Planning of Site Investigation for Subsurface Shallow Gas Pressure: A Smart Sampling Strategy

Zheng Guan[1], Zijun Cao[2], Yi Hong[3], and Yu Wang[4]

[1]*PhD student, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. Email: zhengguan3-c@my.cityu.edu.hk*
[2]*Professor, State Key Laboratory of Water Resources and Hydropower Engineering Science, Institute of Engineering Risk and Disaster Prevention, Wuhan University, 299 Bayi Road, Wuhan 430072, P. R. China. Email: zijuncao@whu.edu.cn*
[3]*Associate Professor, Key Laboratory of Offshore Geotechnics and Material of Zhejiang Province, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, Zhejiang 310058, P.R. China. Email: yi_hong@zju.edu.cn*
[4]*Professor, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. Email: yuwang@cityu.edu.hk*

Abstract: Subsurface gas is widely found in coastal areas all over the world at a shallow depth (about tens to one hundred meters). The main content of subsurface shallow gas is methane (>95%). Therefore, the presence of shallow gas may lead to a significant risk for underground construction and offshore engineering. Currently, there is no scientific and quantitative method in engineering practice for planning of site investigation for gas pressure. To address this issue, this paper presents a smart sampling strategy for planning of site investigation using innovative data analytic methods (e.g., Bayesian compressive sampling, BCS, and information entropy). The smart sampling strategy automatically determines the minimal number of investigation points (e.g., modified CPTs) and their optimal locations. Simulated non-Gaussian and non-stationary gas pressure data are used to illustrate and validate the smart sampling strategy, and the method is shown to perform reasonably well.

Keywords: Planning of site investigation; Bayesian method; Cone penetration test; Compressive sampling; Information entropy

1. Introduction

Subsurface gas is widely found in coastal areas all over the world at a shallow depth (about tens to one hundred meters) (e.g., Li and Lin 2010). The main content of subsurface shallow gas is methane (>95%). Therefore, the presence of shallow gas may lead to a significant risk for underground construction and offshore engineering, such as the potential occurrence of explosion and fire accident during tunnel construction when using tunneling boring machine (e.g., Wang et al. 2018). The high gas pressure leads to an increasing possibility of explosion and risk for geotechnical construction. Therefore, the spatial distribution of shallow gas pressure (i.e., shallow gas pressure profile along horizontal direction) should be properly delineated during site investigation for subsequent hazard analysis and mitigation. In geotechnical engineering practice, cone penetration test (CPT) equipment incorporated with a gas pressure transducer, namely modified CPT is commonly used for measuring gas pressure (e.g., Li et al. 2009). As a rule of thumb, a large number of modified CPT soundings provides an increasing reliability of the interpreted shallow gas pressure profile. However, increasing the number of modified CPT soundings also significantly increases the time and cost of site investigation. The tradeoff between the expenditures of site investigation and reliability of interpreted gas pressure profile suggests a need to develop a scientific method for properly planning a site investigation for gas pressure (i.e., determine the minimal number of modified CPTs and their optimal locations for achieving a target level of reliability).

Although the importance of planning of site investigation is widely recognized, quantitative or scientific method is not available for sampling strategy in geotechnical engineering practice. Current geotechnical design codes and manuals around the world only provide some empirical guidelines for the planning of site investigation. For example, 15-40m, 20-200m, and 25-75m of spacing of investigation points for high-rise buildings, roads, and dams are suggested in Eurocode 7-2 (CEN 2007). However, little scientific evidence is available to support these guidelines. Another key issue for these guidelines is that the number of investigation points is determined in advance. However, little information is available for a certain site to predetermine the proper number of investigation points.

