*Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)*
*Edited by* Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly
©2022 ISGSR Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_05-003-cd



# Machine Learning of Subsurface Geological Model for Assessment of Reclamation Induced Consolidation Settlement

Chao Shi[1] and Yu Wang[2]

[1]Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China.
E-mail: chaoshi6-c@my.cityu.edu.hk
[2]Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China.
E-mail: yuwang@cityu.edu.hk

**Abstract:** Reclamation is an effective method to create buildable lands for congested coastal megacities such as Hong Kong and Macau. The greatest geotechnical risk associated with reclamation works is consolidation, which is a time-dependent process of pore water expulsion and ground settlement. An accurate evaluation of consolidation requires a sound understanding of spatial distribution of subsurface soil layer boundaries and spatial variability of soil consolidation parameters from limited site-specific measurements such as cone penetration tests. It is common practice to determine subsurface stratigraphic boundaries using straight lines to connect the same stratigraphy revealed from adjacent measurements, and assume deterministic soil consolidation parameters for consolidation analysis. This simplified practice gains popularity among engineering practitioners due to its convenience for implementation. However, great difficulties may occur when complex geology (e.g., interbedded soil layers) is encountered. More importantly, a false interpretation of subsurface stratigraphy from limited data may fail to identify the most critical design scenario, thus pose significant risks to safety and serviceability of a geotechnical system. In this study, a unified framework is proposed to assess reclamation induced consolidation settlement with explicit consideration of stratigraphic uncertainty and spatial variability of consolidation parameters. Consolidation settlements associated with different combinations of geological realizations and geotechnical random field samples are calculated using the classical 1D consolidation theory. Performance of the proposed unified framework is demonstrated using an illustrative example. Results indicate that the framework can provide accurate evaluation of ground differential settlement with quantified uncertainty.

Keywords: Probabilistic analysis; Geological uncertainty; Convolutional neural network; Bayesian Compressive Sensing.

## 1 Introduction

Reclamation is considered an effective way of creating buildable lands in coastal metropolitan cities. The major geotechnical risk associated with land reclamation is the primary consolidation of fine-grained materials under the deadweight of future superstructures. In definition, consolidation is a time-dependent process of excess pore water pressure expulsion under a sustained working load and can be significantly affected by the spatial distribution of soil consolidation parameters (e.g., permeability) and spatial distribution of soil drainage boundaries. In addition, consolidation is often accompanied with differential ground settlements, which is a key design indicator for regulating reclamation design (e.g., Burland 1997). An accurate evaluation of reclamation induced ground settlement requires effective tools to simultaneously delineate spatial distribution of subsurface soil layer boundaries and model spatial variability of soil parameters in a quantitative and objective manner.

Cone Penetration Test (CPT) is considered the most direct in-situ testing method for investigating subsurface soil conditions. Engineers normally rely on CPT measurements (i.e., cone pressure, sleeve friction and pore pressure) for deriving consolidation parameters such as preconsolidation pressure based on the assumption that soil parameters are deterministic and homogeneous for a single soil layer. This simplified practice does not consider intrinsic variability of soil parameters and can possibly lead to a false interpretation of consolidation mechanism. Regarding soil layer boundaries, linear interpolation, which simply connects the same soil layer boundaries between adjacent measurements, is the most commonly adopted due to its convenience. For simple soil layer patterns, this linear practice may be acceptable. However, large difficulties may be encountered for complex subsurface stratigraphy (e.g., Shi and Wang 2021a).

To address the abovementioned challenge, a data-driven framework is proposed to assess reclamation induced ground settlement with explicit consideration of stratigraphic uncertainty and spatial variability of soil parameters. Multiple geological realizations of subsurface stratigraphy are generated from limited site-specific data and a training image, reflecting prior geological knowledge, using a stochastic simulation tool. Multiple random field samples (RFSs) of geotechnical properties associated with each geological realization are modelled using Bayesian Compressive Sensing (BCS). Performance of the proposed unified framework for the assessment of reclamation induced ground settlement is demonstrated using an illustrative example.

