Computers and Geotechnics 118 (2020) 103331

Contents lists available at ScienceDirect

Computers and Geotechnics

journal homepage: www.elsevier.com/locate/compgeo

Research Paper

On the optimization of site investigation programs using centroidal Voronoi tessellation and random field theory

Linchong Huang[a], Shuai Huang[b], Zhengshou Lai[b,*]

[a] School of Aeronautics and Astronautics Engineering, Sun Yat-sen University, Guangzhou 510275, China
[b] School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China

ARTICLE INFO

Keywords:
Site investigation
Centroidal Voronoi tessellation
Random field
Nonstructural grid
Irregular site geometry

ABSTRACT

This work proposes a centroidal Voronoi tessellation (CVT)-based site investigation scheme, which is applicable to arbitrary site geometries and arbitrary numbers of investigation locations. A modified Lloyd algorithm is developed to generate CVT-based site investigation programs. With the data from prior site investigations, the random field statistics of a site property are estimated using a Markov chain Monte Carlo simulation-based Bayesian inference approach, and the site property at each location is estimated using Kriging interpolation. The performance of the CVT-based site investigation programs and optimization are demonstrated using two illustrative examples, namely, a square site and an irregular site.

1. Introduction

In geotechnical engineering, important site properties may include the soil density, compression index, friction angle, permeability, and liquefaction potential, among others. These properties could be crucial for many engineering applications. The fact that site properties are spatially variable has long been recognized by geotechnical engineers and researchers [1–5]. The spatial variability of a site property could result from various factors, such as fluctuations in material constituents, random alignments during the transport and deposition of soil grains, variable historic loading conditions. The main aim of site characterization is, for structure design purposes, to characterize the geotechnical ground model, including to determine or estimate the values of a site property at every location within the site, and to distinguish the distribution of soil layers to which those parameters may be related. Hence, this process has become an essential task for many geotechnical engineering applications.

Despite the spatial variability of site properties, the values of a site property at all locations are generally correlated with or dependent on each other. Such autocorrelation features provide a good basis for the development of site characterization approaches. One option is to treat the site characterization problem as a function approximation task consisting of existing data. In this fashion, Samui and Sitharam [6] applied an artificial neural network model to map the standard penetration test results at every site location and took existing data from borehole measurements as training inputs. Another category of approaches that is more prevalent than others is known as random field (RF) theory, which has been shown to be rather effective in modeling spatially variable and autocorrelated site properties [7–12]. The application of RF theory has already led to numerous excellent studies in geotechnical engineering. For example, RF theory has been applied to simulate soil properties in probabilistic stability analyses of slopes and footings [13], to model the hydrodynamics of seabed pipelines [14], to evaluate static liquefaction in dilative sand fill [15], to conduct the areal mapping of the effectiveness of soil compaction [5], and to delineate contaminated areas [12].

By using RF theory, the values of a site property at every site location can be discretized into a list of random variables. These random variables are constrained by a probability distribution type, a spatial autocorrelation function, and the corresponding statistical parameters (e.g., mean, standard deviation, and scales of fluctuation). Site characterization thus determines the RF statistics associated with a particular site property. Generally, site characterization is a multistep process that can be divided into six stages: (1) desk study, (2) site reconnaissance, (3) in situ investigation, (4) laboratory testing, (5) interpretation of site observation data, and (6) inference of soil properties, rock properties or underground stratigraphy [16]. Among these stages, the in situ site investigation is the primary stage. At this stage, the site properties are measured at designated locations so that the site properties at every other location within the site can be approximated based on selected characterization models [17].

The accuracy of site characterization is significantly affected by the

* Corresponding author.
E-mail address: laizhengsh@mail.sysu.edu.cn (Z. Lai).

