Canadian Geotechnical Journal

Selection of Standard Penetration Test Number for Geotechnical Investigation of a Vertical Cross-section Considering Spatial Variability and Correlation in Soil Properties

Journal:	<i>Canadian Geotechnical Journal</i>
Manuscript ID	cgj-2020-0440.R2
Manuscript Type:	Article
Date Submitted by the Author:	23-Dec-2020
Complete List of Authors:	Guan, Zheng; City University of Hong Kong Wang, Yu; City University of Hong Kong, Dept of Civil and Architectural Engineering
Keyword:	Site investigation, Bayesian compressive sampling, Random field, Standard Penetration test, Parametric study
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Selection of Standard Penetration Test Number for Geotechnical Investigation of a Vertical Cross-section Considering Spatial Variability and Correlation in Soil Properties

Zheng Guan¹ and Yu Wang²

¹ PhD student, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. (Tel): 852-3442-6492 (Fax): 852-3442-0427 (Email): zhengguan3-c@my.cityu.edu.hk

² Professor, Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. (Tel): 852-3442-7605 (Fax): 852-3442-0427 (Email): yuwang@cityu.edu.hk (Corresponding author)

ABSTRACT

Standard Penetration test (SPT) is a widely used in-situ test for characterizing variation of subsurface soil properties, and results of site investigation are usually simplified as a 2D vertical cross-section for subsequent geotechnical design and construction. Current geotechnical design codes and guidelines only provide general recommendations for selection of an appropriate number of in-situ tests (e.g., SPT) (e.g., the greater variability of subsurface conditions, the larger number of SPTs required to obtain sufficient underground information). No quantitative or rational method is available for selecting the appropriate number of SPTs considering spatial variability and correlation in subsurface conditions. A comprehensive parametric study is carried out in this study to investigate the influence of spatial variability in subsurface conditions on the minimum SPT number needed for satisfying an accuracy requirement of site investigation. Random field is adopted to model spatial variation and correlation in soil properties in typical site conditions, and 2D Bayesian compressive sampling is used to interpolate sparse SPT data. Based on the parametric study results, a statistical chart is developed for geotechnical engineers to conveniently select appropriate number of SPTs in a vertical cross-section. Real SPT data from New Zealand are used to illustrate and validate the proposed method.

Keywords: Site investigation; Bayesian compressive sampling; Random field; Standard Penetration test; Parametric study

36 Introduction

37 A primary purpose of geotechnical site investigation is to collect subsurface data for
38 determining engineering properties of subsurface soils and delineating underground
39 stratigraphy (e.g., Clayton 2001; Cao et al. 2016; Rix et al. 2018). In geotechnical site
40 investigation, Standard Penetration test (SPT) is a widely used in-situ test for characterizing
41 horizontal and vertical variation of soil properties, and the results of site investigation are
42 usually simplified as a 2D vertical cross-section for subsequent planning, design, and
43 construction of geotechnical engineering projects (e.g., Skempton 1986; Clayton 1995; Wang
44 et al. 2019a). It is widely recognized that the accuracy of site investigation results is greatly
45 affected by the number of investigation points (e.g., SPT) (e.g., AASHTO 2018). However, no
46 quantitative guideline is available for selecting an appropriate number of SPTs needed for
47 satisfying an accuracy requirement of site investigation. Current geotechnical design codes
48 only provide general recommendations for selecting the appropriate number of in-situ tests
49 (e.g., SPT) (e.g., CEN 2007; New York State DOT 2013; FHWA 2017; Michigan DOT 2019).
50 For example, Craig (2004) states that “*The greater the degree of variability of the ground*
51 *conditions, the greater the number of boreholes or pits required.*”, but it is unclear what test
52 number exactly shall be for “the greater the degree of variability”.

53 In geotechnical engineering practice, the importance of SPT number has long been
54 recognized qualitatively by geotechnical engineers, but the number of SPTs is selected mainly
55 based on engineering judgement (e.g., Christian 2004; Wang et al. 2015). Overreliance on
56 engineering judgement might lead to inconsistent or biased site investigation results. Note that
57 risk associated with the underground conditions has been recognized as the greatest risk for
58 most civil engineering projects, and such risk is directly affected by the accuracy of
59 investigation results (e.g., Ground Board 1991; Clayton 2001). When an inadequate number of
60 SPTs is performed, the site investigation result might be inaccurate, or even misleading, leading

to significant increase in the likelihood of change orders, claims, cost and schedule overruns during construction of the projects (e.g., Boeckmann and Loehr 2016). Prezzi et al. (2011) investigated more than 300 construction contracts in USA and concluded that the main reason for geotechnical change orders was insufficient geotechnical investigation. Although the accuracy of site investigation results can be improved by simply increasing the SPT number, the resources (e.g., time, money, and manpower) required also significantly increase. Therefore, it is essential to develop a quantitative method for properly selecting the minimum number of SPTs required for satisfying an accuracy requirement of site investigation.

Effects of the number of investigation points on reliability-based design have received considerable attention in recent geotechnical literature (e.g., Jaksa et al. 2005; Cho 2007; Goldsworthy et al. 2007; Ching and Phoon 2012; Fenton et al. 2015; Wang et al. 2019b; Guan and Wang 2019; Crisp et al. 2019). For example, Loehr et al. (2015) concluded that the number of measurements required to limit the possibility of producing under-reliable geotechnical designs greatly depends on the variability of sites. Ching and Phoon (2012) investigated the influence of the amount of site investigation information on reliability-based design of geotechnical structures. However, no rational method is available for quantitatively determining the minimum number of SPTs for geotechnical investigation of a vertical cross-section.

This paper aims to address the issue described above and proposes a rational method for quantitatively determining the minimum number of SPTs required for satisfying an accuracy requirement of site investigation in a 2D vertical cross-section. The proposed method is developed based on a comprehensive parametric study that covers typical site conditions in geotechnical practice. In the parametric study, spatial variability of soil properties within a vertical cross-section is modelled by a 2D random field, and interpolation of sparse SPT data along both vertical and horizontal direction is performed using 2D Bayesian compressive

sampling (BCS). This paper is organized as follows. First, framework of the parametric study is provided, followed by 2D random field modeling to simulate a wide variety of typical site conditions. Then, 2D BCS for interpolation of sparse SPT data is briefly introduced, and analysis of the parametric study result is performed, leading to development of a statistical chart for SPT number selection. Real SPT data are used to illustrate the proposed method.

Framework for parametric study

It has been widely acknowledged that the number of SPTs should be selected to reflect spatial variability in site conditions (e.g., FHWA 2017; AASHTO 2018). Generally speaking, to achieve the same level of accuracy in site investigation results, the required number of SPTs for a highly variable site condition shall be much larger than that for a uniform site condition. Then, in order to select an appropriate number of SPTs for a given site, influence of spatial variability in site conditions on the required SPT number for achieving a target accuracy should be quantified. In this study, an extensive parametric study is performed to quantify the relationship between accuracy of SPT interpretation results and SPT numbers considering spatial variability and correlation at site conditions typically encountered in engineering practice. As shown in Fig. 1, the parametric study mainly involves seven steps, which are described below.

To evaluate accuracy of the interpretation results from a given number of SPTs, the true site condition should be completely known, but is often not available in engineering practice. On the other hand, spatial variability of soil properties can be modelled explicitly using random fields (e.g., Vanmarcke 1977), and SPT data are often used to directly estimate various soil properties (e.g., shear wave velocity, cohesion, and friction angle) in engineering practice (e.g., Mayne 2006). Therefore, in Step 1 of this study, a wide variety of typical site conditions are modelled using 2D SPT data random fields. In addition, note that accuracy of the results

interpreted from the same number of SPTs might vary significantly for different random field samples (i.e., 2D high-resolution vertical cross-sections of SPT data) generated from a given random field (i.e., a typical site condition). To consider variability in the accuracy obtained from different random field samples, many (e.g., 100) random field samples are used for each random field. In other words, many (e.g., 100) possible 2D high-resolution vertical cross-sections of SPT data are generated for a typical site condition in Step 2.

In Steps 3-6, accuracies of the interpretation results from different numbers of SPTs are first evaluated for a typical site condition (i.e., a given SPT data random field). For a given random field sample (i.e., a given high-resolution cross-section of SPT data), measurement data from different numbers of boreholes are extracted from the given cross-section of SPT data in Step 3 to mimic different numbers, M_{x_1} (e.g., $M_{x_1}=5, 8, \dots, 30$), of boreholes drilled along a horizontal direction (e.g., x_1 direction) where SPTs are performed. Note that SPT is generally performed with a depth interval, $d_v=1.0\text{m}-5.0\text{m}$, and an equal spacing along the horizontal direction is usually adopted in engineering practice (e.g., Mayne et al. 2002; Zhao and Wang 2019; Guan et al. 2020). In Step 4, the extracted measurement data for each borehole number with different depth intervals (e.g., $d_v=1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}, 5\text{m}$) are used to interpolate cross-section of SPT data, which has the same dimension with the given cross-section of SPT data using an interpolation method (e.g., 2D Bayesian compressive sampling (BCS)). 2D BCS is adopted in this study because it is a non-parametric and data-driven method which can interpolate a high-resolution of SPT data directly from sparse measurement data (e.g., Wang et al. 2017; Montoya-Noguera et al. 2019). As the genuine cross-section of SPT data is completely known, the accuracy of a cross-section of SPT data interpreted from different combinations of M_{x_1} and d_v can be quantified by measuring a relative error, RE, between the cross-section of SPT data interpreted from sparse SPT measurement data and the true one in

Step 5. Then, Steps 3-5 are repeated many (e.g., 100) times, using a different SPT cross-section generated from a given 2D random field each time, leading to 100 REs at different combinations of M_{x_1} and d_v . In Step 6, statistics of relative errors (e.g., mean relative error, μ_{RE}) are calculated from each of 100 REs, and μ_{RE} is used as the accuracy of interpretation result from a given combination of M_{x_1} and d_v for a typical site condition.