## 2 Proposed unified framework for consolidation settlement analysis

The proposed framework for reclamation induced settlement mainly consists of four key steps. The first step involves the interpretation of soil types from CPT measurements based on derived soil behavior type index, $I_c$. The second step mainly deals with the development of 2D subsurface geological cross-sections from interpreted soil types and prior geological knowledge using a data-driven algorithm, iterative convolution eXGBoost (IC-XGBoost). Following the principle of stochastic simulation, multiple geological realizations can be generated. For each geological cross-section, multiple RFSs of relevant geotechnical parameters are generated using BCS in the third step. In step 4, the reclamation induced ground settlement is assessed using different combinations of geological realizations and RFSs of soil properties under a framework of Monte Carlo simulation (MCS). In the following sections, only key components of the proposed framework are discussed.

### 2.1 CPT-based soil classification and interpretation of consolidation parameters

Soil types of each CPT sounding can be determined based on the soil behavior type (SBT) index, $I_c$, proposed by Robertson and Cabal (2012). The index is calculated based on measured cone pressure $q_c$, sleeve friction $f_s$, and pore pressure $u_2$ following the below equation:

$$I_c = ((3.47 - logQ_t)^2 + (logF_r + 1.22)^2)^{0.5} \tag{1}$$

where $Q_t$ and $F_r$ denote normalized penetration resistance and normalized friction ratio. With derived $I_c$ values, it is possible to classify soil types based on the SBT chart established by Robertson (2012). For instance, soil is classified to be silty clay to clay if $I_c$ varies between 2.95 and 3.6.

In addition, consolidation parameters can also be derived from CPT measurements. For example, pre-consolidation pressure ($\sigma_p$') and over-consolidation ratio ($OCR$) can be calculated using the below equations:

$$\sigma'_p = k \times (q_t - \sigma_{vo}) \tag{2}$$

$$OCR = k \times \frac{(q_t - \sigma_{vo})}{\sigma'_{v0}} \tag{3}$$

where $q_t$ represents cone pressure corrected for pore pressure; $\sigma_{v0}$ and $\sigma_{v0}$' are total and effective vertical stresses; $k$ is a constant and ranges between 0.2 and 0.5. In this study, an average value of 0.35 is adopted for $k$ (Kulhawy and Mayne 1990). Apart from $\sigma_p$' and $OCR$, other parameters relevant to consolidation analysis are determined from laboratory tests.