https://doi.org/10.1016/j.compgeo.2019.103331
Received 5 June 2019; Received in revised form 29 October 2019; Accepted 31 October 2019

extent of the in situ site investigation program. The acquisition of additional site investigation data generally results in more accurate site characterization results. However, only a small portion (e.g., on the order of 1/100,000 or less) of the total site area/volume can typically be sampled for a practical engineering project due to the required commitment of financial and human capital [18–21]. It thus becomes a challenging and demanding problem to achieve a balance between the site characterization accuracy and site investigation effort. In fact, increasing amounts of research have recently been performed on optimizing site investigation programs. For example, Gong et al. [22] developed a maximum likelihood approach to characterize soil property statistics and thus to optimize the site investigation program in terms of the improved prediction of tunneling-induced ground settlement. Li et al. [23] presented an approach using 3D Kriging for geotechnical sampling schemes and Monte Carlo methods for probabilistic slope stability analyses to investigate the optimum sampling locations and cost-effective design of a slope. Gong et al. [24] developed a biobjective optimization framework for improving site investigation programs by considering the robustness of characterized statistics and the number of site investigation locations. Yang et al. [25] compared the performance of nine different schemes (i.e., with various locations, spacings and depths of monitoring sections) at characterizing the spatial variability of soil properties for a slope problem using a probabilistic back-analysis framework.

Usually, site investigation programs are designed to follow some regular pattern. For example, due to the simplicity in its implementation, a structural grid of investigation locations is commonly used (e.g., [26–29,24]). Nevertheless, despite the recent progress achieved in the optimization of site investigation programs, most of the aforementioned studies were only concerned with site investigation programs based on regular patterns; such programs could have at least two limitations. First, the number of investigation locations may not change continuously. Taking a square isotropic site as an example, an investigation program using a 6×6 mesh grid would result in 36 investigation locations, while a program using a 7×7 mesh grid would result in 49 locations. There is a considerable jump in the number of locations from the grid of 6×6 to that of 7×7, whereas the optimum site investigation program could exist with a grid somewhere between 6×6 and 7×7. Although a program with an anisotropic mesh grid (e.g., 5×7, 6×7, 5×8, or 6×8, see Fig. 1a) could be adopted, such programs may not be optimal for an isotropic site because they are not uniformly distributed in each dimension [30,23,25]. Second, it is not straightforward to apply a structural grid to a site with an irregularly shaped domain geometry (see Fig. 1b). There could exist situations in which the soil properties of an irregularly shaped site are to be characterized (e.g., structures with irregular layout, or airport, public square, farms with irregular geometries). There is, however, rare literature or techniques on designing site investigation programs with an arbitrary number of investigation locations or arbitrary site geometries. To approach such gap, in this work, centroidal Voronoi tessellation (CVT) is utilized to derive site investigation programs. CVTs partition a plane into regions based on the distances to points in a specific subset of the plane [31]. CVTs have been extensively applied in image compression, quadrature methods, finite difference methods, resource distributions, cellular biology, statistics, and the territorial behavior of animals [31]. In this work, the centroids of CVT cells are taken as site investigation locations. Such a CVT-based site investigation program is advantageous because it can accommodate an arbitrary number of investigation locations and is applicable to arbitrary site geometries. To evaluate the performance of the resulting CVT-based site investigation programs, two types of robustness, (1) the signal-to-noise ratio (SNR) of characterized RF statistics and (2) the SNR of the Kriging-interpolated site property at every location, are employed. The results are compared with site investigation programs containing random investigation locations or structural grids. The optimum site investigation program is then approached using the obtained CVT-based site investigation programs through a biobjective optimization process.

The remainder of this paper is structured as follows. Section 2 briefly presents the theories pertaining to site characterization, including RF theory, the theory of Kriging interpolation, and the back-calculation of RF statistics. Section 3 describes the Lloyd algorithm employed to generate the CVT-based site investigation programs and the robustness metrics used to evaluate the performance of the developed site investigation programs. Section 4 reports the results of the CVT-based site investigation and optimization for two illustrative examples, namely, a square site and an irregularly shaped site. Section 5 presents some further discussions and Section 6 summarizes the concluding remarks of this work.

2. Site characterization

2.1. Random field theory

One primary objective of site characterization is to map the site properties of interest onto the site domain. In this respect, RF theory provides an effective way to model spatially variable site properties at every site location. Herein, the site property of interest is assumed to present as a stationary RF process such that its statistical properties (i.e., the mean, standard deviation and scales of fluctuations) are constant over the entire field. By using RF theory, the site domain is