To explore the effect of spatial variability for a wide variety of typical site conditions on the relationship between the SPT number and level of accuracy, Steps 2-6 are repeated many times for various random fields simulated in Step 1, leading to many relationships among M_{x_1} , d_v and μ_{RE} for various site conditions. Then, statistical analysis is carried out to quantify the influence of spatial variability in site condition on the relationship among M_{x_1} , d_v and μ_{RE} . Based on the parametric study result, a statistical chart is further established for facilitating geotechnical engineers to select the minimum number of SPTs needed for satisfying an accuracy requirement of site investigation. In the parametric study described above, generation of 2D SPT data random field and 2D BCS are two key modules, which are described in Section 3 and 4, respectively.

Random field simulation of typical site conditions

It is well-recognized that the soil properties are spatially correlated and varying along both horizontal and vertical direction owing to the complex geological processes that soils have gone through during their formation (e.g., Fenton and Griffiths 2008; Wang et al. 2016). In geotechnical engineering, spatial variability of soil properties may be modelled explicitly using random fields (e.g., Vanmarcke 1977, 2010; Fenton 1999; Stuedlein et al. 2012). Therefore, in this study, a wide variety of typical site conditions commonly encountered in engineering

practice are modelled using 2D SPT data random fields. For example, consider the normalized SPT value $(N_1)_{60}$, which is usually expressed as (e.g., Boulanger and Idriss 2014):

$$(1) \quad (N_1)_{60} = C_N N_{60}$$

where N_{60} is the corrected SPT N value; $C_N = (\sigma_{atm} / \sigma_{v0}')^{0.5} \leq 2$, $\sigma_{atm} = 100$ kPa, σ_{v0}' = effective vertical overburden stress. $(N_1)_{60}$ has been widely used to estimate various soil properties such as friction angle, preconsolidation stress, and cyclic resistance ratio (e.g., Mayne 2006).

The $(N_1)_{60}$ data is often taken to be normally distributed since the normal distribution is commonly used for modelling probability distribution of soil properties (e.g., Fenton and Griffiths 2008; Wang and Xu 2015). To model a 2D Gaussian random field for $(N_1)_{60}$ data, three parameters are needed, including mean $\mu_{(N_1)_{60}}$, standard deviation $\sigma_{(N_1)_{60}}$, and spatial scale of fluctuation (i.e., vertical scale of fluctuation, λ_v , and horizontal scale of fluctuation, λ_h). In addition, variability of $(N_1)_{60}$ data may be conveniently described by the coefficient of variation, which is expressed as (e.g., Vanmarcke 1977):

$$(2) \quad COV_{(N_1)_{60}} = \frac{\sigma_{(N_1)_{60}}}{\mu_{(N_1)_{60}}}$$

The spatial scale of fluctuation describes the distance within which $(N_1)_{60}$ data tend to be significantly correlated in a 2D cross-section (e.g., Vanmarcke 1977). An anisotropic single exponential correlation function (SECF) is adopted in this study, and it is expressed as below (e.g., Vanmarcke 2010):

$$\rho(\Delta x_1, \Delta x_2) = \exp \left(-2 \sqrt{\frac{(\Delta x_1)^2}{\lambda_v^2} + \frac{(\Delta x_2)^2}{\lambda_h^2}} \right)$$

183

184 where Δx_1 and Δx_2 are the absolute distance between any two $(N_1)_{60}$ data points along the
 185 horizontal and vertical direction, respectively. For a given $(N_1)_{60}$ random field, the spatial
 186 variability of a typical site condition is explicitly specified by $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h .

187 To perform a comprehensive parametric study, a wide variety of site conditions
 188 frequently encountered in engineering practice is simulated. Numerous studies have been
 189 carried out to evaluate spatial variability of SPT data, and the typical values of $\mu_{(N_1)_{60}}$,
 190 $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h have been reported in literature, as summarized in Table 1 (e.g., Phoon
 191 and Kulhawy 1999; Zhang and Chen 2012; Oguz et al. 2019; Balasubramani and Dodagoudar
 192 2020). In general, $\mu_{(N_1)_{60}}$ is ranging from 0.1 to 100, while $\text{COV}_{(N_1)_{60}}$ is ranging from 1% to
 193 100%; λ_v and λ_h are ranging from 0.1m to 10m and 1m to 100m. Using these typical values,
 194 a wide variety of site conditions can be simulated using 2D random fields. Table 2 summarizes
 195 the random field parameters used in the parametric study, and it covers typical ranges reported
 196 in literature. In addition, different sizes of cross-sections (e.g., Depth $\times$ Width, $D \times W = 9\text{m} \times$
 197 50m , $9\text{m} \times 100\text{m}$, $15\text{m} \times 250\text{m}$, ..., $90\text{m} \times 1000\text{m}$) are also used in random field simulation.
 198 To consider many possible 2D vertical cross-sections of $(N_1)_{60}$ data within a given site
 199 condition, 100 random field samples (i.e., 2D high-resolution vertical cross-sections of $(N_1)_{60}$
 200 data) are generated for each random field using circulant embedding algorithm (e.g., Dietrich
 201 and Newsam 1993).

202 2D Bayesian compressive sampling (BCS) is another key module in the proposed
 203 framework for interpolating sparse SPT measurement data along vertical and horizontal
 204 directions simultaneously (e.g., Zhao et al. 2018), as described in the following section.

2D Bayesian compressive sampling (BCS)

2D BCS is a novel sampling method in signal or image processing which integrates Bayesian formulation with compressive sampling (CS) in a two-dimensional space (e.g., Ji et al. 2008, 2009; Huang et al. 2014; Wang and Zhao 2016, 2017). In geotechnical engineering, 2D BCS has been used for interpretation of spatial varying 2D geo-data from sparse measurement (e.g., Zhao et al. 2018). Let a $k_{x_2} \times k_{x_1}$ matrix $\mathbf{F}$ represents the spatial variation of SPT data (i.e., 2D high-resolution vertical cross-section of $(N_1)_{60}$ data) along depth (i.e., x_2 direction) and horizontal direction (i.e., x_1 direction) with a total number of elements $K = k_{x_2} \times k_{x_1}$, and a matrix $\hat{\mathbf{F}}$ which have the same dimension as $\mathbf{F}$ denotes the cross-section of $(N_1)_{60}$ data interpreted from M_{x_1} SPT boreholes conducted in $\mathbf{F}$. SPT measurement data are expressed by an $M_{x_2} \times M_{x_1}$ matrix $\mathbf{Y}$ ($M_{x_2} < k_{x_2}$, $M_{x_1} < k_{x_1}$) with a total number of elements $M = M_{x_2} \times M_{x_1}$. Note that M_{x_2} represents the number of measurement data in each borehole along depth which can be calculated directly by $D/d_v + 1$. In the context of CS, $\mathbf{F}$ may be represented by a weighted sum of K orthonormal 2D basis functions, $\mathbf{B}_t^{2D}$, $t=1, 2, \dots, K$, which is expressed as (e.g., Candès et al. 2006):

$$(4) \quad \mathbf{F} = \sum_{t=1}^K \mathbf{B}_t^{2D} \omega_t^{2D}$$

where ω_t^{2D} represents the weight coefficient corresponding to $\mathbf{B}_t^{2D}$. Note that the K in Eq. (4) is the total number of data points within a cross-section. Because the spatially varying and auto-

correlated signals (e.g., 2D high-resolution vertical cross-section of $(N_1)_{60}$ data, $\mathbf{F}$) are compressible, most of ω_t^{2D} are close to zero (e.g., Tipping 2001; Zhao et al. 2018).

Let a K -dimensional vector $\boldsymbol{\omega}^{2D} = [\omega_1^{2D}, \omega_2^{2D}, \dots, \omega_K^{2D}]$ represents K weight coefficients for different basis functions. Under a CS framework, most components of $\boldsymbol{\omega}^{2D}$ have a negligible amplitude (e.g., Tipping 2001). Therefore, when elements in $\boldsymbol{\omega}^{2D}$ with relatively large magnitude are estimated from sparse SPT measurement data $\mathbf{Y}$, the high-resolution cross-section of $(N_1)_{60}$ data can be properly reconstructed. The mathematical relationship between $\mathbf{Y}$ and ω_t^{2D} can be expressed as:

$$(5) \quad \mathbf{Y} = \boldsymbol{\Psi}_{x_2} \mathbf{F} \boldsymbol{\Psi}_{x_1} = \sum_{t=1}^K \boldsymbol{\Psi}_{x_2} \mathbf{B}_t^{2D} \boldsymbol{\Psi}_{x_1} \omega_t^{2D} = \sum_{t=1}^K \mathbf{A}_t^{2D} \omega_t^{2D}$$