### 2.2 Stratigraphic uncertainty modelling by IC-XGBoost2D

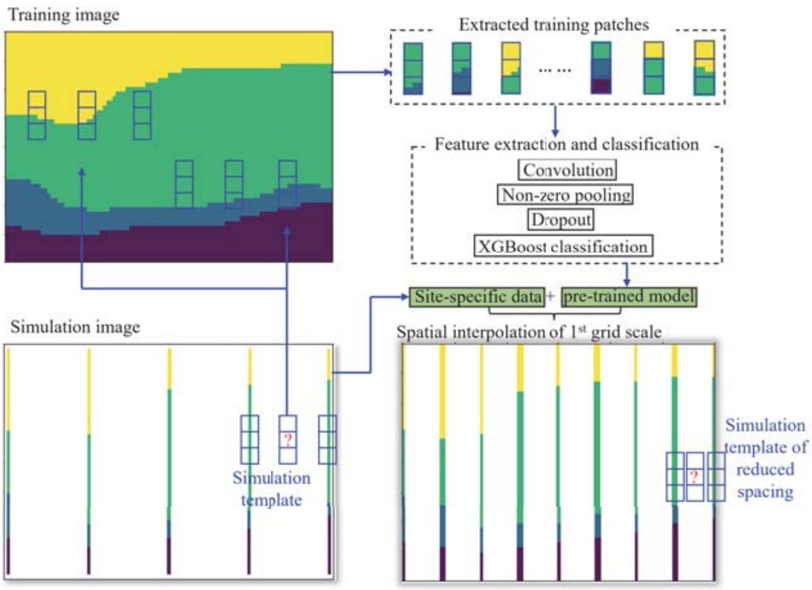
IC-XGBoost is a stochastic simulation tool for delineating 2D subsurface geological cross-sections based on limited site-specific data and a single training image reflecting prior geological knowledge. A qualified training image can be borrowed from nearby site with the similar geological settings (Heim 1990). Fig. 1 shows the key components of IC-XGBoost algorithm. For the spatial interpolation of the 1st grid scale in Fig. 1a, a simulation template, i.e., three distant columns of cells, compatible to the simulation image is first determined and transferred to the single training image to extract all the potential stratigraphic patterns. The collected training patches are then further processed and refined via a series of operations such as convolution, non-zero pooling and dropout before feeding to XGBoost for building a classification model. The trained classification model is then combined with the site-specific data for the determination of all soil types in the middle of any two adjacent line measurements, thus complete the spatial interpolation of the 1st grid scale. Subsequently, simulation template of reduced spacing is adaptively determined from the 1st interpolation result for the 2nd round of spatial interpolation. As shown in Fig. 1b, the whole process repeats iteratively until all the unknown soil types in the simulation image are determined, completing a geological cross-section or a realization $Z_r$. By changing random seeds, multiple geological realizations can be generated and the most probable interpolation $Z_{mp}(\boldsymbol{x})$ can be derived by assigning each spatial point $\boldsymbol{x}$ with soil type of the highest frequency. The stratigraphic uncertainty associated with $Z_{mp}(\boldsymbol{x})$ can be quantified using a dispersion measure, $Dp(\boldsymbol{x})$:

$$Dp(\boldsymbol{x}) = \frac{\sum_{r=1}^{N_r} \boldsymbol{I}(Z_r(\boldsymbol{x}) \neq Z_{mp}(\boldsymbol{x}))}{N_r} \tag{4}$$

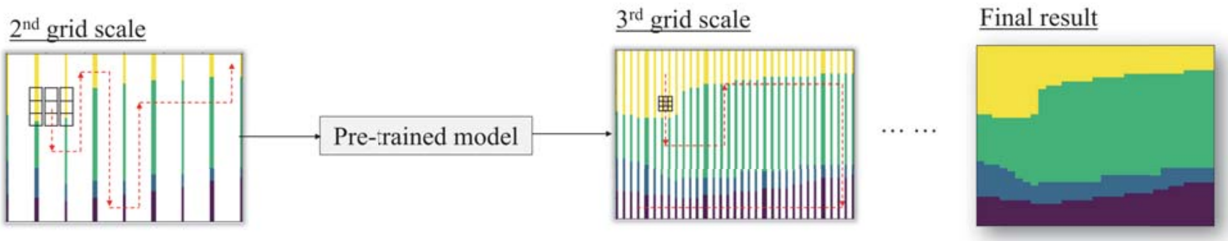
where $N_r$ denotes the number of realizations; $\boldsymbol{I}(\cdot)$ is an indicator function and has a value of one if the condition within the parenthesis is true, or zero otherwise. For a training example where the ground truth geological cross-section is available, the accuracy, $Acc$, associated with $Z_{mp}(\boldsymbol{x})$ can be calculated as follows:

$$Acc = \frac{\sum_{i=1}^{N_h \times N_v} \boldsymbol{I}(Z_T(\boldsymbol{x}_i) = Z_{mp}(\boldsymbol{x}_i))}{N_h \times N_v} \tag{5}$$

where $N_h$ and $N_v$ stand for total sampled points in the horizontal and vertical directions, respectively; $Z_T$ denotes the test image. Note that in real engineering applications, the ground truth is unknown and therefore, the prediction accuracy cannot be calculated. In this study, the test cross-section is used for validation purpose only.