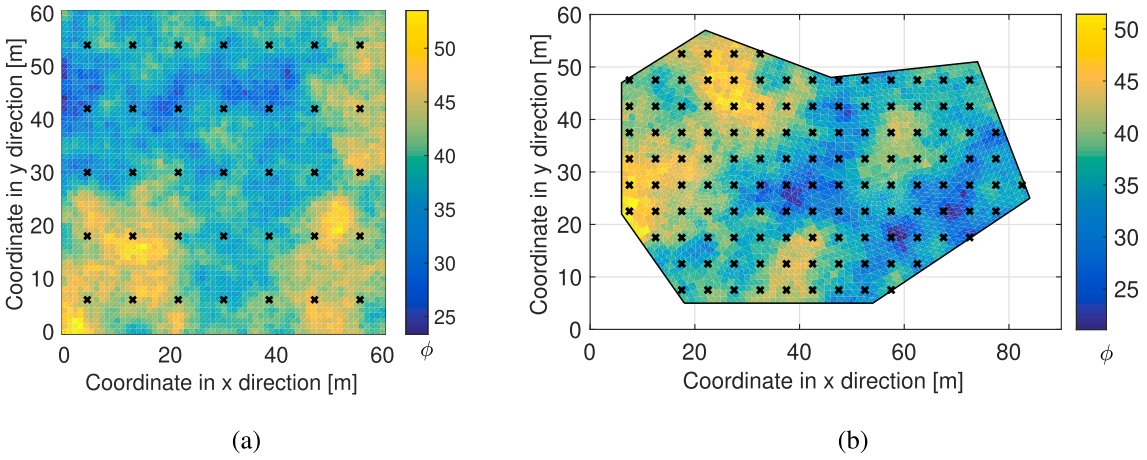


Fig. 1. Limitations of site investigation programs with structural grids: (a) an anisotropic site investigation program with a 5×7 structural grid and (b) a site with an irregular domain geometry. The background contour plot represents the soil friction angle ϕ (generated hypothetically based on RF theory), and the crosses indicate the investigation locations of a site investigation program.

discretized into a mesh so that the site property is approximated by a finite set of random variables. Each random variable associates with a mesh element and defines the site property within that element. For simplicity, the midpoint method, which was first introduced by Der Kiureghian and Ke [32], is adopted in this work for site discretization and approximation. In other words, the site property within a mesh element is assumed to be constant and equal to the site property at the centroid of that element.

Let $\vec{s}(\vec{x})$ denote the RF process of the site property of interest, where $\vec{x}$ represents the spatial coordinates and $\vec{s} = [s_1, s_2, ..., s_n]$ represents the site property at n mesh elements. Considering a two-dimensional case with a Gaussian distribution, the RF statistical parameters include the mean μ, standard deviation σ, scale of fluctuation in the x direction θ_x and scale of fluctuation in the y direction θ_y. The random variables $\vec{s}$ thus conform to a multivariate Gaussian distribution

$$\vec{s} \sim N(\mu, \vec{C}) \tag{1}$$

such that the joint probability distribution could be given as

$$f(\vec{s}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \vec{C}}} \exp\left[-\frac{1}{2}\left(\vec{s} - \mu\right)\vec{C}^{-1}(\vec{s} - \mu)^T\right] \tag{2}$$

where $\vec{C}$ represents the covariance matrix, in which each component is given as $C_{ij} = \rho_{ij}\sigma^2$, and ρ_{ij} represents the correlation between the variables i and j. A simple exponential correlation function, which is adopted in this work, could be given as

$$\rho(\tau_x, \tau_y) = \exp\left[-2\left(\frac{\tau_x}{\theta_x} + \frac{\tau_y}{\theta_y}\right)\right] \tag{3}$$

where τ_x and τ_y are the separation distances between the centroids of elements i and j projected onto the x and y directions, respectively.

2.2. Kriging interpolation

Given the values of a site property at certain locations (e.g., measured from prior site investigations), the mapping of this site property becomes a conditional random field (CRF). Kriging theory provides a way to estimate the joint probability distribution of the CRF. Denoting the values of the site property at measured locations $\vec{x}_p$ as known data $\vec{s}_p$, the site property values $\vec{s}_n$ at unmeasured locations $\vec{x}_n$ follow a multivariate Gaussian distribution given as

$$\vec{s}_n | \vec{s}_p \sim N(\vec{\mu}_{n|p}, \vec{C}_{n|p}) \tag{4}$$

in which

$$\vec{\mu}_{n|p} = \vec{\mu}_n + \vec{C}_{np}\vec{C}_{pp}^{-1}\left(\vec{s}_p - \vec{\mu}_p\right) \tag{5}$$

$$\vec{C}_{n|p} = \vec{C}_{nn} - \vec{C}_{np}\vec{C}_{pp}^{-1}\vec{C}_{pn} \tag{6}$$

