# 2D high-resolution SBT image for subsurface soil zonation from limited CPTs using Bayesian compressive sampling

Yue Hu[1] and Yu Wang[2]
*Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong SAR*
yuehu47-c@my.cityu.edu.hk[1], yuwang@cityu.edu.hk[2]

**ABSTRACT:** A novel method is presented in this paper for subsurface soil classification and zonation in a two-dimensional (2D) vertical cross-section using a limited number of cone penetration tests (CPT). CPT is usually performed vertically, and the number of CPT soundings in a site is often limited in geotechnical engineering practice due to time and budget constraints. It is, therefore, difficult to properly interpolate CPT results along horizontal direction for depicting spatial variation of geo-material properties. The method presented in this study aims to address this difficulty, and it consists of three modules: 2D interpolation of CPT data using 2D Bayesian compressive sampling, determination of soil behavior type (SBT) using Robertson chart at every location in the 2D cross-section, and soil layer/zone delineation using an edge detection method. High-resolution images of CPT data and SBT information in the 2D vertical cross-section can be obtained. Soil layer/zone boundaries are delineated automatically. The proposed method is illustrated using a simulated example. It shows that in the simulated example the method performs well even when only five sets of CPT soundings are available.

**Keywords:** site characterization; subsurface soil cross-section; Bayesian estimation; compressive sensing

## *1. Introduction*

Delineation of subsurface soil stratification and zonation is an important task in site investigation for subsequent geotechnical design process. Cone penetration test (CPT) is a commonly used in-situ test for assessing engineering properties of subsurface soils [1-5]. It consists of pushing a cylindrical steel probe into the ground and provides nearly continuous measurements (e.g., 2-5cm measurement interval) of penetration tip resistance, sleeve friction, and pore water pressure to evaluate the property variations of subsurface soils along depth. When the probe penetrates through different soil layers with different mechanical properties, the apparatus records distinct CPT readings. Therefore, CPT can well reveal subsurface soil stratification and has been a popular tool in geotechnical site investigation [6-14].

Note that the previous works abovementioned are only limited to one-dimensional (1D) profile analysis, i.e., along the depth. However, subsurface soil layers/zones were formed during complex geological processes (e.g., weathering, erosion, transportation, and metamorphic processes) and both the boundaries of soil layers/zones and soil properties (e.g., CPT sounding measurements) within soil layers/zones may vary vertically and horizontally. To depict the varying nature of soil properties along the horizontal direction, two- or three-dimensional (2D or 3D) geotechnical analysis is often performed. In geotechnical engineering practice, a 2D vertical cross-section is commonly used to characterize the geology condition at a site, e.g., spatial variations of soil types along both vertical and horizontal directions [15-16].

Due to time and budget constraints, the number of investigation points (e.g., CPT soundings) at a site is often limited, especially in small to medium-sized projects. To interpret a 2D vertical cross-seciton, geotechnical engineers often delineate the stratification and zonation through linear interpolation between investigation points with supports of geological, geomorphological, as well as geophysical knowledges of the region. A common operation is to connect the boundaries that separate different soil layers in adjacent investigation points by straight lines [17], as shown in Fig. 1. Note that the actual shapes of soil layer boundaries are not necessarily linear. In addition, direct linear interpolation method cannot handle the situation when different numbers of soil layers are observed at different soundings, a scenario that is frequently encountered in site investigation. For example, in Fig. 1 four soil layers (i.e., clay, silt mixtures, sand mixtures and sand) are observed from CPT1, while only three soil layers (i.e., clay, sand mixtures and sand) are found from CPT2. It is very challenging to predict the soil types for the triangle zone filled with question marks because three sides of the triangles encounter three different soils. The issue in Fig. 1 underscores the need of a rational and objective method for identifying stratification and zonation in a 2D vertical cross-section from limited soundings.

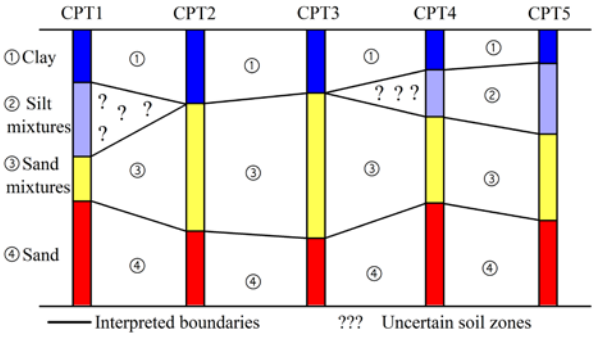



**Figure 1.** Challenge in subsurface soil zonation and stratification

This paper presents a CPT-based method that addresses the above challenge in non-parametric and data-driven manners. CPT data are directly interpolated in a 2D vertical cross-section without assuming a

parametric spatial autocorrelation function. The delineation is derived automatically and subjective human manipulation shown in Fig. 1 is not required. The method is illustrated using a simulated example. The effect of amount of CPT soundings is also investigated in the simulated example.

## 2. The proposed method

The proposed method consists of three modules: (1) 2D interpolation of normalized CPT soundings data (e.g., normalized friction ratio $F_R$ and normalized cone resistance $Q_t$) using 2D Bayesian compressive sampling (2D BCS); (2) soil behavior type (SBT) classification at each location using the interpolation results and a SBT chart; and (3) delineation of soil layer or zone boundaries using an edge detection method. The three modules are introduced in details as follows.