(a) Spatial interpolation of 1st grid scale




(b) Iterative prediction

**Figure 1.** Key components of IC-XGBoost (after Shi and Wang 2021b)

### 2.3 Modelling soil property spatial variability from limited measurements

BCS is a Bayesian supervised learning algorithm developed for spatial interpolation of spatially varying geo-properties from limited measurements (Wang and Zhao 2017, Wang et al. 2017). BCS is established based on the fact that geotechnical signals are compressive and can be concisely represented as a weighted summation of a limited number of pre-specified basis function. Accordingly, a spatially varying 2D geotechnical signal $\boldsymbol{F}$ can be represented as follows:

$$\boldsymbol{F} = \sum_{t=1}^{N_{x_h} \times N_{x_v}} \boldsymbol{B}_t^{2D} \omega_t^{2D} \tag{6}$$

where $N_{xh}$ and $N_{xv}$ denote total sampled points in the horizontal and vertical directions; $\boldsymbol{B}_t^{2D}$ stands for the $t$-th basis function, whose corresponding weight is $\omega_t^{2D}$. Note that $\boldsymbol{B}_t^{2D}$ can be constructed using readily available tools in popular programming software such as python package scipy.fftpack.dct. Due to the compressibility of a signal, most elements of $\omega_t^{2D}$ are zero and those non-zero components can be derived from limited site-specific data $\boldsymbol{Y}$, which is a sub-matrix of $\boldsymbol{F}$:

$$\boldsymbol{Y} = \sum_{t=1}^{N_{x_h} \times N_{x_v}} \boldsymbol{\psi}_{x_h} \boldsymbol{B}_t^{2D} \boldsymbol{\psi}_{x_v} \omega_t^{2D} = \sum_{t=1}^{N_{x_h} \times N_{x_v}} \boldsymbol{A}_t^{2D} \omega_t^{2D} \tag{7}$$

where $\boldsymbol{\psi}_{xh}$ and $\boldsymbol{\psi}_{xh}$ stand for problem-specific measurement matrices, denoting the positions of the measured $\boldsymbol{Y}$ in $\boldsymbol{F}$. With the estimated weight coefficient $\widehat{\omega}_t^{2D}$, the original 2D signal can be approximated as $\widehat{\boldsymbol{F}}$:

$$\widehat{\boldsymbol{F}} = \sum_{t=1}^{N_{x_h} \times N_{x_v}} \boldsymbol{B}_t^{2D} \widehat{\omega}_t^{2D} \tag{8}$$

Note that the reconstructed 2D signal contains significant statistical uncertainty as the weight coefficient is estimated from limited site-specific measurements. The associated statistical uncertainty can be explicitly considered under a Bayesian framework:

$$p(\hat{\boldsymbol{\omega}}^{2D}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\hat{\boldsymbol{\omega}}^{2D}) \cdot p(\hat{\boldsymbol{\omega}}^{2D})}{p(\boldsymbol{Y})} \tag{9}$$

where $p(\boldsymbol{Y}|\hat{\boldsymbol{\omega}}^{2D})$ is the likelihood function representing the occurrence probability of $\boldsymbol{Y}$ given $\hat{\boldsymbol{\omega}}^{2D}$; $p(\hat{\boldsymbol{\omega}}^{2D})$ is the prior probability density function (PDF); $p(\boldsymbol{Y})$ is the evidence term which ensures the integration of the posterior probability $p(\hat{\boldsymbol{\omega}}^{2D}|\boldsymbol{Y})$ equals to one. The explict mathematical forms in Eq. (9) are referred to Zhao and Wang (2020). More specifically, multiple RFSs of $\hat{\boldsymbol{\omega}}^{2D}$ can be generated using Markov Chain Monte Carlo (MCMC) simulation such as Gibbs sampling. Once RFSs of $\hat{\boldsymbol{\omega}}^{2D}$ are obatained, multiple 2D signals $\hat{\boldsymbol{F}}$ can be generated following Eq. (6).