where $\vec{C}$ represents the covariance matrix and the subscripts n and p indicate the indexes of measured and unmeasured locations, respectively. Here, the components of the covaraiance matrix $\vec{C}$ are calculated in the same way as those in Eq. (2). Then, the joint probability distribution is calculated as

$$f\left(\vec{s}_n \middle| \vec{s}_p\right) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \vec{C}_{n|p}}} \exp\left[-\frac{1}{2}(\vec{s}_n - \vec{\mu}_{n|p})\vec{C}_{n|p}^{-1}(\vec{s}_n - \vec{\mu}_{n|p})^T\right] \tag{7}$$

Directly sampling the realizations of the random variables $\vec{s}$ from the joint probability distribution given in Eq. (2) or Eq. (7) could be difficult if there is a large number of random variables. In this respect, some simple yet effective methods, such as the covariance matrix decomposition method or Hoffman method [33–35], can be used to generate such autocorrelated (and conditional) random variables. Another option is the Markov chain Monte Carlo (MCMC) simulation method [36,37,24], which will be briefly described in the following section.

2.3. Back-calculation of random field statistics

With prior data from site investigations, the RF statistics of the site property of interest are determined for a site characterization task. While the mean and standard deviation of a site property can be estimated using a moment-based approach [38,24,4], the scales of fluctuation are much more difficult to estimate. The scales of fluctuation are often characterized using empirical approaches. A simple and useful first approach is to estimate the semivariogram from the available data and optimally fit the estimated semivariogram with a theoretical semivariogram [39,4].

Due to the limited availability of information from previous site investigations, it is not always possible to obtain the actual values of the mean, standard deviation, and scales of fluctuation. The deviation between the estimated and actual values is called the statistical uncertainty. There are two schools of thoughts on how to model such statistical uncertainty [38]: frequentist thought and Bayesian thought. In the work of Ching et al. [38], it was indicated that Bayesian thought in general performs better than frequentist thought in terms of consistency. In particular, the MCMC method is recommended when the amount of information is very limited. In this work, Bayesian thought is adopted to estimate the RF statistics based on the data from previous site investigations.

Following the Bayesian inference framework, the probability of observing site property $\vec{s}_p$, denoted as $f\left(\vec{s}_p, \vec{\theta}\right)$, can be given by

$$f\left(\vec{s}_p \middle| \vec{\theta}\right) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \vec{C}_{pp}}} \exp\left[-\frac{1}{2}\left(\vec{s}_p - \mu\right)\vec{C}_{pp}^{-1}(\vec{s}_p - \mu)^T\right] \tag{8}$$

where $\vec{s}_p$ represents the observed values of random variables $\vec{s}$ at investigation locations $\vec{x}_p$ and $\vec{\theta}$ repents the RF statistics (i.e., $\vec{\theta} = [\mu, \sigma, \theta_x, \theta_y]$ in this case). Then, the posterior probability density function (PDF) of the statistics of the geotechnical property, denoted as $f\left(\vec{\theta} \middle| \vec{s}_p\right)$, can be computed following [24] as

$$f\left(\vec{\theta} \middle| \vec{s}_p\right) = kf\left(\vec{s}_p \middle| \vec{\theta}\right) f\left(\vec{\theta}\right) \tag{9}$$

where $f\left(\vec{\theta}\right)$ represents the prior PDF of the RF statistics and k represents a normalization factor that guarantees unity for the cumulative probability.

Additionally, realizations of statistics $\vec{\theta}$ could be sampled using a MCMC simulation Gilks et al. [36], Li et al. [37], Gong et al. [24]. The MCMC simulation procedure based on the Metropolis-Hastings algorithm is available from the aforementioned studies and is briefly presented below for completeness.

Step 1 Initialize the starting point of the Markov chain with arbitrary values of variable $\vec{\theta}$, which could be calculated from empirical approaches or estimated from empirical experience.

Step 2 Sample a candidate realization $\vec{\theta}^*$ from a jump function $f\left(\vec{\theta}^* \middle| \vec{\theta}\right)$. In this work, the jump function is assumed to have a multivariate normal distribution with a mean $\vec{\theta}$ and preset

covariance matrix $\vec{C}_{\theta}$.