This paper presents a smart sampling strategy for planning of site investigation for subsurface gas pressure based on the work by Guan et al. (2020). The sampling strategy is goal-oriented and progressive that gradually increase the number of modified CPTs until the reliability of interpreted shallow gas pressure profile achieves the target level. The presented smart sampling strategy uses information from previously conducted modified CPT soundings to wisely choose the next optimal location of modified CPT. Thus, even without the prior knowledge about a certain site, the smart sampling strategy can efficiently and adaptively determine the minimal number of modified CPT soundings and their corresponding optimal locations. The presented sampling strategy is illustrated and validated using simulated non-Gaussian and non-stationary gas pressure data.

2. Methodology

The overall philosophy of the smart sampling strategy is to progressively increase the number of modified CPTs until achieving the target reliability level of interpreted shallow gas pressure profile. Bayesian compressive

sampling (BCS) method is used in this study to interpret a complete gas pressure from sparse modified CPT measurements, and the reliability of the interpreted gas pressure profile is quantified in terms of coefficient of variation. Note that BCS is a non-parametric method, and it can rationally and effectively deal with non-Gaussian and non-stationary gas pressure data, which are commonly encountered in geotechnical practice (e.g., Wang and Zhao 2017; Montoya-Noguera et al. 2019; Wang et al. 2019).

The preliminary modified CPTs might be far apart because the target reliability level is achieved by conducting additional modified CPTs in a progressive manner. Such a strategy is also consistent with the geotechnical engineering experience. Note that investigation points are generally sparse during the phase of preliminary design. If the results of preliminary investigation results show that the subsurface condition is significantly uncertain, additional investigation points are performed subsequently. After the starting point of sampling strategy is determined, the initial number of modified CPTs with the optimal locations is performed for interpreting the gas pressure profile. Then, if the reliability of the interpretation result does not satisfy the target level, additional modified CPTs are iteratively performed one by one until the target level of reliability is achieved. In other words, for each iteration, only one modified CPT is conducted at an optimal location, which is selected based on information from previously conducted modified CPT soundings, and the new gas pressure measurement, together with previously measured gas pressure, are used for re-interpreting the gas pressure profile. The iteration process stops when the target reliability level of interpretation results achieves.

2.1 Framework of smart sampling strategy

As shown in Fig. 1, the framework of the smart sampling strategy involves five steps. In Step 1, the target reliability level (i.e., target coefficient of variation, COV_T) shall be first determined depending on the tolerable risk level of the project, expected variability of subsurface conditions, and other project-specific consideration. Then, an initial number, M of modified CPTs is obtained based on engineering experience and existing geotechnical guidelines (e.g., Look 2007; Rix et al., 2018) in Step 2. After that, M modified CPTs are conducted for interpreting a complete gas pressure profile using BCS method, and the reliability level in terms of COV_M is obtained from the interpretation result in Step 3.

In the context of BCS method, a N-length vector $\hat{X} = [\hat{X}_1, \hat{X}_2, ..., \hat{X}_N]$ represents the mean gas pressure profile interpreted from sparse modified CPT measurements, and the vector $\sigma = [\sigma_1, \sigma_2, ..., \sigma_N]$ indicates the corresponding standard deviation. The reliability of the interpreted gas pressure profile (e.g., the coefficient of variation, COV_M) is defined as:

$$COV_M = \frac{\max(\sigma_1, \sigma_2, ..., \sigma_N)}{\mu_{\hat{X}}} \times 100\% \qquad (1)$$

where $\mu_{\hat{X}}$ represents the mean value of the interpreted gas pressure profile.