### 2.4 Sequential modelling of soil property spatial variability for each soil type

Fig. 2 shows the framework for the integratoin of IC-XGBoost with BCS. For each generated geological realization, the RFS of CPT properties (e.g., cone pressure) can be sequentially determined for each soil layer. As shown in Fig. 2, a binary transformation is applied to each geological realization to extract stratigraphic boundaries for each soil layer. Subsequently, site-specific CPT properties within each soil layer are sampled and serve as input for BCS algorithm, which is used to interpolate spatially varying CPT data within each soil domain. Finally, the interpolated CPT properties from all different soil layers are collected and combined to yield a full 2D RFS of CPT properties.

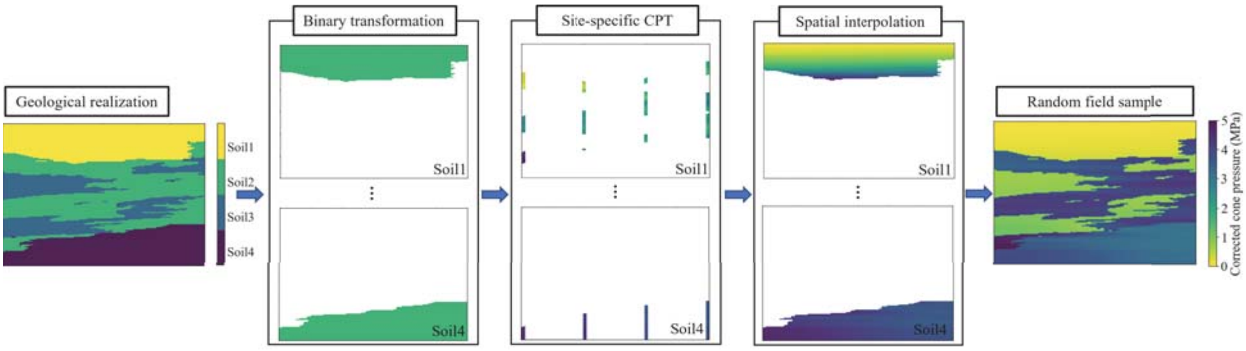



**Figure 2.** Sequential generation of random field samples of CPT cone pressure for a single realization of geological cross-section

### 2.5 Consolidation settlement assessment of fine-grained materials

After generating multiple geological realizations and RFSs of geotechnical properties, the associated consolidation settlement can be evaluated using the classical 1D consolidation theory. In this study, only the primary consolidation settlement is considered:

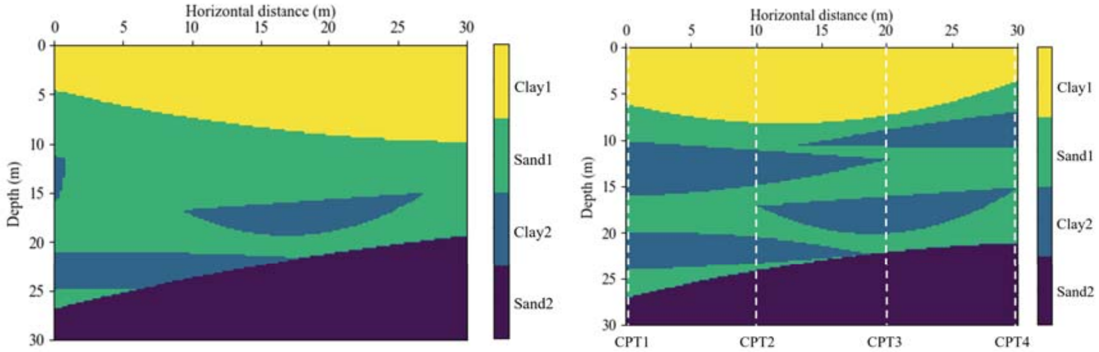
$$s_p = \frac{c_r}{1+e_0} \times H_0 \times \log\left(\frac{\sigma'_{v0}+\Delta\sigma}{\sigma'_{v0}}\right), \text{ for } \sigma'_{v0} + \Delta\sigma < \sigma'_p \tag{10}$$