Step 3 Calculate the ratio $\frac{f\left(\vec{\theta}^{*}\middle|\vec{s}_p\right)}{f\left(\vec{\theta}\middle|\vec{s}_p\right)}$, where $f\left(\vec{\theta}\middle|\vec{s}_p\right)$ and $f\left(\vec{\theta}^{*}\middle|\vec{s}_p\right)$ are the posterior PDFs of $\vec{\theta}$ and $\vec{\theta}^{*}$, respectively (see Eq. (9)).

Step 4 Sample an acceptance ratio α from a uniform distribution of U (0, 1).

Step 5 Determine whether the candidate realization $\vec{\theta}^{*}$ is acceptable: if $\min\left[\frac{f\left(\vec{\theta}^{*}\middle|\vec{d}\right)}{f\left(\vec{\theta}\middle|\vec{d}\right)}, 1\right] \geqslant \alpha$, then the candidate realization $\vec{\theta}^{*}$ is accepted and $\vec{\theta} = \vec{\theta}^{*}$; otherwise, the candidate realization $\vec{\theta}^{*}$ is rejected. Note that numerical truncation errors could exist for small $f\left(\vec{\theta}\middle|\vec{d}\right)$ or $f\left(\vec{\theta}^{*}\middle|\vec{d}\right)$. This situation could be improved by using a log transformation such that $\log\left(\frac{f\left(\vec{\theta}^{*}\middle|\vec{d}\right)}{f\left(\vec{\theta}\middle|\vec{d}\right)}\right) = \log f\left(\vec{\theta}^{*}\middle|\vec{d}\right) - \log f\left(\vec{\theta}\middle|\vec{d}\right)$ [24].

Step 6 Repeat steps 2–5 until a target number of realizations is obtained.

3. Site investigation programs

3.1. CVT and the modified Lloyd algorithm

Recognizing the limitations of structural grid-based site investigation programs, a novel site investigation scheme based on CVT is proposed in this work. A Voronoi tessellation (sometimes also called a Voronoi diagram) constitutes the partitioning of a plane into regions based on the distances to points in a specific subset of the plane [31]. Often, the points are called Voronoi seeds, and the regions are called Voronoi cells. Each Voronoi seed corresponds to a Voronoi cell, which consists of all points closer to that seed than to any other. CVT is a special type of Voronoi tessellation in which the seed of a Voronoi cell is also its centroid (i.e., the arithmetic mean or center of mass). CVT can change the topology of the mesh, leading to elements that are more nearly equilateral. A number of algorithms can be used to generate CVTs, including the Lloyd algorithm [40], probabilistic methods [41] or quasi-Newton methods [42]. Among these methods, the Lloyd algorithm is simple yet remains the most widely used and is thus adopted in this work. The procedures of the Lloyd algorithm are described as follows.

Step 1 Initialize n Voronoi seeds with random positions in the site domain.

Step 2 Compute the Voronoi diagram (i.e., the polygon vertexes of Voronoi cells) based on the current positions of the n seeds.

Step 3 Compute the centroid of each Voronoi cell.

Step 4 For each Voronoi cell, reposition the Voronoi seed according to its corresponding centroid.

Step 5 Repeat steps 2–4 until a convergence criterion is reached.

In reference to step 2 in which the Voronoi diagram is computed, algorithms such as the sweepline algorithm can be used [43]. There are also built-in functions in commercial software (e.g., *voronoi* in Matlab) or open-source packages (e.g., *scipy.spatial.Voronoi* in Python) that are available for such a task. Note that most algorithms (e.g., *voronoi* in Matlab) are developed to compute the Voronoi diagram for an infinite space (see the graph sketched in Fig. 2a). Herein, a simple modification is proposed for computing Voronoi diagrams that are bounded by a finite site domain. First, four auxiliary seeds that are far from the site domain are added to the original list of Voronoi seeds. For example, the points on the left, right, top and bottom sides of a domain centroid with a separation distance of five times the domain range in the x or y directions can be used as auxiliary seeds (see Fig. 2b). Then, a Voronoi diagram is computed based on the new list of Voronoi seeds for an infinite space. Next, for each Voronoi cell corresponding to the original list of Voronoi seeds, the polygon intersection between the Voronoi cell and the site domain is computed. Finally, a bounded Voronoi diagram (see Fig. 2c) is obtained by replacing the Voronoi cells with the corresponding polygon intersections.