When COV_M is larger than COV_T, the reliability of interpretation result is smaller than the allowable level, and thus an additional modified CPT is performed at an optimal location for further improving the reliability of interpretation result in Step 4. Note that the optimal location is selected wisely based on previously conducted modified CPT soundings using information entropy-based method. Subsequently, a new measurement from the additional conducted modified CPT together with previous modified CPT measurements are used to re-interpret the complete gas pressure profile. Steps 3-4 are repeatedly performed until the interpretation result achieves the target level (i.e., $COV_M \leq COV_T$), leading to the minimal number of modified CPTs, M, and their corresponding optimal locations in Step 5.
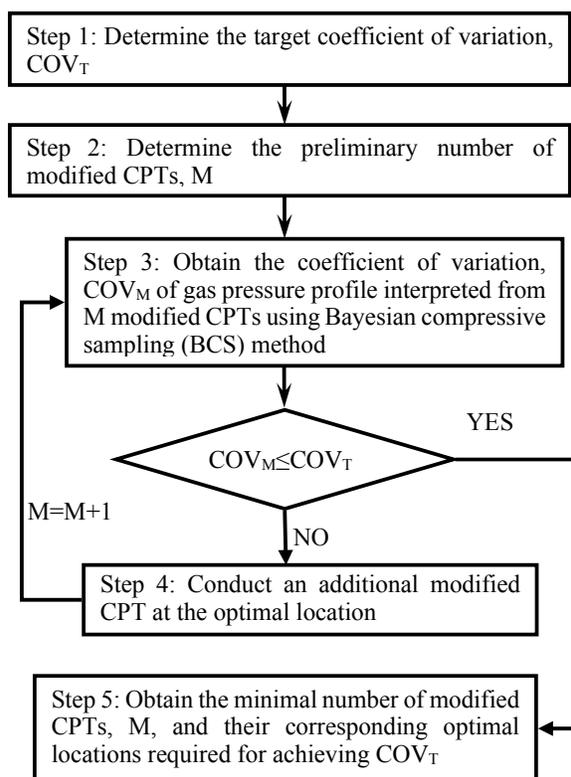

Figure 1. Framework of smart sampling strategy

The determination of the optimal modified CPT locations is a key element in smart sampling strategy, as described in the following subsection.

2.2 Determination of optimal modified CPT locations using information entropy

The gas pressure is varying and correlated along horizontal direction due to complex geological and biological process during the formation of shallow gas, and thus a random process might be used to model the

variation of gas pressure along the horizontal direction (e.g., Vanmarcke 1977). Information entropy is commonly used to quantify the uncertainty of a random process (e.g., Cover and Thomas 2012). The information entropy H of an interpreted gas pressure profile, $\hat{X}$ is expressed as (e.g., MacKay 2003):

$$H(\hat{X}) = -\int p(\hat{X})\ln[p(\hat{X})]d\hat{X} \qquad (2)$$

where $p(\hat{X})$ indicates the probability density function of $\hat{X}$. Using information entropy, the optimal location for an additional modified CPT is the one that maximizes the reduction of current information entropy $H(\hat{X})$ to an updated information entropy $H(\hat{X}_{new})$ after adding an extra modified CPT measurement.

When there are no pre-existing modified CPT measurements, the reduction of information entropy is approximately equal to the information entropy of measurements y, and expressed as:

$$\Delta = H(\hat{X}) - H(\hat{X}_{new}) \approx H(y) \qquad (3)$$

For such a case, Zhao and Wang (2019) show that when the measurement data are equally spaced, the reduction reaches the maximum value. This indicates that the optimal sampling strategy for the preliminary modified CPTs is equally spaced along the horizontal direction.

Note that BCS provides the full probability distribution of interpreted gas pressure profile. Detailed equations and derivations of BCS are referred to Wang and Zhao (2017). Combining the results of BCS and Eq. (2), when one additional modified CPT measurement is taken at the location with the maximum variance, the information entropy reduction reaches the maximum value (Zhao and Wang 2019). In other words, the optimal location of an additional modified CPT is the one with maximum variance along the gas pressure profile interpreted from previous modified CPT measurements. The result is also consistent with common sense in geotechnical site investigation, which indicates that additional investigation points should be performed at the locations with the largest uncertainty. In the next section, a set of simulated gas pressure data is used to illustrate and validate the smart sampling strategy.