$$s_p = \frac{c_c}{1+e_0} \times H_0 \times \log\left(\frac{\sigma'_{v0}+\Delta\sigma}{\sigma'_p}\right) + \frac{c_r}{1+e_0} \times H_0 \times \log\left(\frac{\sigma'_p}{\sigma'_{v0}}\right), \text{ for } \sigma'_{v0} < \sigma'_p \text{ and } \sigma'_{v0} + \Delta\sigma > \sigma'_p \tag{11}$$

where $e_0$ denotes initial void ratio; $H_0$ is the initial thickness of a fine-grained soil layer; $\Delta\sigma$ is applied vertical surcharge at the ground surface; $c_c$ and $c_r$ are virgin compression and recompression index, respectively. With calcuated total primary consolidation settlement, assocaited angular distortion $\beta$ can be calculated as follows:

$$\beta = \frac{\delta_{AB}}{l_{AB}} + \frac{\delta_{BC}}{l_{BC}} - \omega \tag{12}$$

where $\delta_{AB}$ and $\delta_{BC}$ are differential settlements between reference points; $l_{AB}$ and $l_{BC}$ are incremental horizontal distances; $\omega$ is tilt. Positive and negative $\beta$ values refer to hog and sag displacment modes, respectively.

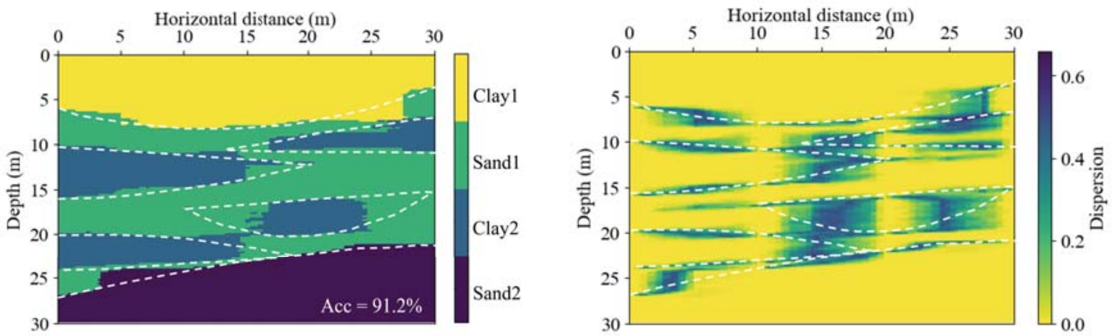



(a) Training image (b) Test image

**Figure 3.** Training and test image.

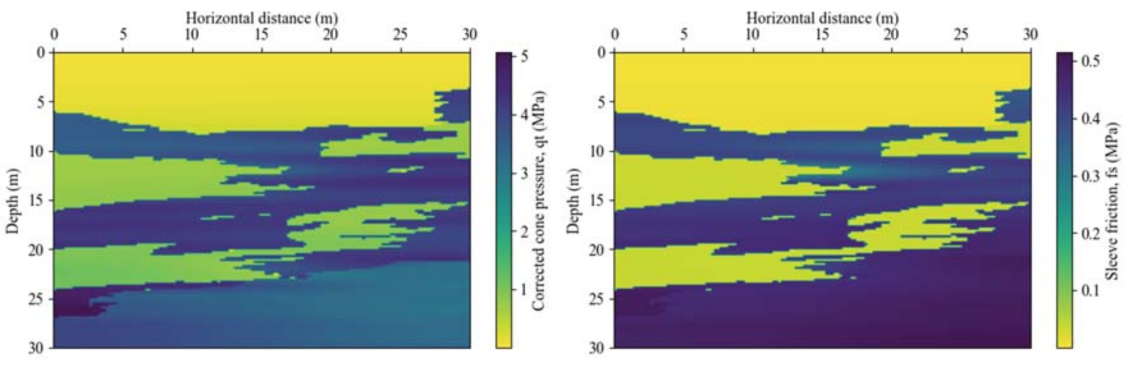
## 3 Illustrative example

In this section, a pair of geological cross-sections modified from a real reclamation site in Hong Kong were collected to demonstrate performance of the proposed method. Fig. 3 shows the extracted training and test images, which may be considerd as synthetic due to various modifications. Both images have a totol horizontal distance of 30m and a total depth of 30m. The resolutions for both directions are taken as 0.2m, resulting in a total of 22500 points. For illustration, four CPTs were taken from the test image as site-specific measurements. In addition, $q_t$ and $f_s$ profiles along CPT soundings are