### 2.1. 2D interpolation of CPT data using Bayesian compressive sampling

Two-dimensional Bayesian compressive sampling (2D BCS) is an application of compressive sampling/sensing (CS) in a 2D space. CS is a novel sampling theory in signal processing, and it asserts that a spatially varying signal (e.g., a CPT profile along depth) can be recovered from sparse sampling points [18-20]. In the context of 2D BCS, a 2D signal **F** (e.g., CPT data variations in a 2D vertical cross-section), which is spatially varying along coordinates $x_1$ and $x_2$ (e.g., depth and horizontal directions), is represented by a matrix of $N_{x_1} \times N_{x_2}$. Mathematically, **F** can be expressed as weighted summation of a series of orthonormal 2D basis functions [21-23];

$$\mathbf{F} = \sum_{t=1}^{N_{x_1} \times N_{x_2}} \mathbf{B}_t^{2D} \omega_t^{2D} \tag{1}$$

where $\mathbf{B}_t^{2D}$ is fixed 2D basis function (e.g., discrete wavelet function, discrete cosine function); $\omega_t^{2D}$ is the corresponding weight coefficient. Note that for many natural signals (e.g., spatially autocorrelated geotechnical properties), most $\omega_t^{2D}$ have negligibly small values except for a limited number of non-trivial ones with significantly large magnitudes [22, 24]. Therefore, **F** can be approximated properly as $\hat{\mathbf{F}}$ if the non-trivial components of $\omega_t^{2D}$ are identified and estimated using sparse measurements data **Y** in the context of CS/BCS. **Y** is a sub-matrix of **F** with a dimension of $M_{x_1} \times M_{x_2}$ ($M_{x_1} \le N_{x_1}$, $M_{x_2} \le N_{x_2}$). The relation between **Y** and $\omega_t^{2D}$ is expressed as [21-23]:

$$\mathbf{Y} = \mathbf{\Psi}_{x_1} \mathbf{F} \mathbf{\Psi}_{x_2} = \sum_{t=1}^{N_{x_1} \times N_{x_2}} \mathbf{A}_t^{2D} \omega_t^{2D} \tag{2}$$

in which $\mathbf{\Psi}_{x_1}$ and $\mathbf{\Psi}_{x_2}$ are problem-specific measurement matrices with dimensions of $M_{x_1} \times N_{x_1}$ and $N_{x_2} \times M_{x_2}$, respectively, reflecting the positions of rows and columns of measured data **Y** in **F**. $\mathbf{\Psi}_{x_1}$ and $\mathbf{\Psi}_{x_2}$ can be adapted from an identity matrix. $\mathbf{A}_t^{2D}$ is a submatrix of 2D basis function $\mathbf{B}_t^{2D}$, i.e., $\mathbf{A}_t^{2D} = \mathbf{\Psi}_{x_1} \mathbf{B}_t^{2D} \mathbf{\Psi}_{x_2}$ and has the same dimension as **Y**.

Using the Eq. (2), the non-trivial coefficients can be estimated through maximum likelihood estimation [25] for $F_R$ and $Q_t$ respectively. To distinguish from the underlying true coefficients $\omega_t^{2D}$, $\hat{\omega}_t^{2D}$ is used to denote the estimated ones. The best estimate of the $\hat{\omega}_t^{2D}$ vector is derived under Bayesian framework and expressed as [17, 18]:

$$\boldsymbol{\mu}_{\hat{\omega}^{2D}} = \left(\mathbf{J} + \mathbf{D}\right)^{-1} \boldsymbol{V}_{tr} \tag{3}$$

where **J** is a matrix with element $J_{t,s} = tr[\mathbf{A}_t^{2D} (\mathbf{A}_s^{2D})^T]$, ($t$, $s$ = 1, 2, …, $N_{x_1} \times N_{x_2}$). "*tr*" represents trace operation in linear algebra. **D** is a diagonal matrix with diagonal elements $D_{t,t} = \alpha_t$ ($t$ = 1, 2, …, $N_{x_1} \times N_{x_2}$) in which $\alpha_t$ are non-negative parameters to be determined by maximum likelihood algorithm [22-25]. $\boldsymbol{V}_{tr}$ is a column vector with a length of $N_{x_1} \times N_{x_2}$:

$$\boldsymbol{V}_{tr} = \left\{ tr[\mathbf{Y}(\mathbf{A}_1^{2D})^T], tr[\mathbf{Y}(\mathbf{A}_2^{2D})^T], \ldots, tr[\mathbf{Y}(\mathbf{A}_{N_{x_1} \times N_{x_2}}^{2D})^T] \right\}^T \tag{4}$$

Using the best estimate of $\hat{\omega}_t^{2D}$ together with corresponding 2D basis function, the best estimate of 2D signal **F** can be obtained as $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$ (e.g., complete CPT data in a 2D cross-section), which is expressed as [21-23]:

$$\boldsymbol{\mu}_{\hat{\mathbf{F}}} = E(\hat{\mathbf{F}}) = \sum_{t=1}^{N_{x_1} \times N_{x_2}} \mathbf{B}_t^{2D} E(\hat{\omega}_t^{2D}) = \sum_{t=1}^{N_{x_1} \times N_{x_2}} \mathbf{B}_t^{2D} \mu_{\hat{\omega}_t^{2D}} \tag{5}$$

in which $\mu_{\hat{\omega}_t^{2D}}$ is the best estimate or expectation of $\hat{\omega}_t^{2D}$. Detailed derivation of the Bayesian formulation is referred to Zhao et al. [22].

In contrast to conventional geostatistics (e.g., ordinary Kriging), the presented method can preserve non-stationarity of CPT data (e.g., data from multiple types of soils with different statistics) without a de-trending process [26, 27]. The anisotropic pattern of CPT data in the 2D vertical cross-section is automatically preserved in a non-parametric manner [28].