3. Simulated data example

In geotechnical engineering practice, gas pressure data are usually non-Gaussian and non-stationary (e.g., Guan et al. 2020). Therefore, the smart sampling strategy is illustrated and validated using a set of gas pressure P_g data along the horizontal direction, which is generated from a non-Gaussian and non-stationary 1D random field. Consider, for example, a lognormal random field with a quadratic trend in mean, $\mu = -1.8\times10^{-6}x^2 + 1.8\times10^{-3}x + 0.2$ (MPa) together with constant standard deviation σ=0.1MPa and exponential correlation function with correlation length λ =50m is used to simulate non-Gaussian and non-stationary gas pressure P_g random field. These random field parameters values are consistent with the real gas pressure data

reported in literature (e.g., Li et al. 2009; Wang et al. 2018). The simulated P_g profile is ranging from 0m to 1023m with a resolution of 1m along the horizontal direction, as shown in Fig. 2.
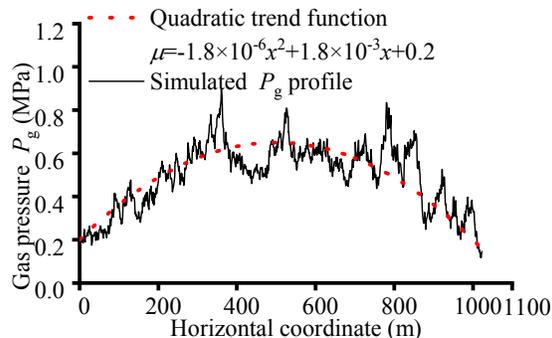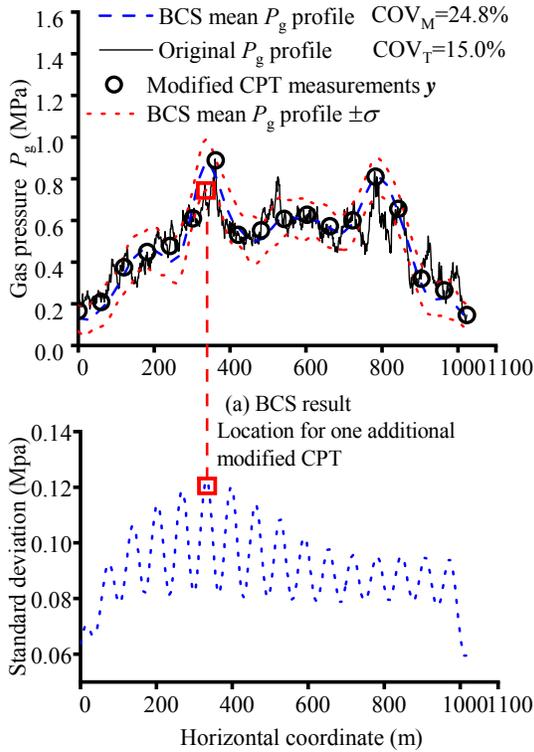

Figure 2. Simulated gas pressure P_g data

Consider, for example, characterization of the spatial variation of this gas pressure P_g profile using the smart sampling strategy. As shown in Fig. 1, the target coefficient of variation, COV_T, shall be first determined. Suppose that COV_T is determined as 15% based on the tolerable risk level of the project (i.e., Step 1). In Step 2, an initial spacing of modified CPT is taken as 60m based on the geotechnical design codes and guidelines (e.g., Rix et al., 2018). In total, $1023/60+1\approx18$ modified CPTs are first performed with equal interval along the horizontal direction. To mimic the conduction and interpretation procedure of these 18 modified CPTs, 18 P_g data are taken from the original gas pressure profile and used as the input to BCS for interpreting the complete gas pressure profile. The interpretation result is shown in Fig. 3(a). In this figure, the dashed line and solid line represent the mean P_g profile interpreted from 18 P_g measurement data and original P_g profile respectively, together with the modified CPT measurement data y denoted by open circles. The reliability of the interpreted P_g profile is also shown in this figure by two dotted lines, i.e., mean P_g profile $\pm\sigma$. The coefficient of variation, COV_M is calculated as 24.8% for M=18, which is larger than the COV_T=15% (i.e., Step 3). Therefore, additional modified CPTs should be carried out to further improve the reliability of the interpretation result.