modelled using random field theory. Both $q_t$ and $f_s$ follow lognormal distribution. In addition, the autocorrelation structure for both parameters is taken as exponential function form. The random field parameters adopted in this study are referred to Shi and Wang (2021c).




(a) Most probable interpolation (b) Dispersion

**Figure 4.** Results of IC-XGBoost algorithm




(a) Best estimate of corrected cone pressure (b) Best estimate of sleeve friction

**Figure 5.** BCS results

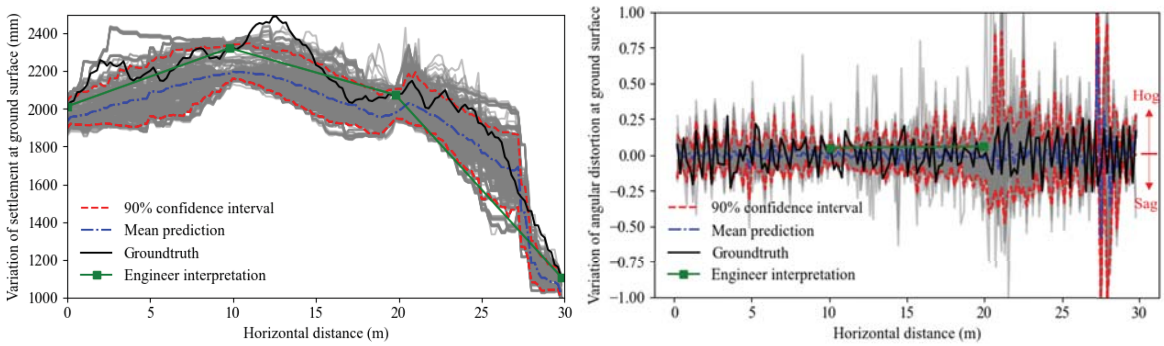
### 3.1 Modelling results for subsurface stratigraphy and soil spatial variability

The time required for generating a single stochastic realization is less than 10 seconds using a computer with Intel(R) Core (TM) i7-10750H and 16GB RAM. Fig. 4 shows the results from IC-XGBoost algorithm conditioning on site-specific data and a single training image. The derived most probable interpolation in Fig. 4a has an accuracy of 91.2% and the associated stratigraphic patterns reasonably capture the overall pattern of soil layer boundaries shown in Fig. 3b. Fig. 4b shows the quantified stratigraphic uncertainty associated with the most probable interpolation using a dispersion plot. Clearly, areas of larger interpolation uncertainties (or high dispersion values) mainly cluster around the interpolated soil layer boundaries.

Fig. 5 shows the BCS modelling results of CPT $q_t$ and $f_s$ for a single geological realization. Fig. 5a shows the spatial distribution of mean $q_t$ profile. Clearly, the fine-grained materials (i.e., Clay1 and Clay2) have relatively low $q_t$ values in comparison with those of coarse-grained materials. In addition, Fig. 5b shows the mean of spatially varying $f_s$ values. There is a significant difference in terms of magnitudes for different soil layers.