Regarding the convergence criterion, this work adopts the condition that is triggered when the summation of all distances between the seeds and centroids normalized by that of the initial configuration falls below a preset threshold (e.g., 1e-8 in this work). It should be noted that the Lloyd algorithm may not be particularly efficient in terms of convergence, although it is simple to implement. Both the number of seeds and the site geometry could have an effect on the convergence rate of the Lloyd algorithm. For CVTs with a large number of Voronoi seeds, the Lloyd algorithm may converge slowly. One option to accelerate the convergence is to over-relax the seeds by moving them ω times the distance to their corresponding centroid. Typically, a value that is slightly less than 2 could be used for ω.

3.2. Robustness of a site investigation

To evaluate the effectiveness of a site investigation program, some metrics that can quantify the site characterization accuracy are required. Unfortunately, the geotechnical community has not commonly accepted a quantitative, consistent, spatially sensitive method for such purposes [18]. One option, as proposed by Lloret-Cabot et al. [44], is to evaluate the reduction in the spatial uncertainty relative to the original (unconditional) RF by comparing the input variance and the estimation (i.e., Kriging) variance, both as a function of position. Another similar proposal is to use the prediction variance reduction factor (PVRF) [12]. The PVRF is calculated by first obtaining the average prediction variance across the site and then evaluating the change in this average variance upon drilling and sampling at a new borehole location. To consider both the variance and the mean of the RF, Gong et al. [24] adopted the signal-to-noise ratio (SNR) of the RF statistics to evaluate the site investigation robustness and further applied it to the optimization of site investigation programs.

This work also adopts the SNR to quantify the site characterization uncertainty and to evaluate the robustness of site investigation programs. For a particular random variable, the SNR is defined as [45,24]

$$SNR = 10\log_{10}\left(\frac{\mu_v^2}{\sigma_v^2}\right) \tag{10}$$

where μ_v and σ_v represent the mean and standard deviation of this random variable, respectively, and the subscript v is used to indicate any arbitrary random variable v. Herein, two types of robustness are considered: 1) the robustness of the characterized RF statistics and 2) the robustness of the characterized site property at every location within the site (hereafter referred to as the point site property). The robustness of RF statistics is calculated as

$$SNR_1 = w_\mu SNR_\mu + w_\sigma SNR_\sigma + w_{\theta_x} SNR_{\theta_x} + w_{\theta_y} SNR_{\theta_y} \tag{11}$$

in which

$$SNR_\mu = 10\log_{10}\left(\frac{\mu_\mu^2}{\sigma_\mu^2}\right) \tag{12}$$

$$SNR_\sigma = 10\log_{10}\left(\frac{\mu_\sigma^2}{\sigma_\sigma^2}\right) \tag{13}$$

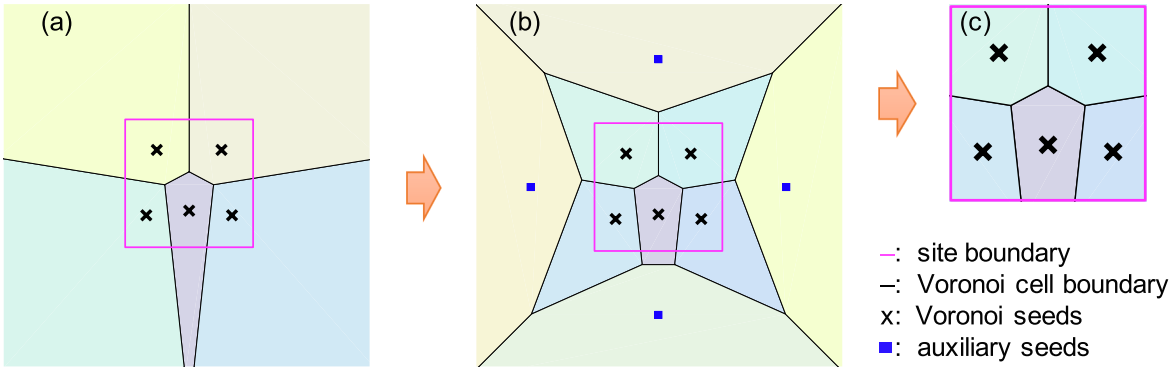


Fig. 2. Usage of auxiliary seeds to generate bounded Voronoi tessellations in the Lloyd algorithm: (a) Voronoi tessellation without auxiliary seeds, (b) Voronoi tessellation with auxiliary seeds, and (c) the final bounded Voronoi tessellation.