where a $M_{x_2} \times k_{x_2}$ matrix $\boldsymbol{\Psi}_{x_2}$ and a $k_{x_1} \times M_{x_1}$ matrix $\boldsymbol{\Psi}_{x_1}$ indicate the locations of SPT measurements $\mathbf{Y}$ in $\mathbf{F}$ along depth and horizontal direction, respectively, and both can be constructed from an identity matrix. For example, all elements in the first row of $\boldsymbol{\Psi}_{x_2}$ are equal to zero except the element at the location of the first SPT measurement along depth which is equal to one. Using Eq. (5), $\boldsymbol{\omega}^{2D}$ can be approximately estimated as $\hat{\boldsymbol{\omega}}^{2D} = [\hat{\omega}_1^{2D}, \hat{\omega}_2^{2D}, \dots, \hat{\omega}_K^{2D}]$ under a Bayesian framework (e.g., Ji et al. 2008; Fang et al. 2012). Zhao et al. (2018) have shown that $\hat{\boldsymbol{\omega}}^{2D}$ follows a multivariate Student's t distribution. The mean $\hat{\boldsymbol{\omega}}^{2D}$, denoted as $\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{2D}}$, is expressed as (e.g., Zhao et al. 2018):

$$(6) \quad \boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{2D}} = \mathbf{H} \mathbf{V}_{tr} = (\mathbf{J} + \mathbf{D})^{-1} \mathbf{V}_{tr}$$

where $\mathbf{D}$ is an $K \times K$ diagonal matrix which can be estimated by maximizing the likelihood of $\mathbf{Y}$ (e.g., Ji et al. 2009; Wang and Zhao 2017); $\mathbf{J}$ is a $K \times K$ matrix with elements $J_{t,s} = \text{tr}[\mathbf{A}_t^{2D}(\mathbf{A}_s^{2D})^T]$ ($t, s = 1, 2, \dots, K$); a K -dimensional column vector $\mathbf{V}_{tr} = \{\text{tr}[\mathbf{Y}(\mathbf{A}_1^{2D})^T], \text{tr}[\mathbf{Y}(\mathbf{A}_2^{2D})^T], \dots, \text{tr}[\mathbf{Y}(\mathbf{A}_K^{2D})^T]\}^T$; “tr” indicates the trace of a matrix in linear algebra. Using Eq. (6), the mean of $\hat{\mathbf{F}}$, denoted as $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$, can be obtained as (e.g., Zhao et al. 2018):

$$(7) \quad \boldsymbol{\mu}_{\hat{\mathbf{F}}} = \sum_{t=1}^K \mathbf{B}_t^{2D} \mu_{\hat{\omega}_t^{2D}}$$

where $\mu_{\hat{\omega}_t^{2D}}$ represents the mean value of t -th weight coefficient. In this study, $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$ is considered as the best estimated 2D high-resolution vertical cross-section of $(N_1)_{60}$ data from sparse SPT measurements. Note that previous studies (e.g., Wang et al. 2017, Hu et al. 2020) have shown that BCS method outperforms kriging when measurement data from a specific site are sparse and limited. When plenty of measurement data points are available, BCS and kriging provide similar results.

After the cross-section of $(N_1)_{60}$ data is interpreted from sparse SPT measurements, the accuracy of the interpretation results may be quantified by calculating the relative error, RE between $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$ and $\mathbf{F}$ which is expressed as:

$$(8) \quad \text{Relative error: RE} = \frac{\|\boldsymbol{\mu}_{\hat{\mathbf{F}}} - \mathbf{F}\|}{\|\mathbf{F}\|} = \frac{\sqrt{\sum_{i=1}^{k_{x_2}} \sum_{j=1}^{k_{x_1}} (\mu_{\hat{f}_{i,j}} - f_{i,j})^2}}{\sqrt{\sum_{i=1}^{k_{x_2}} \sum_{j=1}^{k_{x_1}} (f_{i,j})^2}}$$

where $\mu_{\hat{f}_{i,j}}$ and $f_{i,j}$ represent the entry in the i -th row and j -th column of the matrix $\mu_{\hat{f}}$ and F , respectively; $\|\bullet\|$ represents the Frobenius norm operation in linear algebra, which is defined as the square root of the sum of squares of all matrix elements. Note that the definition of the relative error adopted in this study is widely used in signal and image processing for quantifying the accuracy of reconstructed signal and image (e.g., Ji et al. 2009; Kuzin et al. 2015; Marques et al. 2018).

Using simulated typical site conditions and 2D BCS, a comprehensive parametric study is carried out to quantify the influence of spatial variability on the relationship among M_{x_1} , d_v and μ_{RE} . Details of the parametric study are described in the following sections.

278

279 **Parametric study using simulated SPT data**

280 *Steps 1-2: Simulate typical site conditions using random fields*

In Step 1, typical site conditions are modelled by 2D random fields. The total number of combinations is $1 \times 6 \times 7 \times 6 = 252$ among the mean $\mu = 10$, coefficient of variation, $COV = 5\%$, 10% , 20% , 30% , 40% , 50% , vertical scale of fluctuation, $\lambda_v = 0.5\text{m}$, 1m , 2m , 4m , 5m , 6m , 10m and horizontal scale of fluctuation $\lambda_h = 5\text{m}$, 10m , 20m , 40m , 50m , 100m . The total number of combinations is $1 \times 3 \times 5 \times 3 = 45$ among the mean $\mu = 100$, coefficient of variation, $COV = 10\%$, 80% , 100% , vertical scale of fluctuation, $\lambda_v = 0.5\text{m}$, 2m , 5m , 8m , 10m and horizontal scale of fluctuation $\lambda_h = 5\text{m}$, 15m , 75m . The total number of combinations is $1 \times 2 \times 5 \times 5 = 50$ among the mean $\mu = 0.1$, coefficient of variation, $COV = 10\%$, 50% , vertical scale of fluctuation, $\lambda_v = 0.5\text{m}$, 2m , 5m , 8m , 10m and horizontal scale of fluctuation $\lambda_h = 1\text{m}$, 10m , 30m , 75m , 100m . Therefore, a total number of $252 + 45 + 50 = 347$ (N_1)₆₀ random fields is simulated to cover various typical sites in engineering practice. For each random field, 100 2D

high-resolution vertical cross-sections of $(N_1)_{60}$ data are generated in Step 2. Consider, for example, a site condition that consists of medium sand with medium variability and is modelled by a 2D Gaussian $(N_1)_{60}$ random field with $\mu_{(N_1)_{60}} = 10$, $COV_{(N_1)_{60}} = 30\%$, $\lambda_v = 5\text{ m}$ and $\lambda_h = 75\text{ m}$. In Step 2, 100 high-resolution cross-sections of $(N_1)_{60}$ data are generated for this simulated random field. For example, a cross-section of $(N_1)_{60}$ data with $D \times W = 15\text{ m} \times 250\text{ m}$ and a resolution of 0.1m and 0.5m along the vertical and horizontal direction, respectively. In total, the generated cross-sections of $(N_1)_{60}$ data have about $150 \times 500 = 75000$ data points. One hundred random field samples (i.e., $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{100}$) are generated for this simulated $(N_1)_{60}$ random field that represents medium sand. As an illustration, four of them (i.e., $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$) are shown by contour maps in Fig. 2.

Steps 3-6: Evaluate accuracy of the results interpreted from different numbers of SPTs for a given site condition

In Steps 3-6, accuracies of site investigation result from different numbers of SPTs are first evaluated for a typical site condition (i.e., a given $(N_1)_{60}$ random field). To mimic different numbers $M_{x_1} = 5, 8, 10, 12, 15, 18, 20, 22, 25, 28, 30$ of boreholes with SPTs performed using different depth interval $d_v = 1\text{ m}, 1.5\text{ m}, 2.5\text{ m}, 3\text{ m}, 5\text{ m}$ for a given high-resolution cross-section of $(N_1)_{60}$ data, the measurement data corresponding to the different number of SPTs are extracted from the given cross-section of $(N_1)_{60}$ data in Step 3. For example, consider $M_{x_1} = 5$ boreholes with SPTs performed with a depth interval $d_v = 1.5\text{ m}$ for the cross-section of $(N_1)_{60}$ data, $\mathbf{F}_1$, as shown in Fig. 2(a). The number, M_{x_2} , of SPTs performed in each borehole is calculated as $D/d_v + 1 = (15\text{ m}/1.5\text{ m}) + 1 = 11$. To mimic $M_{x_1} = 5$ boreholes in $\mathbf{F}_1$, a total number, $M = M_{x_2} \times M_{x_1} = 11 \times 5 = 55$, of SPT $(N_1)_{60}$ data points are extracted from $\mathbf{F}_1$ in Step 3.

In Step 4, the extracted SPT $(N_1)_{60}$ data points for each borehole number are used as input to the 2D BCS for interpolating a cross-section of SPT data, which has the same resolution with the given original cross-sections of $(N_1)_{60}$. Then, the accuracy (i.e., relative error, RE) of SPT data cross-sections interpreted from different M_{x_1} values are quantified using Eq. (8) in Step 5. To illustrate Steps 4-5, different borehole numbers $M_{x_1}=5, 8, 10, 12, 15, 18, 20, 22, 25, 28, 30$ and different depth intervals $d_v=1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}$ are repeatedly performed for the $\mathbf{F}_1$ shown in Fig. 2(a). The cross-sections of SPT data with a dimension of 150×500 interpreted from $M_{x_1}=5, 10, 20, 30$ boreholes with a depth interval $d_v=1.5\text{m}$ are shown by contour maps in Fig. 3. In Fig. 3, SPT measurement data are denoted by open circles. It is found from Fig. 3 that the interpreted cross-sections become increasingly similar to the true one (see $\mathbf{F}_1$ in Fig. 2(a)), both locally and globally, as the number of SPTs increases. Relative errors of the cross-sections interpreted from the total number of measurement data $M = M_{x_2} \times M_{x_1} = 11 \times 5, 11 \times 10, 11 \times 20, 11 \times 30 = 55, 110, 220, 330$ are calculated as 23.2%, 17.2%, 15.0% and 14.0%, respectively. Moreover, the $(N_1)_{60}$ profiles along horizontal direction at the depth of 6m interpreted from $M = 55, 110, 220, 330$ are shown in Fig. 4. It is observed that the $(N_1)_{60}$ profile gradually converges to the true $(N_1)_{60}$ profile as the SPT number increases. The accuracies of cross-sections of $(N_1)_{60}$ interpreted from different borehole numbers with different SPT depth intervals are summarized in Fig. 5. Note that the SPT number $M_{x_2}=16, 11, 7, 6$ along each borehole are calculated for $d_v=1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}$. Fig. 5 shows that the relative error generally decreases as M_{x_1} or M_{x_2} increases.