### 2.2. Determination of soil behavior type (SBT) using SBT chart

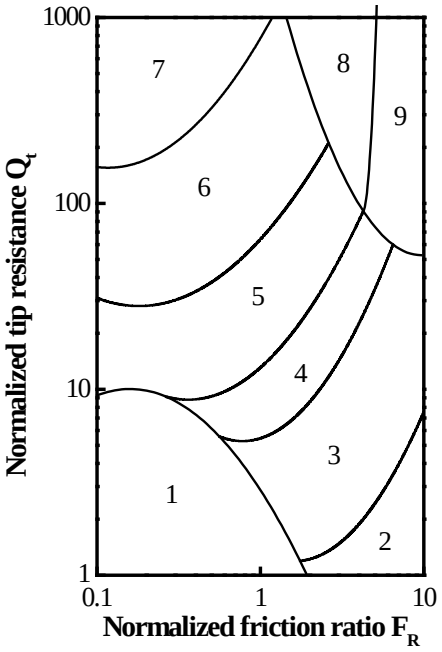
Once the 2D vertical cross-sections of normalized CPT data are interpolated, soil classification at each location can be performed using a SBT classification chart [6, 7]. An example of an SBT chart is shown in Fig. 2. The horizontal and vertical axis are $F_R$ and $Q_t$ respectively. The expressions of $F_R$ and $Q_t$ are given as below:

$$\begin{aligned} F_R &= \frac{f_s}{q_t - \sigma_{v0}} \times 100\% \\ Q_t &= \frac{q_t - \sigma_{v0}}{\sigma'_{v0}} \end{aligned} \tag{6}$$

where $\sigma_{v0}$ and $\sigma'_{v0}$ are vertical total stress and vertical effective stress respectively; $f_s$ is sleeve friction; $q_t$ is corrected cone resistance, expressed as:

$$q_t = q_c + (1-a)u \quad (7)$$

where $q_c$ is original cone resistance; $a$ is cone area ratio; $u$ is pore pressure for cosideration of pore water pressure condition. The whole chart is divided into nine areas, corresponding to nine SBTs ranging from sensitive, fine-grained to very stiff fine-grained soils, respectively. SBT at each location can be determined using a $F_R$ and $Q_t$ data pair from their respective interpolated 2D cross-sections at the same location. By checking which area the $F_R$ and $Q_t$ data pair is located on the SBT chart (see Fig. 2), the SBT at the corresponding location is determined. In a similar fashion, the SBT at every location in the 2D vertical cross-section is obtained, leading to a 2D SBT map or image for the vertical cross-section.




**Soil behavior type classification**

| Area | Description |
|---|---|
| 1 | Sensitive, fine-grained |
| 2 | Organic soils (peats) |
| 3 | Clays (clay to silty clay) |
| 4 | Silt mixtures (clayey silt to silty clay) |
| 5 | Sand mixtures (silty sand to sandy silt) |
| 6 | Sands (clean sand to silty sand) |
| 7 | Gravelly sand to sand |
| 8 | Very stiff sand to clayey sand |
| 9 | Very stiff, fine-grained |

**Figure 2.** Soil behavior type classification chart (after [6])

## *2.3. Delineation of soil layer or zone boundaries using an edge detection method*

Boundaries of different soil layers and zones are delineated to clearly define the stratification and zonation. In general, the locations where the SBT index changes abruptly in the SBT map may be interpreted as the boundaries of different soil units. For example, if two adjacent elements in the SBT map are classified as 5 (i.e., sand mixtures) and 6 (i.e., sands) respectively, the line between these two elements may be interpreted as the boundary between sand mixtures and sands zones (see, for example, black bold line in Fig. 3a). Determination of soil layers/zones boundaries is therefore equivalent to locating the abrupt changes of SBT values in an SBT map or image.

In image processing, locating abrupt change may be achieved through edge detection methods [29]. For example, edge detection techniques can mathematically find the locations where the first order derivative of the image intensity (e.g., SBT value in this case) reaches a maximum [30]. In this study, first order derivatives along both vertical and horizontal directions at each location are calculated by convolving the SBT map with an edge detection operator, such as a Canny filter [30]. The convolution operation results in a 2D SBT gradient map, where those locations with maximum gradients are interpreted as the boundary. A binary matrix with the same dimension as the SBT map can be obtained, with "Y" representing the locations of boundaries, as shown in Fig. 3b. Using the above edge detection method, soil layer or zone boundaries can be delineated automatically from the SBT map. Note that the edge detection process can be implemented readily using commercial software, such as the "edge" function in image processing toolbox of MATLAB.

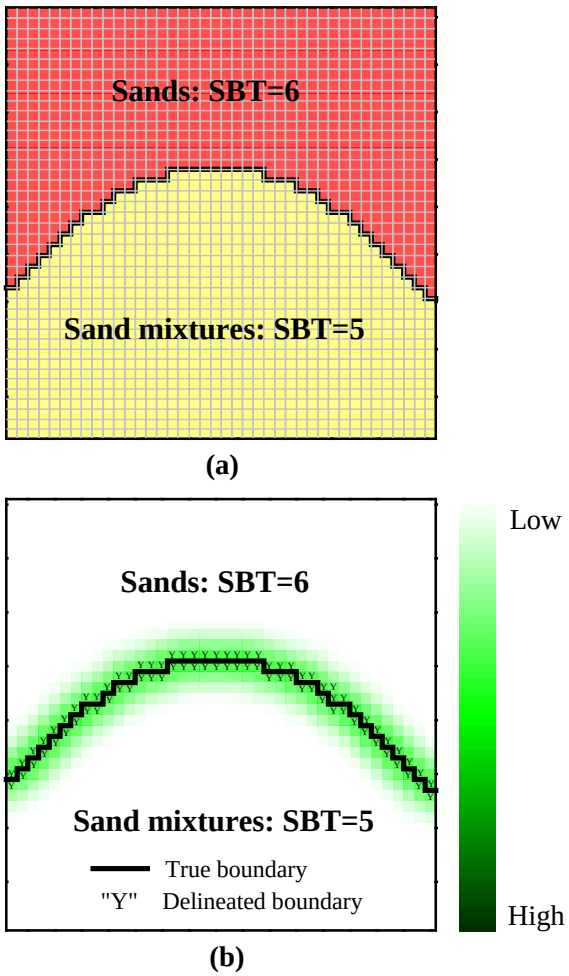



**Figure 3.** Example of soil layer delineation in soil behavior type (SBT) map: (a) SBT map with two soil layers and boundary (black bold line) between soil layers; (b) SBT gradient map and boundary delineation by edge detection

Fig. 4 briefly shows a flowchart summarizing implementation procedures of the proposed method. For

details, readers are referred to [23]. A simulated example is used to illustrate the method in the following section.