$$SNR_{\theta_x} = 10\log_{10}\left(\frac{\mu^2_{\theta_x}}{\sigma^2_{\theta_x}}\right) \tag{14}$$

$$SNR_{\theta_y} = 10\log_{10}\left(\frac{\mu^2_{\theta_y}}{\sigma^2_{\theta_y}}\right) \tag{15}$$

where the subscripts μ, σ, θ_x, and θ_y represent the mean, variable, scale of fluctuation in the x direction, and scale of fluctuation in the y direction, respectively, and w_μ, w_σ, w_{θ_x}, and w_{θ_y} are the corresponding weights. In this work, all these weights are taken as unity for simplicity. The robustness of the point site property is calculated as

$$SNR_2 = \frac{1}{n}\sum_{i=0}^{n} SNR_{s_i} = \frac{1}{n}\sum_{i=0}^{n} 10\log_{10}\left(\frac{\mu^2_{n|p,i}}{\sigma^2_{n|p,i}}\right) \tag{16}$$

where $\mu_{n|p,i}$ and $\sigma_{n|p,i}$ are the mean (i.e., Eq. (5)) and standard deviation (i.e., Eq. (6)), respectively, of the characterized site property at every site location $\vec{x}_i$ based on Kriging interpolation, n is the number of mesh elements in the RF discretization, and i indicates the element index.

4. Illustrative examples

In this section, the performance of the CVT-based site investigation programs will be evaluated using two illustrative examples, including a square site and an irregular site. In both examples, the friction angle is characterized, and the hypothetical fields of the friction angle generated by RF theory are used for illustration purposes. To generate the hypothetical fields of the friction angle, the sites are first discretized into quadrilateral meshes. Then, the friction angles of every mesh element are determined using midpoint approximation [32] and Cholesky decomposition [33]. The RF statistical parameters used to generate the hypothetical fileds of the friction angle are as follows: mean $\mu = 35^\circ$, standard deviation $\sigma = 5.25^\circ$, and scales of fluctuation in the x and y directions θ_x, $\theta_y = 25$ m, respectively. First, the detailed CVT, MCMC simulation, site investigation and optimization results will be presented using the square site example. Then, similar analyses are conducted on the irregular site example to address the applicability and performance of the CVT-based site investigation scheme for a site with an irregular domain geometry.

4.1. CVT and Lloyd algorithm results

The square site example involves a square geometry with a width of 60 m. The aim of site investigation is to characterize the friction angle of the soil at this site. First, the CVT and Lloyd algorithm results are presented. Fig. (3) shows three CVTs that are computed from three different initial configurations with 15 seeds. For each initial configuration, the 15 seeds are randomly placed throughout the site domain. Three primary observations can be made based on these results. First, a CVT with a certain number of seeds is not necessarily unique; there are multiple CVT solutions whose layouts basically depend on the initial configuration of the seeds. Second, two different initial configurations may lead to the same CVT, as indicated by the first two CVTs in Fig. 3. Finally, although there are multiple CVT solutions, the centroids of the final CVTs all form a rather systematically ordered distribution within the site domain. As will be illustrated later, systematically ordered site investigation programs could improve the site characterization accuracy.

Another point of interest for the CVT and Lloyd algorithm is the convergence profile. In this regard, the residuals (i.e., the summations of all distances between seeds and their corresponding centroids normalized by the distance at the initial stage) as a function of the iteration number are plotted in Fig. 4. The residuals are obtained from the three CVT calculation trials with 15 seeds discussed above, as already shown in Fig. 3. At the very beginning, the Voronoi tessellations experience great changes (e.g., some edges of a Voronoi cell could vanish, and new edges could be generated). Thus, the residual-iteration plot exhibits some fluctuations at this very beginning stage, after which the Lloyd algorithm roughly follows a linear convergence trend, and the layouts of the Voronoi tessellations are gradually finalized.

To obtain insight into the effects of the seed number on the CVT and Lloyd algorithm, the results of one example CVT with 100 seeds and the corresponding convergence profile are presented in Fig. 5. The 100 centroids of this CVT do not precisely form a 10×10 structural grid, which is another CVT solution. This is consistent with the fact that there are multiple CVT solutions for a certain seed number. For the convergence profile in the residual-iteration plot, the case with 100 seeds also presents a linear trend towards convergence.