As mentioned in Section 2, the accuracy of interpretation results from a given number of SPTs might vary significantly for different random field samples. To consider variation among different random field samples generated from a given site condition, Steps 3-5 are repeated 100 times using different cross-sections of $(N_1)_{60}$ generated from a given 2D random

field, leading to 100 REs at different combinations of M_{x_1} and M_{x_2} , as shown in Fig. 6. In Step 6, the mean relative error μ_{RE} is calculated from 100 REs at different combinations of $M_{x_2}=16, 11, 7, 6$ and $M_{x_1}=5, 8, 10, 12, 15, 18, 20, 22, 25, 28, 30$. The μ_{RE} corresponding to different combinations of M_{x_1} and M_{x_2} are summarized in Fig. 7. The results show that the accuracy of interpretation results increases as the borehole number increases or the depth interval decreases. A quantitative relationship among M_{x_1} , d_v and μ_{RE} has been established in Fig. 7 for the given $(N_1)_{60}$ random field (i.e., a site condition with medium dense sand).

Step 7: Investigate the influence of spatial variability and correlation in soil properties

To explore the effect of spatial variability in soil properties for a wide variety of typical site conditions on the relationship between the number of SPTs and level of accuracy, the accuracies of interpretation result from different numbers of boreholes with SPTs performed with different depth intervals should be obtained for various site conditions. Therefore, Steps 2-6 are repeated for each of the 2D $(N_1)_{60}$ random fields using eleven different borehole numbers ($M_{x_1}=5, 8, 10, 12, 15, 18, 20, 22, 25, 28, 30$) and five different depth intervals ($d_v=1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}, 5\text{m}$). For the 252 random fields with the mean $\mu=10$ and $D \times W = 15\text{m} \times 250\text{m}$, the total number of combinations is $252 \times 2 \times 6 = 3024$ among $d_v = 1.5, 3.0$ and $M_{x_1} = 5, 10, 15, 20, 25, 30$. For the remaining 95 random fields together with other values shown in Table 2 (i.e., $d_v = 1.0\text{m}, 2.5\text{m}$ and 5m , $M_{x_1} = 8, 12, 18, 22, 28$, depth of vertical cross-section = $9\text{m}, 30\text{m}, 45\text{m}, 90\text{m}$, width of vertical cross-section = $50\text{m}, 100\text{m}, 1000\text{m}$), 1064 combinations are generated. In total, 4088 μ_{RE} are obtained at different combinations of d_v and M_{x_1} for different random fields.

In total, 4088 μ_{RE} are obtained at different combinations of d_v and M_{x_1} . As mentioned in Section 2, the spatial variability in soil properties is represented explicitly by three random field parameters: coefficient of variation, $COV_{(N_1)_{60}}$, vertical scale of fluctuation, λ_v , and horizontal scale of fluctuation, λ_h . As summarized in Table 2, eight different coefficient of variation (i.e., $COV_{(N_1)_{60}} = 5\%, 10\%, 20\%, 30\%, 40\%, 50\%, 80\%, 100\%$), eight different vertical scale of fluctuations (i.e., $\lambda_v = 0.5m, 1m, 2m, 4m, 5m, 6m, 8m, 10m$), and ten different horizontal scale of fluctuations (i.e., $\lambda_h = 1m, 5m, 10m, 15m, 20m, 30m, 40m, 50m, 75m, 100m$) are considered in the parametric study. Details of the parametric study results are provided in the following section.

Parametric study results

Results of the parametric study (i.e., 4088 μ_{RE} at different combinations of d_v and M_{x_1}) are plotted in a dimensionless plot in Fig. 8. Note that a given M_{x_1} value represents a borehole spacings along horizontal direction for a given width of a cross-section. For consistency between horizontal and vertical directions, borehole spacing, d_h , rather than M_{x_1} , is used in Fig. 8, which can be readily calculated by the width of site divided by $(M_{x_1} - 1)$. The horizontal axis of Fig. 8 is the normalized borehole spacing, d_N , which is defined as:

$$(9) \quad d_N = \frac{d_h}{\lambda_h} \times \frac{d_v}{\lambda_v}$$

where d_h is borehole spacing along horizontal direction; d_v represents the vertical interval of SPT data along depth; λ_v and λ_h are vertical and horizontal scale of fluctuation, respectively.

The vertical axis of Fig. 8 is the normalized mean relative error which is defined as mean relative error μ_{RE} normalized by the coefficient of variation, $COV_{(N_1)_{60}}$. It can be seen from Fig. 8 that 4088 data pairs of d_N and $\mu_{RE} / COV_{(N_1)_{60}}$ exhibit an obvious trend between d_N and $\mu_{RE} / COV_{(N_1)_{60}}$. The value of $\mu_{RE} / COV_{(N_1)_{60}}$ increases rapidly from about 0.4 to about 1.3 as d_N increases from about 0 to about 2.5, and it converges to an asymptotic value of about 1.3 when d_N is larger than 2.5. The effect of spatial variability and correlation in soil properties (i.e., $COV_{(N_1)_{60}}$, λ_v and λ_h) on the relationship among μ_{RE} , d_v and d_h is also quantified in Fig. 9, as discussed in the following subsections.

Effect of coefficient of variation

Coefficient of variation, $COV_{(N_1)_{60}}$, is a statistical measure of the dispersion of $(N_1)_{60}$ data points around the mean within a site, and it is a useful indicator of the variability within a typical site condition (e.g., Fenton and Griffiths 2008). General speaking, the larger value of $COV_{(N_1)_{60}}$, the higher variability within a site. It can be seen from Fig. 9 that the coefficient of variation, $COV_{(N_1)_{60}}$ of a site has a great effect on the relationship among μ_{RE} , d_v and d_h . For a given accuracy of interpretation result (i.e., μ_{RE}) and spatial scale of fluctuation (i.e., λ_v and λ_h), the borehole spacing d_v or depth interval d_h decreases as the $COV_{(N_1)_{60}}$ increases. It is reasonable because a large value of $COV_{(N_1)_{60}}$ indicates that $(N_1)_{60}$ data within a cross-section have a large variability over the mean value, and thus a small borehole spacing and a small depth interval are required to capture the variation of $(N_1)_{60}$ data along vertical and horizontal direction. The result is also consistent with the geotechnical common sense that much closer investigation point spacing should be adopted at sites where underground condition is suspected to be more variable.

Effect of horizontal scale of fluctuation

The horizontal scale of fluctuation reflects spatial correlation of $(N_1)_{60}$ data along horizontal direction in a random field (e.g., Vanmarcke 1977). It represents the horizontal distance within which $(N_1)_{60}$ data are significantly correlated. $(N_1)_{60}$ data separated by a distance larger than λ_h might show little correlation. It is observed from Fig. 9 that horizontal scale of fluctuation also has a significantly effect on the relationship among μ_{RE} , d_v and d_h . For given μ_{RE} , λ_v and $COV_{(N_1)_{60}}$, d_v or d_h increases as the λ_h increases. When horizontal scale of fluctuation increases, variation of $(N_1)_{60}$ data becomes increasingly correlated along horizontal direction. For such case, a relatively small number of SPT measurement may be sufficient to produce a reliable interpretation result. On the contrary, when λ_h is small, the $(N_1)_{60}$ data become increasingly uncorrelated or highly fluctuated, and the $(N_1)_{60}$ random field along horizontal direction tends to be increasingly ragged. Therefore, an increasing number of SPT measurements is needed to properly characterize the cross-section of $(N_1)_{60}$ data.

Effect of vertical scale of fluctuation

The vertical scale of fluctuation describes spatial correlation of $(N_1)_{60}$ data along depth, and it represents a vertical distance within which $(N_1)_{60}$ data are significantly correlated along depth (e.g., Fenton and Griffiths 2008). The influence of vertical scale of fluctuation on the relationship among μ_{RE} , d_v and d_h is similar to that of the horizontal scale of fluctuation. For given μ_{RE} , λ_h and $COV_{(N_1)_{60}}$, d_v or d_h increases as the λ_v increases. It means that the required number of SPTs decreases as the vertical scale of fluctuation increases. Note that $(N_1)_{60}$ data points along depth become increasingly correlated as λ_v increases. For such case, a relatively small number of SPTs is needed to properly characterize the variation of $(N_1)_{60}$ data along

depth. In summary, $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h all have a significant influence on the relationship among μ_{RE} , d_v and d_h . A quantitative relationship between the accuracy of interpretation result (i.e., μ_{RE}) and the SPT number is developed in the next subsection.