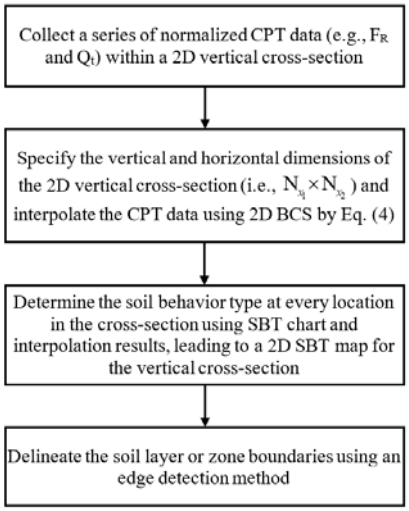



**Figure 4.** Implementation procedures of the method

## 3. Simulated example

A synthesized 2D vertical cross-section with a thickness of 12.7m and a width of 25.5m is used, as shown in Fig. 5. Four different soil types are involved, i.e., clay, silt mixtures, sand mixtures and sand, with the corresponding SBT values of 3 to 6, respectively. Note that the Layers 1, 3 and 4 extend throughout the horizon, and the Layer 2 exists only around left and right hand sides of the cross-section. It should be clarified that the subsurface condition is unknown in practice, and the synthesized cross-section in Fig. 5 is for validation purpose only in this study. In this study, normalized CPT data, i.e., $Q_t$ and $F_R$, are first simulated within each soil unit with a resolution of 0.1m along both depth and horizontal direction using a 2D random field generator with exponential autocorrelation structure [31]. The random field parameters (e.g., mean, standard deviation (SD), horizontal correlation length $\lambda_h$ and vertical correlation length $\lambda_v$) used for each layer are summarized in the Table 1 [32]. The simulation produces 2D cross-sections of $Q_t$ and $F_R$, respectively, each of which is stored in a matrix with a dimension of 128×256. One pair of realizations of $F_R$ and $Q_t$ is shown in Fig. 6. The simulated $F_R$ and $Q_t$ cross-sections consider not only the spatially varying soil properties but also the spatially varying soil layer boundaries.

**Table 1.** Parameters used in random field simulation

| Layer | Mean of | | SD of | | $\lambda_h$ | | $\lambda_v$ | |
|---|---|---|---|---|---|---|---|---|
| | $\ln F_R$ | $\ln Q_t$ | $\ln F_R$ | $\ln Q_t$ | $\ln F_R$ | $\ln Q_t$ | $\ln F_R$ | $\ln Q_t$ |
| Layer1 | 1.7 | 2.6 | 0.15 | 0.15 | 15m | 15m | 4m | 6m |
| Layer2 | 0.9 | 3 | 0.15 | 0.15 | 30m | 30m | 4m | 6m |
| Layer3 | 0.3 | 3.6 | 0.15 | 0.15 | 25m | 25m | 4m | 6m |
| Layer4 | -0.3 | 4.4 | 0.15 | 0.15 | 20m | 20m | 4m | 6m |

Suppose only five CPT soundings are performed in the 2D vertical cross-section, as denoted by black dash lines (e.g., M1-M5) in Fig. 6. CPT profiles at M1-M5 (see Fig. 7) will be used as input to the 2D BCS interpolation, and the complete $F_R$ and $Q_t$ data over the 2D vertical cross-section (i.e., Figs. 6a&6b) are used for comparison

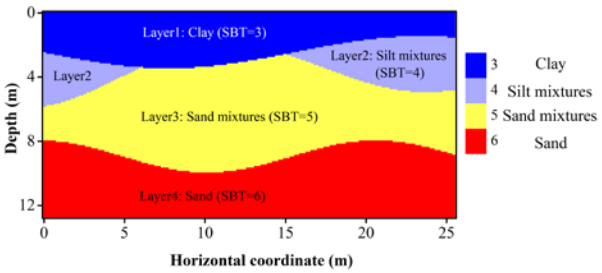



**Figure 5.** A synthesized 2D cross-section with different soil zones

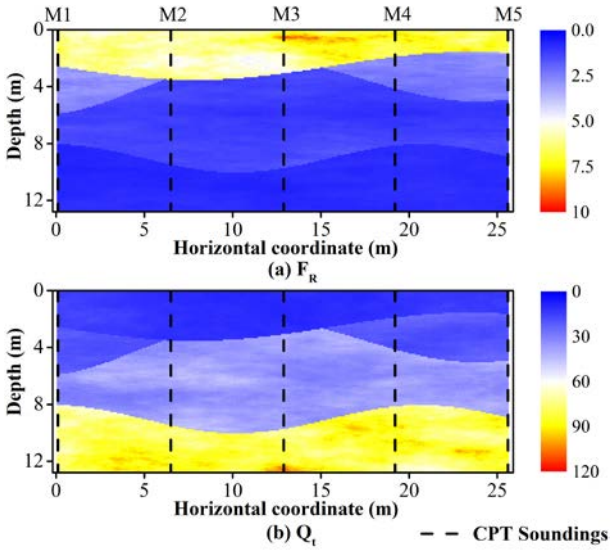



**Figure 6.** Simulated $F_R$ and $Q_t$ data in a 2D vertical cross-section with spatially varying soil layer boundaries

and validation purposes only. Note that in this simulated example the zonation boundaries are clearly defined within the 2D vertical cross-section whereas in real case transitional regions may be expected. The transitional effect or thin layer effect can also be considered in this method.