In Step 4, one additional modified CPT is conducted at the corresponding optimal location. As discussed in subsection 2.2, the optimal location is the one with the maximum variance along with the interpreted P_g profile. The standard deviation of the interpreted P_g profile is shown in Fig. 3(b), it is found from this figure that the maximum standard deviation is located at the 333m. Therefore, an additional modified CPT should be conducted at the location of 333m. Then, the P_g data at the location of 333m is taken from the original P_g profile and used together with the initial 18 P_g data to re-interpret the complete P_g profile. The P_g profile interpreted from M=18+1=19 is shown in Fig. 4 using similar legends in Fig. 3. Fig. 4 shows that the interval

between two dotted lines becomes smaller, which suggests the increasing reliability of the interpretation result. In addition, the mean P_g profile also increasingly converges to the true one. After adding one additional CPT measurement, the COV_M is reduced from 24.8% to 22.0%, which is still larger than the target value, and thus additional modified CPT measurements is required.

Steps 3-4 are repeatedly performed until the target coefficient of variation is achieved. The evolution of COV_M with the number of additional modified CPT measurements is summarized in Fig. 5 by open circles. It is found from Fig. 5 that COV_M generally decreases as the number of additional modified CPT increases. When the number of additional modified CPT increases to 8 (i.e., M=18+8=26), the COV_M decreases to 14.5%, which is smaller than COV_T=15%. In other words, the total number, M=26 of modified CPTs is required for achieving the target reliability level. The P_g profile interpreted from 26 modified CPT measurements is shown in Fig 6. It is observed that the interval between two dotted line in Fig. 6 is significantly smaller than the one interpreted from 18 P_g data. It indicates that the reliability of interpreted P_g profile is greatly improved after performing extra modified CPTs. Using the smart sampling strategy, the minimal number of modified CPTs and their optimal locations are wisely determined for this non-Gaussian and non-stationary gas pressure data.

(a) BCS result

(b) Standard deviation of interpreted P_g profile

Figure 3. Gas pressure P_g profile interpreted from 18 initial modified CPT measurements
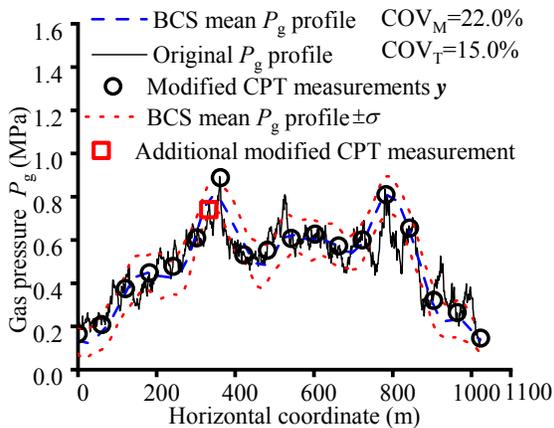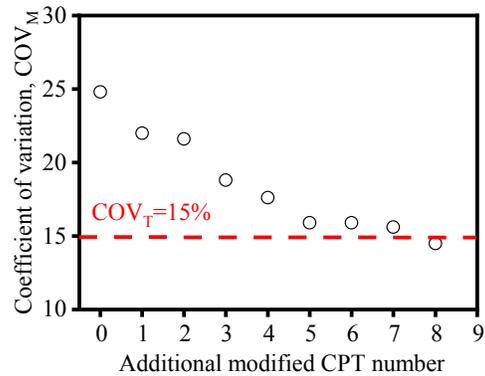

Figure 5. Variation of COV_M with the number of additional modified CPT measurements