Quantitative relationship between accuracy of interpretation result and SPT number

It can be seen from Fig. 8 that the relationship between normalized relative error, $\mu_{\text{RE}} / \text{COV}_{(N_1)_{60}}$ and normalized SPT spacing is non-linear. $\mu_{\text{RE}} / \text{COV}_{(N_1)_{60}}$ increases monotonically as d_N increases, and it approaches to an asymptotic value of about 1.3 when d_N is larger than 2.5. One of the simple models for such an asymptotic pattern is an exponential model (e.g., Ott 1994) which is adopted in this study. For this case, the asymptotic exponential form might be written as:

$$(10) \quad y = a + b \times (1 - e^{-cx})$$

where $x = d_N$, $y = \mu_{\text{RE}} / \text{COV}_{(N_1)_{60}}$, a is the initial value at $x = 0$, and the asymptote is $y = a + b$ when x is larger than 2.5. A regression analysis is then performed for all data pairs shown in Fig. 8. The best-fit curve for the 4088 data pairs is shown in Fig. 8 by the solid line with the equation given below and a coefficient of determination, R^2 , of 0.77:

$$(11) \quad y = 0.53 + 0.64 \times (1 - e^{-1.75x})$$

where $y = \mu_{\text{RE}} / \text{COV}_{(N_1)_{60}}$ and $x = d_N$. The corresponding confidence interval (e.g., best-fit curve $\pm$ standard deviation) is also shown in Fig. 8 by two dotted lines which are expressed as:

454

$$455 \quad (12) \quad y=0.64+0.64 \times (1-e^{-1.75x})$$

$$456 \quad y=0.42+0.64 \times (1-e^{-1.75x})$$

457

458 The quantitative relationship between the SPT spacing and level of accuracy, μ_{RE} , is
 459 obtained from Eqs. (11) and (12). Such a relationship provides a rational and quantitative tool
 460 for geotechnical engineers to select the minimum number of SPTs needed to meet the accuracy
 461 requirement of site investigation, considering spatial variability and correlation within a
 462 specific site. It should be clarified that the proposed method is not intended to address the
 463 situation of SPT spacing along horizontal direction and SPT interval along depth equal to zero,
 464 because such as situation does not exist in engineering practice. To facilitate application of
 465 such a relationship in engineering practice, a statistical chart is further developed, as described
 466 in the next section.

467

468 **Effect of random field distribution types and spatial correlation function** 469 **types**

470 In addition to normal distribution, lognormal distribution is also often used to model
 471 geotechnical properties in literature (e.g., Lacasse and Nadim 1996; Wang et al. 2015).
 472 Furthermore, several theoretical correlation functions have been often used in literature to
 473 model spatial variation of geotechnical properties, such as single exponential correlation
 474 function (SECF), Gaussian correlation function (GCF), and cosine exponential correlation
 475 function (CECF) (e.g., Vanmarcke 2010; Li et al. 2015). For GCF and CECF, the
 476 autocorrelation structures are expressed as below (e.g., Li et al. 2015):

477

$$(13) \quad \text{GCF: } \rho(\Delta x_1, \Delta x_2) = \exp\left(-\pi\left(\frac{\Delta x_1^2}{\lambda_v^2} + \frac{\Delta x_2^2}{\lambda_h^2}\right)\right)$$

$$\text{CECF: } \rho(\Delta x_1, \Delta x_2) = \exp\left(-\left(\frac{\Delta x_1}{\lambda_v} + \frac{\Delta x_2}{\lambda_h}\right)\right) \cos\left(\frac{\Delta x_1}{\lambda_v}\right) \cos\left(\frac{\Delta x_2}{\lambda_h}\right)$$

480

481 where Δx_1 and Δx_2 are the absolute distance between any two $(N_1)_{60}$ data points along the
 482 horizontal and vertical direction, respectively. To explore the effect of random field distribution
 483 types and spatial correlation function types on the relationship between the SPT number and
 484 the accuracy of interpretation results, the parametric study described in the previous sections
 485 are repeated for normally and lognormally distributed random fields with different correlation
 486 function types covering the parameters summarized in Table 2. The best-fit curves for normally
 487 and lognormally distributed random fields with three different correlation function types are
 488 shown in Fig. 10. The corresponding equations of best-fit curves are summarized in Table 3.

489 Fig. 10 shows that the results for normally distributed and lognormally distributed
 490 random fields are similar. This indicates that the distribution types of $(N_1)_{60}$ random fields have
 491 negligible effect on the parametric results. Furthermore, Fig. 10 indicates that the correlation
 492 function types have considerable effects on the parametric results. It is also interesting to note
 493 that the commonly used Gaussian random fields with a single exponential correlation functions
 494 have the largest values of $\mu_{RE}/\text{COV}_{(N_1)_{60}}$ for a given $d_N = (d_h/\lambda_h) \times (d_v/\lambda_v)$. In other words,
 495 if the Gaussian random field with a single exponential correlation function is used, the resulting
 496 error is more likely to be larger than other models.

497

Statistical chart for determination of the minimum SPT number

In this section, a statistical chart for selecting the minimum number of SPTs is developed based on the quantitative relationship between d_N and $\mu_{RE} / \text{COV}_{(N_1)_{60}}$. Note that μ_{RE} in $\mu_{RE} / \text{COV}_{(N_1)_{60}}$ may be replaced by a target relative error, RE_T , which represents an accuracy requirement of site investigation. By switching the vertical and horizontal axis of Fig. 8 and replacing μ_{RE} by RE_T , a statistical chart for selecting the minimum number of SPTs is generated, as shown in Fig. 11.

To use the statistical chart in Fig. 11 for selecting the minimum number of SPTs in engineering practice, accuracy requirement (i.e., RE_T) and anticipated site conditions (i.e., $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h) should be determined first. Accuracy requirement (i.e., RE_T) represents the desired accuracy level of the interpreted cross-section of SPT data, which shall be determined based on the acceptable and tolerable risk thresholds of a project, anticipated site conditions, and importance of geotechnical structures. For example, when the type of structures is sensitive to differential settlement or tolerable risk is set at a stringent level, a stringent accuracy requirement (i.e., a relatively small RE_T) should be adopted in engineering practice. $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h reflect the spatial variability of site conditions, which may be obtained from different sources, such as aerial photographs, topographic maps, soil survey maps and data, historical data from previous investigations at the site or nearby sites, previous experiences from the area, published reports and literature about the similar soil properties and site conditions (e.g., Phoon and Kulhawy 1999; FHWA 2017). After RE_T , $\text{COV}_{(N_1)_{60}}$, λ_v and λ_h are determined, normalized target relative error, $\text{RE}_T / \text{COV}_{(N_1)_{60}}$ can be calculated. Then, the corresponding normalized SPT spacing d_N can be obtained directly using the proposed statistical chart (e.g., the best-fit curve in Fig. 11). When the depth interval d_v of SPT is determined, the required SPT spacing along a horizontal direction can be obtained as

$d_h = (d_N \times \lambda_h \times \lambda_v) / d_v$ for a specific site. Note that the developed statistical chart should be used in conjunction with the Bayesian compressive sampling. In addition, the minimum and maximum values of the normalized SPT spacing d_N in all 4088 data pairs are about 0.02 and 15, respectively. Therefore, a d_N range of from 0.02 to 15 is recommended for proper application of the proposed statistical chart in engineering practice. In the next section, a set of real SPT data from New Zealand is used to illustrate and validate the proposed method and statistical chart.

Real SPT data example

In this section, the proposed method and statistical chart are used to determine the minimum number of SPTs for characterizing spatial variation of SPT data (e.g., $(N_1)_{60}$) within a cross-section A-A' at a site in Wellington, New Zealand, as shown by a dashed red line in Fig. 12. Based on the records from previous boreholes conducted at this site, the upper layer of the site ranging from a depth of 0m to about 10m mainly consists of silty sand. The saturated unit weight of silty sand γ_{sat} is about 20kN/m³, and the water table at the site is ranging from 0.1m to 5.2m below the ground surface.

The width of cross-section A-A' is about 354m, while the depth of cross-section A-A' is from about 1.5m to about 9.0m. Suppose that the accuracy requirement for the interpreted cross-section of SPT data, RE_T , is set as 45%. Note that $RE_T=45\%$ selected in this example is for illustrative purposes only. In engineering practice, the target relative error shall be selected in accordance with the acceptable risk thresholds of a project, expected site conditions, and importance of the structures. $COV_{(N_1)_{60}}$, λ_v and λ_h of this site are estimated as 50%, 2.5m, and 60m based on SPT data performed at nearby sites previously, and they are also within the typical ranges of these parameters reported in literature (e.g., Vanmarcke 1977; Phoon and

546 Kulhawy 1999). Then, the normalized target relative error, $RE_T / COV_{(N_1)_{60}}$ can be calculated as
 547 0.9. Using the best-fit curve in Fig. 11, the corresponding normalized SPT spacing, d_N can be
 548 obtained as 0.49. The depth interval d_v of SPT is set as 1.5m (i.e., $M_{x_2} = D / d_v + 1$
 549 $= 7.5\text{m} / 1.5\text{m} + 1 = 6$) which is commonly used in geotechnical practice. As
 550 $d_h = (d_N \times \lambda_h \times \lambda_v) / d_v$, the required borehole spacing d_h is estimated as 49m. Therefore, the
 551 number $M_{x_1} = 354 / 49 + 1 \approx 8$ boreholes with an equal spacing should be conducted for
 552 characterizing variation of SPT data within the cross-section A-A'. The total number of SPT
 553 measurement data needed is $M = M_{x_2} \times M_{x_1} = 6 \times 8 = 48$.

