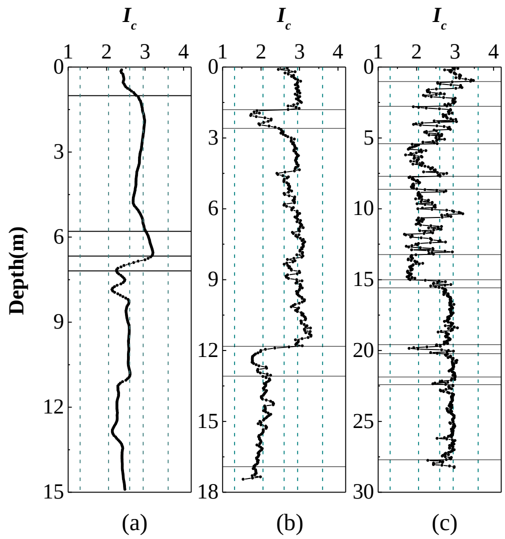

Figure 4. Soil identification results at the testing dataset (a), (b) & (c)

This set of CPT data also has been used to illustrate CPT based soil classification and/or stratification approaches in literature, including Bayesian soil classification and stratification (BSCS) approach based on Robertson SBT chart (Wang et al. 2013), and the wavelet transform modulus maxima (WTMM) method for soil stratification (Ching et al. 2015), which are shown in Fig. 3(b) and (c). In general, the results from the three methods (i.e., the proposed approach, WTMM method, BSCS method) are generally agreed with each other, except that the proposed method identifies some thin soil boundaries which were not identified by WTMM and BSCS. This is because that the proposed the unique Bayesian framework synthetically reflects the engineering judgments and statistical characteristics of I_c. In addition, the potential thin layers identified by statistical characteristics will be filtered out by the defined indicator likelihood function based on Roberson SBT chart, this will be demonstrated in the next section. For the three testing datasets, the proposed algorithm is carried out and the results of the most probable depth of soil boundaries are shown in Fig. 4. Similarly, the boundaries of I_c and each the most probable soil boundary are shown by vertical green dashed lines and horizontal black solid line, respectively. The proposed approach rationally identifies the location of soil layer boundaries based on the profile of I_c.

5. Discussion

This section discusses the impact of engineering judgments based on Roberson SBT chart. The soil stratification results with and without engineering judgments are shown in Fig. 5(a) and (b), respectively. The boundaries between different SBTs based on I_c are shown by vertical green dashed lines and the most probable soil boundaries identified by the proposed approach are shown by horizontal black solid lines. When only statistical features of I_c (i.e., $P(\underline{D}_N, N \mid \underline{\xi}) = p(\underline{\xi} \mid \underline{D}, N)p(\underline{D}_N|N)P(N)$) are available, it can be obviously noted that one potential issue of stratification results (see Fig 5(a)) is the possible misconception of the very thin soil layers. The cause of that it is remarkable sensitive to I_c variation. Especially at depths ranging from 5 to 15 m, numerous potential thin layers are conducted, which is dominated by SBT = 6 soil within this depth range. Therefore, a criterion based on Roberson SBT chart is proposed to reduce the thin layers that exist in the same SBTs zones. The criterion is remarkably effective by comparing results with engineering judgments and without it.

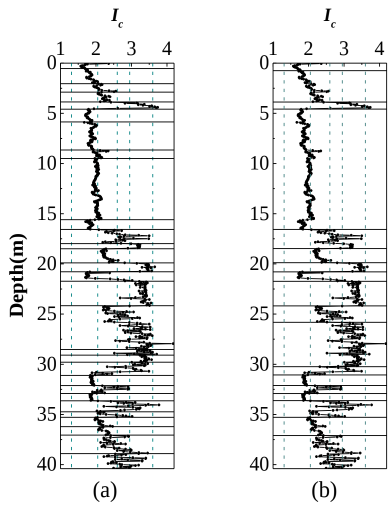

Figure 5. Soil identification results at the training dataset: (a) without engineering judgments and (b) with engineering judgments

6. Conclusion

In this study, a Bayesian framework is developed for soil stratification based on the profile of I_c calculated from cone penetration test measurements. The soil stratification information contained in the profile of I_c is explicitly extracted by proposed Bayesian framework. It is a trade-off between statistical differences in I_c values of adjacent soil layers and engineering judgments based on Roberson SBT chart. The samples of joint posterior distribution of N and $\underline{D}_N$ are obtained by running RJMCMC. In addition, for the purpose of further improve computational saving and ensure the acceptance in the Markov Chain, only the adjacent two layers are changed and maintain the rest of the boundaries invariant. Therefore, this study can simultaneously provide real-time analysis of boundary and quantity of soil layers based on Bayesian inference over a depth-range of interest.

Equations are derived for the proposed approach, and are demonstrated and verified by four real-world cases and contrasted with two other different methods for the training dataset. Results illustrate that RJMCMC for Bayesian framework rationally identify the most probable soil stratigraphy, and to address computational difficulty is greatly promising. The discussion of engineering judgments shows that it is necessary to take the engineering judgments account into the soil stratification.

References

Ang, A.H.-S., and Tang, W.H. 2007. *Probability concepts in engineering: emphasis on applications to civil and environmental engineering*. John Wiley and Sons, New York

Cao, Z.J., Zheng, S., Li, D.Q., and Phoon, K.K. 2019. Bayesian identification of soil stratigraphy based on soil behaviour type index. *Canadian Geotechnical journal*, in press.

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

Dettmer, J., Molnar, S., Steininger, G., Dosso, S.E. and Cassidy, J.F. 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models. *Geophysical Journal International*, 188(2): 719-734.

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

Green, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4): 711-732.

Idriss, I.M., Boulanger, R.W. 2008. *Soil liquefaction during earthquakes.* Monograph MNO-12, Earthquake Engineering Research Institute, Oakland, CA.

Richardson, S., Green, P.J. 1997. On Bayesian analysis of mixtures with an unkown number of components. *Journal of the Royal Statistical Society*, 59(4): 731-792.

Robertson, P.K. 2009. Interpretation of Cone Penetration Tests - A Unified Approach. *Canadian Geotechnical Journal*, 46(46): 1337-1355.

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.

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