### 3.1. 2D interpolation of CPT data in a 2D vertical cross-section

Before implementing 2D interpolation using 2D BCS, five sets of CPT sounding data are used to construct two measurement data matrices **Y** (i.e., one matrix for $F_R$, and the other for $Q_t$). Each CPT profile (i.e., subplot in Fig. 7) has 128 data points, therefore two **Y** matrices with a dimension of 128×5 are constructed by Eq. (2) (i.e., one for $F_R$, and the other for $Q_t$).

After inputting the above information, 2D interpolations are performed for $F_R$ and $Q_t$, respectively, using 2D BCS, and the best estimate $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$ for the 2D $F_R$ and $Q_t$ vertical cross-section are obtained using Eq. (4). Note that the interpolated 2D cross-section has a dimension of 128×256. The interpolation results of $F_R$ and $Q_t$ are demonstrated in Fig. 8.

The 2D colormaps in Fig. 8 are generally consistent with that in Fig. 6. The 2D BCS provides reasonable interpolation results for the non-stationary and anisotropic CPT data. The interpolated 2D cross-section can be regarded as a high-resolution image of CPT data (e.g., one data point per 0.01m$^2$ in this case) recovered from limited sounding data. Fig. 9 compares the original

CPT data profiles at four unsampled locations (i.e., U1-U4 in Fig. 8) with those interpolated ones.

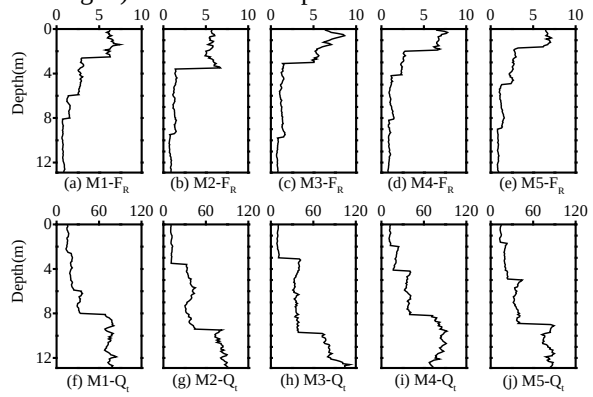



**Figure 7.** Profiles of simulated CPT soundings at M1-M5

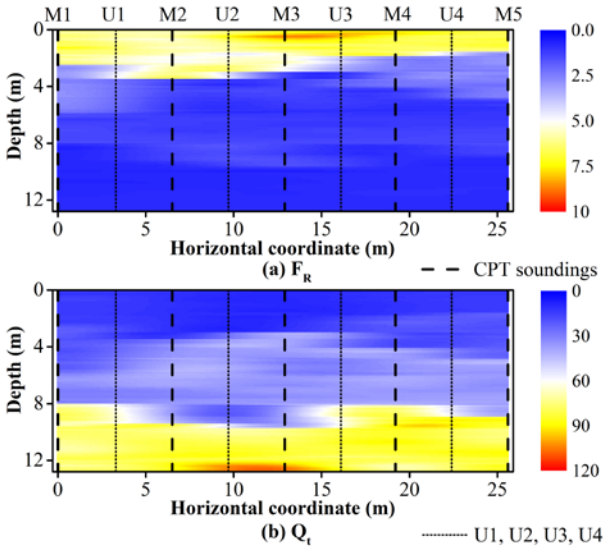



**Figure 8.** Interpolation results of $F_R$ and $Q_t$ from 5 CPT soundings using 2D BCS

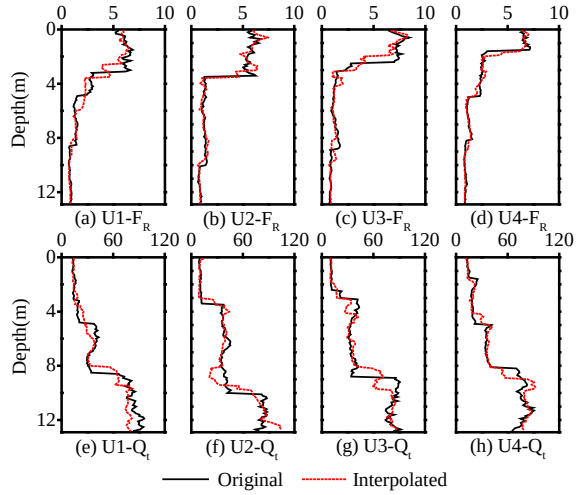



**Figure 9.** Comparisons between the original CPT profiles and interpolated profiles at four unsampled locations U1-U4

The locations of U1 to U4 are marked in Fig. 8 by dotted lines, and they are relatively far away from any of the five CPT soundings. In Fig. 9, the original and interpolated profiles are shown by black solid line and red dotted line, respectively. It shows that the global trends of the CPT data interpolated by BCS are consistent with the original ones, although some details or abrupt changes are not captured accurately. Then, the interpolation results are used to determine the SBT at each of the 128×256 points in the 2D vertical cross-section, as described in the following subsection.