554 To examine the performance of the proposed method for this site, a set of real SPT data
 555 along cross-section A-A' is obtained from New Zealand Geotechnical Database (NZGD)
 556 (NZGD 2020). In total, 94 $(N_1)_{60}$ data points are obtained from 15 boreholes with SPTs
 557 performed at this site, i.e., BH-98066, BH-72342, ..., BH-98070, as shown by solid triangles
 558 in Fig. 12. Based on the result of the proposed method, 8 SPT boreholes shall be performed
 559 with equal spacing of about 50m within the cross-section A-A'. Therefore, 51 $(N_1)_{60}$ data points
 560 from 8 boreholes (i.e., BH-98066, BH-106086, BH-98079, BH-106087, BH-98076, BH-
 561 106089, BH-98074, and BH-98070) are taken as measurement data (i.e., input) for the
 562 proposed method. As shown in Fig. 12 by solid circles, these 8 boreholes have an
 563 approximately equal spacing of 50m within the cross-section A-A'. In other words, these 51
 564 $(N_1)_{60}$ data points are used as the input to 2D BCS for interpolating high-resolution cross-
 565 section A-A' of $(N_1)_{60}$ data. The horizontal and vertical resolutions of cross-section A-A' of
 566 $(N_1)_{60}$ data are set as 1m and 0.1m, respectively, leading a dimension of 76×355 for the cross-
 567 section A-A' of $(N_1)_{60}$. The cross-section A-A' of $(N_1)_{60}$ data interpreted from 51 $(N_1)_{60}$
 568 measurement data using 2D BCS is shown in Fig. 13 by a contour map. The original 94 $(N_1)_{60}$
 569 data points from 15 SPT boreholes and those at corresponding locations estimated from 2D

BCS are shown in Fig. 14 by open triangles and open circles, respectively. The first row of Fig. 14 (i.e., Fig. 14(a)-(h)) is the input for the proposed method (i.e., 8 boreholes with SPTs), while the second row of Fig. 14 (i.e., Fig. 14 (i)-(o)) is 7 SPT boreholes (i.e., BH-72372, BH-98069, BH-98075, BH-98077, BH-98078, BH-98067, and BH-98090) that are used as unsampled locations for the proposed method. It is observed from Fig. 14(i)-(o) that the trend of interpreted $(N_1)_{60}$ values at 7 unsampled locations are generally consistent with the true one, although several predicted $(N_1)_{60}$ values deviate significantly from the measured values. The relative error between original 94 $(N_1)_{60}$ data and those at corresponding locations estimated from 2D BCS is calculated as 44%, which is consistent with the target relative error $RE_T=45\%$. Such consistency indicates that the cross-section A-A' of $(N_1)_{60}$ interpreted from 8 SPT boreholes, which is quantitatively determined using the proposed method, satisfies the accuracy requirement.

Conclusion

A comprehensive parametric study was performed in this study to develop a rational method for selecting the minimum number of SPTs needed for satisfying accuracy requirement in geotechnical site investigation. Based on results of the parametric study, spatial variability and correlation in soil properties (i.e., coefficient of variation $COV_{(N_1)_{60}}$, vertical scale of fluctuation λ_v , and horizontal scale of fluctuation λ_h) have a significant effect on the minimum SPT number required for achieving a target accuracy level on site investigation results, and a quantitative relationship between the SPT number and the accuracy of interpretation results was developed for a wide variety of typical site conditions often encountered in engineering practice. A statistical chart is further developed for facilitating geotechnical engineers to select the minimum SPT number in engineering practice. The proposed method is illustrated using a

594 set of real SPT data at a site in Wellington, New Zealand. The results show that the proposed
595 method performs well.

596 Note that uncertainty in the interpreted cross-section of $(N_1)_{60}$ data from sparse
597 measurement data (e.g., variance of the error) is not considered either in this study. Future
598 research is needed to further extend the proposed method to take into account the uncertainty
599 in the interpreted cross-section of $(N_1)_{60}$ data and evaluate the effect of target relative error of
600 interpreted cross-section of $(N_1)_{60}$ data on reliability index for reliability-based design in
601 geotechnical engineering.

Draft

REFERENCES

- AASHTO. 2018. Standard Specifications for Transportation Materials and Methods of Sampling and Testing, and AASHTO Provisional Standards. American Association of State Highway and Transportation Officials, Washington, DC.
- Balasubramani, D. P., and Dodagoudar, G. R. 2020. Modelling the spatial variability of Standard Penetration Test data for Chennai City using kriging and product-sum model. *Geomechanics and Geoengineering*, doi: 10.1080/17486025.2019.1707884
- Boeckmann, A.Z., and J.E. Loehr. 2016. Influence of Geotechnical Investigation and Subsurface Conditions on Claims, Change Orders, and Overruns. National Cooperative Highway Research Program Synthesis 484, Transportation Research Board, Washington, DC.
- Boulanger, R.W., and Idriss, I. M. 2014. CPT and SPT Based Liquefaction Triggering Procedures. Report No. UCD/CGM-14/01, Center for Geotechnical Modeling, University of California-Davis.
- Candès, E. J., Romberg, J. K., and Tao, T. 2006. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8): 1207-1223.
- Cao, Z., Wang, Y., and Li, D. 2016. Quantification of prior knowledge in geotechnical site characterization. *Engineering Geology*, 203: 107-116.
- CEN (European Committee for Standardization). 2007. Geotechnical design—part 2: ground investigation and testing (Eurocode 7-2). Brussels, Belgium: CEN.
- Ching, J., and Phoon, K. K. 2012. Value of geotechnical site investigation in reliability-based design. *Advances in Structural Engineering*, 15(11): 1935-1945.
- Cho, S. E. 2007. Effects of spatial variability of soil properties on slope stability. *Engineering Geology*, 92(3-4): 97-109.

- 627 Christian, J.T. 2004. Geotechnical Engineering Reliability: How Well Do We Know What We
628 Are Doing? Journal of Geotechnical and Geoenvironmental Engineering, 130: 10(985).
- 629 Clayton, C. R. 1995. The standard penetration test (SPT): methods and use Construction
630 Industry Research and Information Association Report 143, CIRIA, London.
- 631 Clayton, C. R. 2001. Managing geotechnical risk: improving productivity in UK building and
632 construction. Thomas Telford Publishing, London.
- 633 Craig, R. F. 2004. Craig's soil mechanics. CRC press, London.
- 634 Crisp, M., Jaksa, M. B., and Kuo, Y. L. 2019. Toward a generalized guideline to inform optimal
635 site investigations for pile design. Canadian Geotechnical Journal, doi: [10.1139/cgj-](https://doi.org/10.1139/cgj-2019-0111)
636 [2019-0111](https://doi.org/10.1139/cgj-2019-0111).
- 637 Dietrich, C. R., and Newsam, G. N. 1993. A fast and exact method for multidimensional
638 Gaussian stochastic simulations. Water Resources Research, 29(8): 2861-2869.
- 639 Fang, Y., Wu, J., and Huang B. 2012. 2D sparse signal recovery via 2D orthogonal matching
640 pursuit. Science CHINA Inform Sciences, 55(4): 889-897.
- 641 Fenton, G. A. 1999. Random field modeling of CPT data. Journal of geotechnical and
642 geoenvironmental engineering, 125(6): 486-498.
- 643 Fenton, G. A., and Griffiths, D. V. 2008. Risk assessment in geotechnical engineering. John
644 Wiley and Sons, New York.
- 645 Fenton, G. A., Liza, R., Lake, C. B., Menzies, W. T. and Griffiths, D. V. 2015. Statistical
646 sample size for quality control programs of cement-based "solidification/stabilization".
647 Canadian Geotechnical Journal, 52(10):1620-1628.
- 648 FHWA. 2017. Geotechnical Site Characterization. Geotechnical Engineering Circular No. 5,
649 Publication No. NHI-16-072. Federal Highway Administration, U.S. Department of
650 Transportation, Washington, DC.

- 651 Goldsworthy, J. S., Jaksa, M. B., Fenton, G. A., Kaggwa, W. S., Griffiths, V., and Poulos, H.
652 G. 2007. Effect of sample location on the reliability based design of pad foundations.
653 Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards,
654 1(3): 155-166.
- 655 Ground Board, 1991. Inadequate site investigation, Institution of Civil Engineers, Thomas
656 Telford, London.
- 657 Guan, Z., and Wang, Y., 2019. Statistical charts for determining sample size at various levels
658 of accuracy and confidence in geotechnical site investigation. *Géotechnique*, doi:
659 [10.1680/jgeot.18.P.315](https://doi.org/10.1680/jgeot.18.P.315).
- 660 Guan, Z., Wang, Y., Cao, Z., and Hong, Y. 2020. Smart sampling strategy for investigating
661 spatial distribution of subsurface shallow gas pressure in Hangzhou Bay area of China.
662 *Engineering Geology*, <https://doi.org/10.1016/j.enggeo.2020.105711>.
- 663 Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., and Zhang, X. P. 2014. Bayesian
664 nonparametric dictionary learning for compressed sensing MRI. *IEEE Transactions on*
665 *Image Processing*, 23(12): 5007-5019.
- 666 Hu, Y., Wang, Y., Zhao, T., and Phoon, K. K. 2020. Bayesian Supervised Learning of Site-
667 Specific Geotechnical Spatial Variability from Sparse Measurements. *ASCE-ASME*
668 *Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*,
669 6(2), 04020019.
- 670 Jaksa, M. B., Goldsworthy, J. S., Fenton, G. A., Kaggwa, W. S., Griffiths, D. V., Kuo, Y. L.,
671 and Poulos, H. G. 2005. Towards reliable and effective site investigations.
672 *Géotechnique*, 55(2): 109-121.
- 673 Ji, S., Dunson, D., and Carin, L. 2009. Multitask compressive sensing. *IEEE Transactions on*
674 *Signal Processing*, 57(1): 92-106.