### 3.2 Probabilistic consolidation settlement assessment

With generated multiple geological realizations and RFSs of geotechnical properties, the ground surface settlement and associated angular distortion can be derived following Eqs. (10) - (12). As shown in Fig. 6a, the 90% confidence interval can reasonably enclose the ground truth settlement values (i.e., black line). The mean prediction (i.e., blue dashed line) from Monte Carlo simulation reasonably follows the ground truth profile. The discrepancy may possibly be caused by the biased prior geological knowledge reflected in the training image. Fig. 6b shows the derived angular distortion at the ground surface following Eq. (11). Despite of the large variation in calculated angular distortion, the 90% confidence interval can still reasonably enclose the ground truth profile with quantified uncertainty. In comparison, the results from engineer interpretation can only obtain angular distortion values at two sampled points, which emphasizes the importance of incorporating stratigraphic uncertainty and spatial variability of soil properties for assessment of reclamation induced ground settlement.




(a) Ground surface settlement (b) Angular distortion at the ground surface

**Figure 6.** Primary consolidation analysis

## 4 Summary and Conclusions

This study proposed a unified framework for assessment of reclamation induced ground settlement with explicit consideration of stratigraphic uncertainty and spatial variability of soil properties. The subsurface stratigraphy is delineated using IC-XGBoost, which learns typical stratigraphic connectivity from an ensemble training image while conditioning on available site-specific data. Associated spatial variability of soil parameters within each geological realization is modelled using Bayesian Compressive Sensing (BCS). The performance of the proposed method is validated using an illustrative example. Results indicate that the combination of IC-XGBoost and BCS enables an accurate estimation of reclamation induced ground settlement and angular distortion with quantified uncertainty.

#### Acknowledgments

The work described in this paper was supported by grants from the Research Grant Council of Hong Kong Special Administrative Region, China (Project nos. CityU 11202121 and CityU 11213119). The financial support is gratefully acknowledged.

## References

Burland, J. (1997). Assessment of risk of damage to buildings due to tunnelling and excavation. *Earth. Geot. Eng., Balkema*, 1189-1201.

Heim, G.E. (1990). Knowledge of the origin of soil deposits is of primary importance to understanding the nature of the deposit. *Bulletin of the Association of Engineering Geologists*, 27(1): 109-112.

Kulhawy, F.H. and Mayne, P.W., (1990). Manual on estimating soil properties for foundation design (No. EPRI-EL-6800). *Electric Power Research Inst., Palo Alto, CA (USA); Cornell Univ., Ithaca, NY (USA). Geotechnical Engineering Group*.

Shi, C. and Wang, Y. (2021a). Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics, *Canadian Geotechnical Journal*, 58(2): 261-280.

Shi, C. and Wang, Y. (2021b). Development of subsurface geological cross-section from limited site-specific boreholes and prior geological knowledge using iterative convolution XGBoost. *Journal of Geotechnical and Geoenvironmental Engineering*, 147(9), 04021082.

Shi, C. and Wang, Y. (2021c). Assessment of reclamation-induced consolidation settlement considering stratigraphic uncertainty and spatial variability of soil properties. *Canadian Geotechncial Journal*, https://doi.org/10.1139/cgj-2021-0349.

Robertson, P.K. and Cabal, K.L., (2012). Guide to cone penetration testing for geotechnical engineering. *Gregg Drilling & Testing.*

Wang, Y. and Zhao, T., (2017). Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. *Géotechnique*, 67(6), pp.523-536.

Wang, Y., Zhao, T., and Phoon, K. (2017). Direct simulation of random field samples from sparsely measured geotechnical data with consideration of uncertainty in interpretation. *Canadian Geotechnical Journal*, 55(6), pp.862-880.

Zhao, T. and Wang, Y. (2020). Non-parametric simulation of non-stationary non-gaussian 3D random field samples directly from sparse measurements using signal decomposition and markov chain monte carlo (MCMC) simulation, *Reliability Engineering & System Safety*, 203: 107087.