## *3.2. Soil classification using SBT chart*

The SBT at each location in the 2D vertical cross-section can be determined using a SBT chart (see Fig. 2) and corresponding $F_R$ and $Q_t$ data at that location. Similarly, SBT at all 128×256 locations are determined, leading to an SBT map or image. Fig. 10a shows the 2D SBT image obtained from the BCS interpolation results. The underlying soil layer boundaries are also plotted by black solid lines for comparison in Fig. 10a. Fig. 10a shows that most SBTs in the 2D vertical cross-section obtained from the proposed method agree well with the original cross-section (see Fig. 5). To clearly investigate the performance of the proposed method at unsampled locations, four 1D profiles at unsampled locations (i.e., U1-U4 in Fig. 10a) are plotted in Fig. 11. Fig. 11 compares SBT values obtained from the proposed method with their corresponding original SBT values. The original SBT versus depth in these four profiles are shown by black crosses, and the SBT values obtained from proposed method are shown by red diamonds. Fig. 11 shows that most crosses are overlapped with red diamonds. The SBT values obtained from the proposed method with 5 CPT soundings are in good agreement with the original ones at the unsampled locations.

In addition, the accuracy of the SBT image obtained from the proposed method is evaluated quantitatively. The accuracy for each soil layer is quantified by a ratio between the number of points with correctly obtained SBT values and the total number of points in the soil layer. The ratio ranges from 0% to 100%. For example, in the original cross-section shown in Fig. 5, there are 6696 points representing the clay zone (i.e., Layer 1), among which SBT values are correctly obtained at 6525 points when using the proposed method with five CPT soundings. Therefore, the accuracy is calculated as 6525/6696 = 97.5%. Table 2 summarizes the accuracy calculated for the four layers shown in Fig. 5. The second column in Table 2 gives numbers of data points within each of the four layers, while the third column summarizes the corresponding number of SBT values correctly obtained from the proposed method with five CPT soundings. The accuracies are presented in parenthesis. The accuracies are calculated as 97.5%, 82.7%, 94.6% and 93.9% for Layer 1 to 4, respectively, with a total accuracy of 93.6% for four soil layers together. The classification and zonation results obtained from the proposed method are accurate and reasonable.

**Table 2.** Effect of number of CPT soundings on the accuracy of the SBT image

| Layer | Number of points | Number of correctly obtained points | |
|---|---|---|---|
| | | 5 CPT soundings | 20 CPT soundings |
| Layer1 | 6696 | 6525 (97.5%) | 6658 (99.4%) |
| Layer2 | 3599 | 2976 (82.7%) | 3393 (94.3%) |
| Layer3 | 12441 | 11766 (94.6%) | 12364 (99.4%) |
| Layer4 | 10032 | 9417 (93.9%) | 9839 (98.1%) |
| Total | 32768 | 30684 (93.6%) | 32254 (98.4%) |

### 3.3. Boundary delineation for soil layers and zones

Boundaries of soil layers/zones can be delineated automatically using an edge detection method. A Canny filter is used in this study to estimate an SBT gradient map from the SBT map shown in Fig. 10a. By convolving a Canny filter with the SBT image in both directions, the SBT gradient map is obtained, which is colored coded and shown in Fig. 10b. Locations with maximun gradient are denoted by red dots in Fig. 10b. The red dots in Fig. 10b divide approximately the cross-section into five zones, and the SBT values for each zone are consistent with the synthesized cross-section. The original boundaries are also shown by black solid lines in Fig. 10b. In general, the red dots are consistent with the black solid lines. The boundaries and zonation obtained from the proposed method properly identify the spatially varying patterns of soil layers/zones boundaries. The challenge in Fig. 1 is well addressed in an objective manner by the method presented in this study.

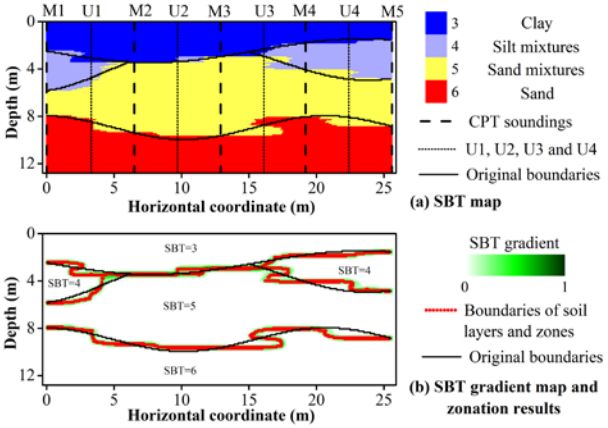



**Figure 10.** Subsurface soil zonation results: (a) SBT map obtained from the proposed method with five CPT soundings; (b) Soil zonation from SBT gradient map

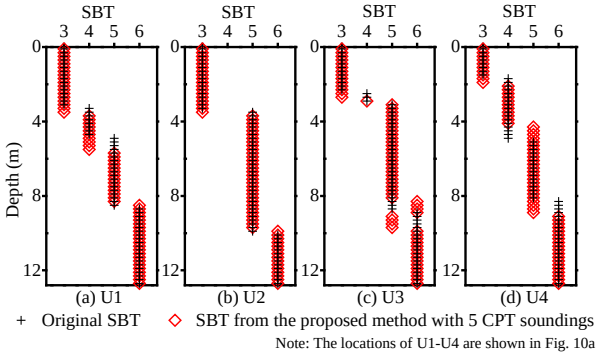



**Figure 11.** Comparison between original SBT profiles and those obtained from the proposed method with 5 CPT soundings

## 4. Effect of adding CPT soundings

It is well-recognized that sample size has important effect on the quality of site characterization [33]. In this Section, the effect of the number of CPT soundings on subsurface soil stratifications and zonation is investigated. One more scenario of CPT sounding scheme, i.e., M=20 is performed for the case shown in Fig. 6.