- 675 Ji, S., Xue, Y., and Carin, L. 2008. Bayesian compressive sensing. *IEEE Transactions on Signal*
676 *Processing*, 56(6): 2346-2356.
- 677 Kuzin, D., Isupova, O., and Mihaylova, L. 2015. Compressive sensing approaches for
678 autonomous object detection in video sequences. In *2015 Sensor Data Fusion: Trends,*
679 *Solutions, Applications (SDF)* (pp. 1-6). IEEE.
- 680 Loehr, J. E., Ding, D., and Likos, W. J. 2015. Effect of Number of Soil Strength Measurements
681 on Reliability of Spread Footing Designs. *Transportation Research Record: Journal of*
682 *the Transportation Research Board*, 2511(1): 37-44.
- 683 Lacasse S, and Nadim F. 1996. Uncertainties in characterizing soil properties. In: *Uncertainties*
684 *in the geologic environment: from theory to practice (GSP 58)*, New York.
- 685 Li, D. Q., Jiang, S. H., Cao, Z. J., Zhou, W., Zhou, C. B., and Zhang, L. M. 2015. A multiple
686 response-surface method for slope reliability analysis considering spatial variability of
687 soil properties. *Engineering Geology*, 187, 60-72.
- 688 Mayne, P. W. 2006. In-situ test calibrations for evaluating soil parameters. In *Characterisation*
689 *and Engineering Properties of Natural Soils—Proceedings of the Second International*
690 *Workshop on Characterisation and Engineering Properties of Natural Soils: Taylor and*
691 *Francis, London.*
- 692 Marques, E. C., Maciel, N., Naviner, L., Cai, H., and Yang, J. 2018. A review of sparse
693 recovery algorithms. *IEEE Access*, 7, 1300-1322.
- 694 Mayne, P. W., Christopher, B. R., and Dejong, J. 2002. *Manual on subsurface investigations.*
695 *Nat. Highway Inst. Sp. Pub. FHWA NHI-01-031. Fed. Highway Administ,*
696 *Washington, DC.*
- 697 Michigan DOT. 2019. *Geotechnical Design Manual.* Michigan Department of Transportation,
698 *Lansing, Michigan.*

- 699 Montoya-Noguera, S., Zhao, T., Hu, Y., Wang, Y., and Phoon, K. K. 2019. Simulation of non-
700 stationary non-Gaussian random fields from sparse measurements using Bayesian
701 compressive sampling and Karhunen-Loève expansion. *Structural Safety*, 79: 66-79.
- 702 New York State DOT. 2013. *Geotechnical Design Manual*. Albany, New York.
- 703 NZGD World Wide Web Address: <https://www.nzgd.org.nz>2020 [accessed at 19/05/2020]
- 704 Oguz, E. A., Huvaj, N., and Griffiths, D. V. 2019. Vertical spatial scale of fluctuation based on
705 standard penetration tests. *Marine Georesources and Geotechnology*, 37(1): 45-56.
- 706 Ott, W. R. 1994. *Environmental statistics and data analysis*. CRC Press, London.
- 707 Phoon, K. K., and Kulhawy, F. H. 1999. Characterization of geotechnical variability. *Canadian*
708 *Geotechnical Journal*, 36(4): 612-624.
- 709 Prezzi, M., B. McCullough, and V.K.D. Mohan. 2011. *Analysis of Change Orders in*
710 *Geotechnical Engineering Work at INDOT*. Publication FHWA/IN/JTRP-2011/10.
711 Joint Transportation Research Program, Indiana Department of Transportation and
712 Purdue University, West Lafayette, Indiana.
- 713 Rix, G.J., Wainaina, N., Ebrahimi, A., Bachus, R.C., Limas, M., Sancio, R., Fait, B., and
714 Mayne, P.W. 2018. *Manual on Subsurface Investigations* (No. NCHRP Project 21-10).
715 Transportation Research Board, Washington, D.C.
- 716 Skempton, A. W. 1986. Standard penetration test procedures and the effects in sands of
717 overburden pressure, relative density, particle size, ageing and overconsolidation.
718 *Géotechnique*, 36(3): 425-447.
- 719 Stuedlein, A. W., Kramer, S. L., Arduino, P., and Holtz, R. D. 2012. Geotechnical
720 characterization and random field modeling of desiccated clay. *Journal of Geotechnical*
721 *and Geoenvironmental Engineering*, 138(11): 1301-1313.
- 722 Tipping, M. E. 2001. Sparse Bayesian learning and the relevance vector machine. *Journal of*
723 *Machine Learning Research*, 1: 211-244.

- 724 Vanmarcke, E. 1977. Probabilistic modeling of soil profiles. *Journal of Geotechnical*
725 *Engineering*, 103(11): 1227-1246.
- 726 Vanmarcke, E. 2010. *Random fields: analysis and synthesis*, World Scientific, London.
- 727 Wang, J. P., and Xu, Y. 2015. Estimating the standard deviation of soil properties with limited
728 samples through the Bayesian approach. *Bulletin of Engineering Geology and the*
729 *Environment*, 74(1): 271-278.
- 730 Wang, Y., Zhao, T., and Cao, Z. 2015 Site-specific probability distribution of geotechnical
731 properties. *Computers and Geotechnics*, 70: 159-168.
- 732 Wang, Y. and Zhao, T. 2016. Interpretation of soil property profile from limited measurement
733 data: a compressive sampling perspective. *Canadian Geotechnical Journal*, 53(9): 1547-
734 1559.
- 735 Wang, Y., and Zhao, T. 2017. Statistical interpretation of soil property profiles from sparse
736 data using Bayesian compressive sampling. *Géotechnique*, 67(6): 523-536.
- 737 Wang, Y., Akeju, O. V., and Zhao, T. 2017. Interpolation of spatially varying but sparsely
738 measured geo-data: A comparative study. *Engineering Geology*, 231: 200-217.
- 739 Wang, Y., Cao, Z., and Li, D. 2016. Bayesian perspective on geotechnical variability and site
740 characterization. *Engineering Geology*, 203: 117-125.
- 741 Wang, Y., Guan, Z., and Zhao, T. 2019b. Sample size determination in geotechnical site
742 investigation considering spatial variation and correlation. *Canadian Geotechnical*
743 *Journal*, 56(7): 992-1002.
- 744 Wang, Y., Hu, Y., and Zhao, T. 2019a. CPT-based subsurface soil classification and zonation
745 in a 2D vertical cross-section using Bayesian compressive sampling. *Canadian*
746 *Geotechnical Journal*, doi: 10.1139/cgj-2019-0131.
- 747 Wang, Y., Zhao, T., and Cao, Z. 2015. Site-specific probability distribution of geotechnical
748 properties. *Computers and Geotechnics*, 70: 159-168.

- 749 Zhang, L., and Chen, J. 2012. Effect of spatial correlation of standard penetration test (SPT)
750 data on bearing capacity of driven piles in sand. Canadian Geotechnical Journal, 49(4):
751 394-402.
- 752 Zhao, T. and Wang, Y. 2019. Determination of efficient sampling locations in geotechnical site
753 characterization using information entropy and Bayesian compressive sampling.
754 Canadian Geotechnical Journal, 56(11): 1622-1637.
- 755 Zhao, T., Hu, Y., and Wang, Y. 2018. Statistical interpretation of spatially varying 2D geo-
756 data from sparse measurements using Bayesian compressive sampling. Engineering
757 Geology, 246: 162-175.

Draft

758 List of Symbols

759 SPT: standard penetration test

760 M_{x_1} : the number of boreholes with SPTs

761 M_{x_2} : the total number of measurement data in each SPT borehole

762 d_v : SPT depth interval

763 d_h : borehole spacing

764 RE: relative error

765 μ_{RE} : mean relative error

766 N_{60} : corrected SPT N value

767 $(N_1)_{60}$: normalized SPT value

768 σ'_{v0} : effective vertical overburden stress

769 $\mu_{(N_1)_{60}}$: mean of $(N_1)_{60}$

770 $\sigma_{(N_1)_{60}}$: standard deviation of $(N_1)_{60}$

771 λ_v : vertical scale of fluctuation

772 λ_h : horizontal scale of fluctuation

773 $COV_{(N_1)_{60}}$: coefficient of variation of $(N_1)_{60}$

774 $\rho(\Delta x_1, \Delta x_2)$: correlation coefficient between any two $(N_1)_{60}$ data points with absolute distance

775 Δx_1 and Δx_2 along the horizontal and vertical direction, respectively

776 D : depth of cross-section of $(N_1)_{60}$ data

777 W : width of cross-section of $(N_1)_{60}$ data

778 $\mathbf{F}$: a $k_{x_2} \times k_{x_1}$ matrix represents 2D high-resolution vertical cross-section of $(N_1)_{60}$ data

779 $\hat{\mathbf{F}}$: the cross-section of $(N_1)_{60}$ data interpreted from M_{x_1} SPT boreholes conducted in $\mathbf{F}$