Figure 4. Gas pressure P_g profile interpreted from 19 modified CPT measurements
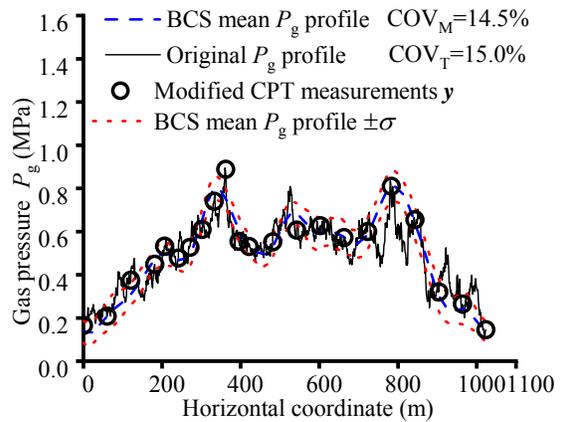

Figure 6. Gas pressure P_g profile interpreted from 26 modified CPT measurements

4. Conclusions

This paper presented a smart sampling strategy for planning of site investigation for subsurface gas pressure. The smart sampling strategy is goal-oriented and progressive that gradually increases the number of modified CPTs until the reliability of interpretation results achieves the target level. After a target reliability level is determined, the minimal number of modified CPTs and their corresponding optimal locations are automatically obtained using the smart sampling strategy. Simulated non-Gaussian and non-stationary gas pressure data are used to illustrate and validate the smart sampling strategy. The results show the method is rational and effective.

5. Acknowledgement

References

CEN (European Committee for Standardization). 2006. *Geotechnical design—part 2: ground investigation and testing (Eurocode 7-2)*. Brussels, Belgium: CEN.

Cover, T. M., and Thomas, J. A. 2012. *Elements of information theory*. John Wiley and Sons. Hoboken, New Jersey.

Guan, Z., Cao, Z., Hong, Y and Wang, Y. 2020. Smart Sampling Strategy for Investigating Spatial Distribution of Subsurface Shallow Gas Pressure in Hangzhou Bay Area of China. *Engineering Geology*, accepted.

Li, L., Zhao, Y., and Yu, L. 2009. Exploration for Quaternary shallow biogenic gas by sealed core drilling and modified CPT. *Coal Geology and Exploration*, 37(5), 72-76, (in Chinese).

Li, Y., and Lin, C. 2010. Exploration methods for late Quaternary shallow biogenic gas reservoirs in the Hangzhou Bay area, eastern China. *AAPG bulletin*. 94(11), 1741-1759.

Look BG. 2007. *Handbook of geotechnical investigation and design tables*. Balkema Taylor & Francis Group, London.

MacKay, D.J., 2003. Information theory, inference and learning algorithms. Cambridge university press, USA.

Montoya-Noguera, S., Zhao, T., Hu, Y., Wang, Y., and Phoon, K. K. 2019. Simulation of non-stationary non-Gaussian random fields from sparse measurements using Bayesian compressive sampling and Karhunen-Loève expansion. *Structural Safety*. 79, 66-79.

Rix, G.J., Wainaina, N., Ebrahimi, A., Bachus, R.C., Limas, M., Sancio, R., Fait, B., and Mayne, P.W. 2018. Manual on Subsurface Investigations (No. NCHRP Project 21-10).

Vanmarcke, E. 1977. Probabilistic modeling of soil profiles. *Journal of Geotechnical Engineering*, 103(11): 1227-1246.

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

Wang, Y., Kong, L., Wang, Y., Wang, M., and Cai, K. 2018. Deformation analysis of shallow gas-bearing ground from controlled gas release in Hangzhou bay of china. *International Journal of Geomechanics*. 18(1), 04017122.

Wang, Y., Zhao, T., Hu, Y., and Phoon, K. K. 2019. Simulation of random fields with trend from sparse measurements without detrending. *Journal of Engineering Mechanics*. 145(2), 04018130.

Zhao, T., and Wang, Y. 2019. Determination of efficient sampling locations in geotechnical site characterization using information entropy and Bayesian compressive sampling. *Canadian Geotechnical Journal*. 56(11), 1622-1637.