Figs. 12a&b compare SBT maps obtained from M=5 and M=20 CPT soundings. It shows that the soil layer/zone boundaries obtained from the proposed method become smooth and approach to the black solid lines when M increases to 20. Fig. 13 shows the estimated SBT profiles at the four unsampled locations for the added M=20 scenario. The legends used in Fig. 11 are also used in Fig. 13. When M increases to 20, more and more black crosses are overlapped by red diamonds. The subsurface soil stratification and zonation obtained from the method are more accurate and reliable when more CPT soundings are available. The fourth column of Table 2 summarizes the accuracy for the M=20 scenarios. As M increases to 20, the accuracy for each layer improves. For example, the accuracy ratio for Layer 1 increases from 97.5% at M = 5 to 99.4% at M = 20.

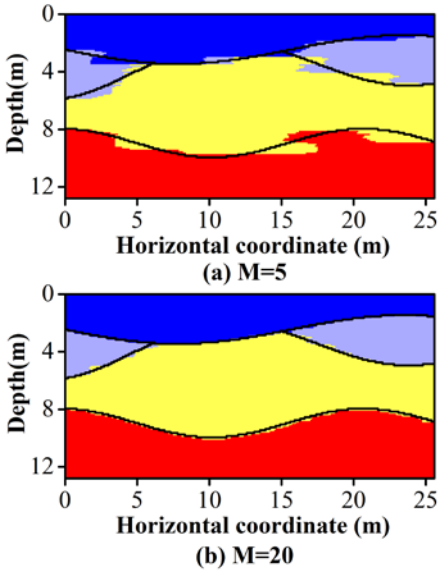



**Figure 12.** Effect of adding CPT soundings on obtained SBT map

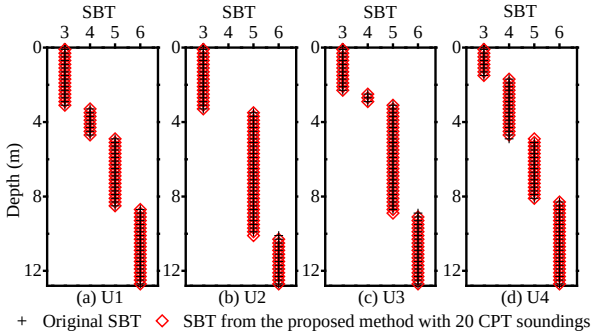



**Figure 13.** Effect of adding CPT soundings on interpolated SBT profiles

## 5. Summary

A CPT-based method is presented in this paper for soil classification and zonation in a two-dimensional (2D) vertical cross-section. The difficulty in interpreting stratification and zonation from limited CPT sounding is addressed in data-driven and non-parametric manners. The method can apply to multiple soil layers simultaneously and bypass the stationary assumption in conventional geostatistics. In addition, it does not require a selection of a parametric form of semi-variogram functions, or an estimation of semi-variogram parameters, for CPT data in a

2D vertical cross-section. The proposed method consists of three modules, and the implementation procedures have been summarized into a flowchart. The method provides high-resolution 2D image for CPT and SBT data in a vertical cross-section, similar to results from geophysical tests.

A simulated example is used to illustrate performance of the method, and the method is showed to perform well. Both spatial variability of CPT data and layer boundary fluctuation are characterized rationally. Moreover, as the number of CPTs increases, the results obtained from the method converge to the underlying cross-section.

## *Acknowledgement*

The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. T22-603/15N and CityU 11213117). The financial supports are gratefully acknowledged.

## *References*

[1] Robertson, P., and Wride, C. "Evaluating cyclic liquefaction potential using the cone penetration test", Canadian Geotechnical Journal, 35(3), pp. 442-459, 1998. https://doi.org/10.1139/t98-017

[2] Fenton, G.A. "Random field modeling of CPT data", Journal of geotechnical and geoenvironmental engineering, 125(6), pp. 486-498, 1999. https://doi.org/10.1061/(ASCE)1090-0241(1999)125:6(486)

[3] Wang, Y., Au, S. K., and Cao, Z. "Bayesian approach for probabilistic characterization of sand friction angles", Engineering Geology, 114(3-4), pp. 354-363, 2010. https://doi.org/10.1016/j.enggeo.2010.05.013

[4] Wang, Y., Fu, C., and Huang, K. "Probabilistic assessment of liquefiable soil thickness considering spatial variability and model and parameter uncertainties", Geotechnique, 67(3), pp. 228-241, 2017. https://doi.org/10.1680/jgeot.15.P.219

[5] Jamshidi Chenari, R., Kamyab Farahbakhsh, H., Heidarie Golafzani, S., and Eslami, A. "Non-stationary realisation of CPT data: considering lithological and inherent heterogeneity", Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, pp. 1-14, 2018. https://doi.org/10.1080/17499518.2018.1447675

[6] Robertson, P. "Soil classification using the cone penetration test", Canadian Geotechnical Journal, 27(1), pp. 151-158, 1990. https://doi.org/10.1139/t90-014

[7] Robertson, P. "Interpretation of cone penetration tests—a unified approach", Canadian Geotechnical Journal, 46(11), pp. 1337-1355, 2009. https://doi.org/10.1139/T09-065

[8] Zhang, Z., and Tumay, M.T. "Statistical to fuzzy approach toward CPT soil classification", Journal of geotechnical and geoenvironmental engineering, 125(3), pp. 179-186, 1999. https://doi.org/10.1061/(ASCE)1090-0241(1999)125:3(179)

[9] Hegazy, Y.A., and Mayne, P.W. "Objective site characterization using clustering of piezocone data", Journal of geotechnical and geoenvironmental engineering, 128(12), pp. 986-996, 2002. https://doi.org/10.1061/(ASCE)1090-0241(2002)128:12(986)