780 K : a total number of elements in $\mathbf{F}$, $K = k_{x_2} \times k_{x_1}$

- 781 $\mathbf{Y}$: an $M_{x_2} \times M_{x_1}$ matrix
- 782 M : a total number of SPT measurement data, $M = M_{x_2} \times M_{x_1}$
- 783 $\mathbf{B}_t^{2D}$: t -th orthonormal basis function
- 784 ω_t^{2D} : the weight coefficient corresponding to $\mathbf{B}_t^{2D}$
- 785 $\boldsymbol{\omega}^{2D}$: $\boldsymbol{\omega}^{2D} = [\omega_1^{2D}, \omega_2^{2D}, \dots, \omega_K^{2D}]$ represents K weight coefficients for different basis functions
- 786 $\hat{\boldsymbol{\omega}}^{2D}$: the approximation of $\boldsymbol{\omega}^{2D}$
- 787 $\boldsymbol{\Psi}_{x_2}$: an $M_{x_2} \times k_{x_2}$ matrix
- 788 $\boldsymbol{\Psi}_{x_1}$: a $k_{x_1} \times M_{x_1}$ matrix
- 789 $\mathbf{D}$: a $K \times K$ diagonal matrix
- 790 $\mathbf{A}_t^{2D}$: an $M_{x_2} \times M_{x_1}$ matrix, $\mathbf{A}_t^{2D} = \boldsymbol{\Psi}_{x_2} \mathbf{B}_t^{2D} \boldsymbol{\Psi}_{x_1}$
- 791 $\mathbf{J}$: a $K \times K$ matrix with elements $J_{t,s} = \text{tr}[\mathbf{A}_t^{2D} (\mathbf{A}_s^{2D})^T]$ ($t, s = 1, 2, \dots, K$)
- 792 $\text{tr}(\bullet)$: trace of a matrix
- 793 $\boldsymbol{\mu}_{\hat{\boldsymbol{\omega}}^{2D}}$: mean of $\boldsymbol{\omega}^{2D}$
- 794 $\boldsymbol{\mu}_{\hat{\mathbf{F}}}$: mean of $\hat{\mathbf{F}}$
- 795 d_N : normalized borehole spacing
- 796

Table 1. Typical ranges of the mean, coefficient of variation and scale of fluctuation of SPT data for a variety of site conditions

SPT data	Mean	Coefficient of variation, COV (%)	Scale of fluctuation (m)	
			Vertical direction: λ_v	Horizontal direction: λ_h
Normalized SPT value $(N_1)_{60}$	0.1-100	1-100	0.1-10	1-100

Note: $(N_1)_{60} = C_N N_{60}$, where N_{60} is the corrected SPT value; $C_N = (\sigma_{atm} / \sigma_{v0}')^{0.5} \leq 2$, $\sigma_{atm} = 100$ kPa, σ_{v0}' = effective vertical overburden stress.

Table 2. Summary of parameters used in parametric study

Parameter	Variation
Mean, $\mu_{(N_1)_{60}}$	0.1, 10, 100
Coefficient of variation, $COV_{(N_1)_{60}}$ (%)	5, 10, 20, 30, 40, 50, 80, 100
Horizontal scale of fluctuation, λ_h (m)	1, 5, 10, 15, 20, 30, 40, 50, 75, 100
Vertical scale of fluctuation, λ_v (m)	0.5, 1, 2, 4, 5, 6, 8, 10
Depth interval, d_v of SPT (m)	1.0, 1.5, 2.5, 3.0, 5.0
The number, M_{x_i} of boreholes with SPTs	5, 8, 10, 12, 15, 18, 20, 22, 25, 28, 30
Depth of vertical cross-section (m)	9, 15, 30, 45, 90
Width of vertical cross-section (m)	50, 100, 250, 1000

Table 3. Summary of the best-fit equations for normally and lognormally distributed random fields with three different correlation function types

Correlation function type	Normal distribution	Lognormal distribution
SECF	$y=0.53+0.64 \times (1-e^{-1.75x})$, $R^2=0.77$	$y=0.53+0.62 \times (1-e^{-1.60x})$, $R^2=0.80$
GCF	$y=1.15-1.15e^{-1.88x}$, $R^2=0.83$	$y=1.14-1.14e^{-1.91x}$, $R^2=0.84$
CECF	$y=0.42+0.73 \times (1-e^{-1.26x})$, $R^2=0.82$	$y=0.42+0.72 \times (1-e^{-1.29x})$, $R^2=0.83$

Note: $y = \mu_{RE} / \text{COV}_{(N_1)_{60}}$; $x = (d_h / \lambda_h) \times (d_v / \lambda_v)$; SECF, GCF and CECF represent single exponential correlation function, Gaussian correlation function and cosine exponential correlation function, respectively.

Draft

Figure Captions

Fig. 1. Flowchart of the extensive parametric study

Fig. 2. Four 2D high-resolution vertical cross-sections of $(N_1)_{60}$ data (i.e., $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$) generated from a 2D random field for a site condition with medium dense sand

Fig. 3. The $(N_1)_{60}$ cross-sections interpreted from different numbers, $M_{x_1}=5, 10, 20, 30$ of boreholes with SPTs performed with depth interval $d_v=1.5\text{m}$ (i.e., $M_{x_2}=11$) using Bayesian compressive sampling for $\mathbf{F}_1$

Fig. 4. The $(N_1)_{60}$ profile at depth = 6m for $\mathbf{F}_1$ interpreted from different numbers, $M_{x_1}=5, 10, 20, 30$ of boreholes with SPTs performed with depth interval $d_v=1.5\text{m}$ (i.e., $M_{x_2}=11$)

Fig. 5. Relative error, RE versus the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v = 1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}$ (i.e., $M_{x_2}=16, 11, 7, 6$) for $\mathbf{F}_1$

Fig. 6. Variation of relative error with the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v = 1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}$ (i.e., $M_{x_2}=16, 11, 7, 6$) for 100 random field samples (i.e., $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{100}$)

Fig. 7. Mean relative error, μ_{RE} , versus the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v = 1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}$ (i.e., $M_{x_2}=16, 11, 7, 6$) within a $(N_1)_{60}$ random field (i.e., a site condition with medium dense sand)

Fig. 8. Dimensionless plot of normalized mean relative error versus normalized SPT spacing

Fig. 9. Normalized mean relative error $\mu_{\text{RE}} / \text{COV}_{(N_1)_{60}}$ versus d_h / λ_h and d_v / λ_v

Fig. 10. Comparison among the best-fit curves for normally and lognormally distributed random fields with single exponential correlation function (SECF), Gaussian correlation function (GCF), and cosine exponential correlation function (CECF)

Fig. 11. Statistical chart for selecting the minimum number of SPTs needed to meet an accuracy requirement of site investigation

Fig. 12. Locations of 15 boreholes with SPTs and a cross-section at Wellington, New Zealand (NZGD 2020)

Fig. 13. High-resolution A-A' cross-section of $(N_1)_{60}$ interpreted from 8 boreholes with SPTs using Bayesian compressive sampling

Fig. 14. Comparison between 94 original $(N_1)_{60}$ data points and those at corresponding locations interpreted from 8 boreholes with SPTs using Bayesian compressive sampling

Fig. 1. Flowchart of the extensive parametric study

Fig. 2. Four 2D high-resolution vertical cross-sections of $(N_1)_{60}$ data (i.e., F_1 , F_2 , F_3 , F_4) generated from a 2D random field for a site condition with medium dense sand

Fig. 3. The $(N_1)_{60}$ cross-sections interpreted from different numbers, $M_{x_1} = 5, 10, 20, 30$ of boreholes with SPTs performed with depth interval

$d_v = 1.5\text{m}$ (i.e., $M_{x_2} = 11$) using Bayesian compressive sampling for F_1

Fig. 4. The $(N_1)_{60}$ profile at depth = 6m for F_1 interpreted from different numbers, $M_{x_1}=5, 10, 20, 30$ of boreholes with SPTs performed with depth interval $d_v=1.5\text{m}$ (i.e., $M_{x_2}=11$)

Fig. 5. Relative error, RE versus the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v = 1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}, 5\text{m}$ (i.e., $M_{x_2}=16, 11, 7, 6, 4$) for F_1

Fig. 6. Variation of relative error with the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v=1\text{m}$, 1.5m , 2.5m , 3m (i.e., $M_{x_2}=16$, 11 , 7 , 6) for 100 random field samples (i.e., $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{100}$)

Fig. 7. Mean relative error, μ_{RE} , versus the number, M_{x_1} of boreholes with SPTs performed with different depth intervals $d_v=1\text{m}, 1.5\text{m}, 2.5\text{m}, 3\text{m}, 5\text{m}$ (i.e., $M_{x_2}=16, 11, 7, 6, 4$) within a $(N_1)_{60}$ random field (i.e., a site condition with medium dense sand)

Fig. 8. Dimensionless plot of normalized mean relative error versus normalized SPT spacing

Fig. 9. Normalized mean relative error $\mu_{RE} / COV_{(N_1)_{60}}$ versus d_h / λ_h and d_v / λ_v

Fig. 10. Comparison among the best-fit curves for normally and lognormally distributed random fields with single exponential correlation function (SECF), Gaussian correlation function (GCF), and cosine exponential correlation function (CECF)

Fig. 11. Statistical chart for selecting the minimum number of SPTs needed to meet an accuracy requirement of site investigation

Fig. 12. Locations of 15 boreholes with SPTs and a cross-section at Wellington, New Zealand (NZGD 2020)

Fig. 13. High-resolution A-A' cross-section of $(N_1)_{60}$ interpreted from 8 boreholes with SPTs using Bayesian compressive sampling

Note: The first row (i.e., Fig.14(a)-(h)) is the input to the proposed method (i.e., 8 boreholes with SPTs), as shown by solid circles in Fig. 12

Fig. 14. Comparison between 94 original $(N_1)_{60}$ data points and those at the same locations and interpreted from 8 boreholes with SPTs using Bayesian compressive sampling