[10] Phoon, K. K., Quek, S., and An, P. "Identification of statistically homogeneous soil layers using modified Bartlett statistics". Journal of Geotechnical and Geoenvironmental Engineering, 129(7), pp. 649-659, 2003. https://doi.org/10.1061/(ASCE)1090-0241(2003)129:7(649)

[11] Das, S.K., and Basudhar, P.K. "Utilization of self-organizing map and fuzzy clustering for site characterization using piezocone data". Computers and Geotechnics, 36(1-2), pp. 241-248, 2009. https://doi.org/10.1016/j.compgeo.2008.02.005

[12] Wang, Y., Huang, K., and Cao, Z. "Probabilistic identification of underground soil stratification using cone penetration tests", Canadian Geotechnical Journal, 50(7), pp. 766-776, 2013. https://doi.org/10.1139/cgj-2013-0004

[13] Cao, Z., Zheng, S., Li, D., and Phoon, K. "Bayesian identification of soil stratigraphy based on soil behaviour type index", Canadian Geotechnical Journal, 56(4), pp. 570-586, 2018. https://doi.org/10.1139/cgj-2017-0714

[14] Cao, Z., and Wang, Y. "Bayesian approach for probabilistic site characterization using cone penetration tests", Journal of Geotechnical and Geoenvironmental Engineering, 139(2), pp. 267-276, 2013. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000765

[15] Mayne, P. W., Christopher, B. R., & DeJong, J. Subsurface Investigations--Geotechnical Site Characterization: Reference Manual (No. FHWA-NHI-01-031). United States. Federal Highway Administration. 2002.

[16] Clayton, C. R., Matthews, M. C., & Simons, N. E. Site investigation. Wiley. 1995.

[17] Mayne, P.W., Christopher, B.R., Berg, R., and DeJong, J. "Subsurface investigations—geotechnical site characterization", Publication Number FHWA - NHI - 01 - 031, National Highway Institute, Federal Highway Administration, Washington, DC. 2002.

[18] Candès, E.J., Romberg, J.K., and Tao, T. "Stable signal recovery from incomplete and inaccurate measurements", Communications on pure and applied mathematics, 59(8), pp. 1207-1223, 2006. https://doi.org/10.1002/cpa.20124

[19] Ji, S., Xue, Y., and Carin, L. "Bayesian compressive sensing", IEEE Transactions on Signal Processing, 56(6), pp. 2346-2356, 2008. https://doi.org/10.1109/TSP.2007.914345

[20] Wang, Y., and Zhao, T. "Interpretation of soil property profile from limited measurement data: a compressive sampling perspective", Canadian Geotechnical Journal, 53(9), pp. 1547-1559. 2016. https://doi.org/10.1139/cgj-2015-0545

[21] Fang, Y., Wu, J., and Huang, B. "2D sparse signal recovery via 2D orthogonal matching pursuit", Science China Information Sciences, 55(4), pp. 889-897, 2012. https://doi.org/10.1007/s11432-012-4551-5

[22] Zhao, T., Hu, Y., and Wang, Y. "Statistical interpretation of spatially varying 2D geo-data from sparse measurements using Bayesian compressive sampling", Engineering Geology, 246, pp. 162-175. 2018. https://doi.org/10.1016/j.enggeo.2018.09.022

[23] Wang, Y., Hu, Y., and Zhao, T. "CPT-based subsurface soil classification and zonation in a 2D vertical cross-section using Bayesian compressive sampling", Canadian Geotechnical Journal, 2019. https://doi.org/10.1139/cgj-2019-0131.

[24] Wang, Y., and Zhao, T. "Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling", Géotechnique, 67(6): 523-536, 2017. https://doi.org/10.1680/jgeot.16.P.143

[25] Tipping, M.E., "Sparse Bayesian learning and the relevance vector machine", Journal of machine learning research, 1(Jun), pp. 211-244, 2001.

[26] Wang, Y., Zhao, T., Hu, Y., and Phoon, K. "Simulation of Random Fields with Trend from Sparse Measurements without Detrending", Journal of Engineering Mechanics, 145(2), 04018130, 2019. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001560

[27] Montoya-Noguera, S., Zhao, T., Hu, Y., Wang, Y., and Phoon, K. K. "Simulation of non-stationary non-Gaussian random fields from sparse measurements using Bayesian compressive sampling and Karhunen-Loève expansion", Structural Safety, 79, pp. 66-79. 2019. https://doi.org/10.1016/j.strusafe.2019.03.006

[28] Hu, Y., Zhao, T., Wang, Y., Choi, C., and Ng, C. W. W. "Direct simulation of two-dimensional isotropic or anisotropic random field from sparse measurement using Bayesian compressive sampling", Stochastic Environmental Research and Risk Assessment, 33(8-9), pp. 1477-1496, 2019. https://doi.org/10.1007/s00477-019-01718-7

[29] Ziou, D., and Tabbone, S. "Edge detection techniques-an overview", Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenii, 8, pp. 537-559, 1998.

[30] Canny, J. "A computational approach to edge detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, (6), pp. 679-698, 1986.

[31] Dietrich, C., and Newsam, G. "A fast and exact method for multidimensional Gaussian stochastic simulations", Water Resources Research, 29(8), pp. 2861-2869, 1993. https://doi.org/10.1029/93WR01070

[32] Phoon, K.-K. & Kulhawy, F.H. Characterization of geotechnical variability. Canadian Geotechnical Journal, 36, 612-624,1999. https://doi.org/10.1139/t99-038

[33] Wang, Y., Guan, Z., and Zhao, T. "Sample size determination in geotechnical site investigation considering spatial variation and correlation", Canadian Geotechnical Journal, 56(7), pp. 992-1002. 2019. https://doi.org/10.1139/cgj-2018-0474