Finally, the convergence rates of the Lloyd algorithm for different numbers of seeds are analyzed. For this purpose, the slope of the residual-iteration plot in the semilog diagram is calculated and defined as the convergence rate. A series of CVTs is generated with different numbers of seeds ranging from 10 to 100. The convergence rate as a function of the number of seeds is depicted in Fig. 6. The convergence rate deceases with an increasing number of seeds. The convergence rate and the number of seeds roughly follow an exponential relation, and the number of iterations required to compute a CVT with a given number of seeds can be estimated based on this fitted empirical relation. For example, the convergence rate of the Lloyd algorithm for 40 seeds is approximately 0.01, which requires approximately 800 (i.e., 8/0.01) iterations to achieve a residual of $1e-8$. It should be noted, however, that the exact number of iterations required to generate a CVT is dependent on the initial configuration and is not a unique value.

4.2. Results of the MCMC simulation

With the data obtained from a site investigation, the RF statistics can be estimated using Bayesian inference (i.e., Eq. (9)) and an MCMC

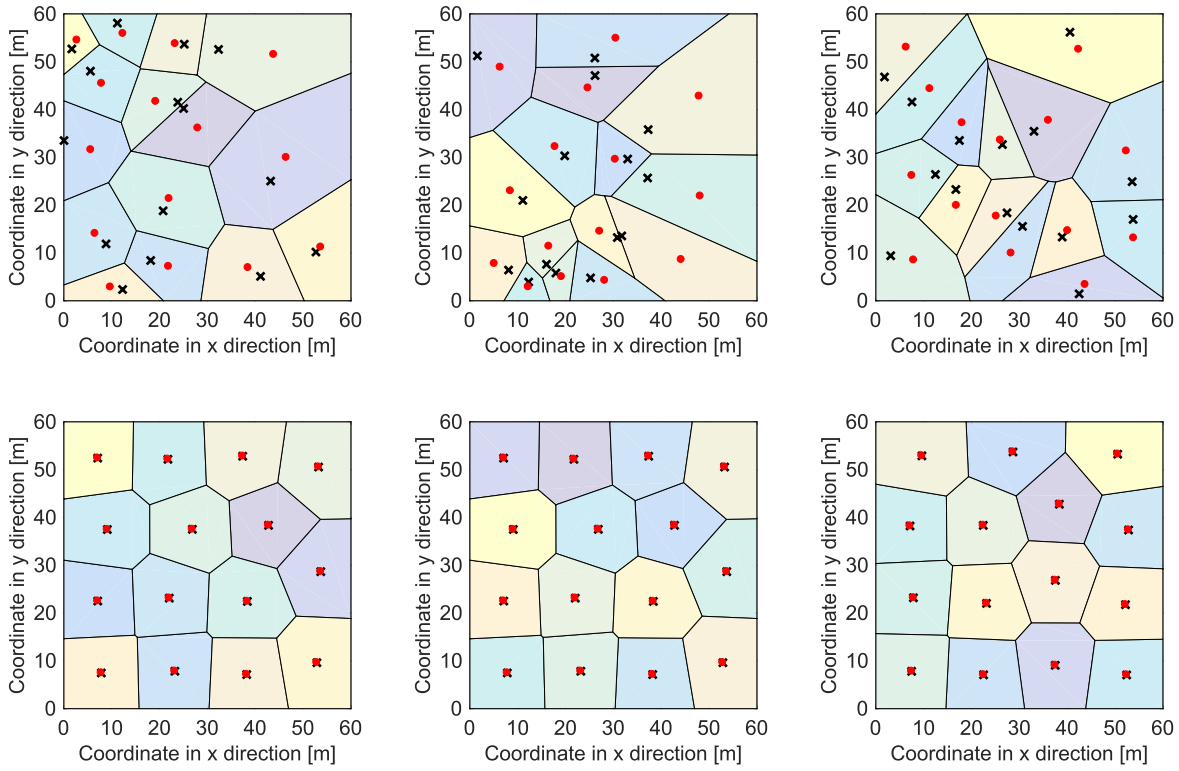


Fig. 3. CVT computed from three different random initial configurations with 15 seeds: the top row presents the initial configurations, while the bottom row presents the corresponding CVTs. The black crosses indicate the Voronoi seeds, and the red dots indicate the centroids. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

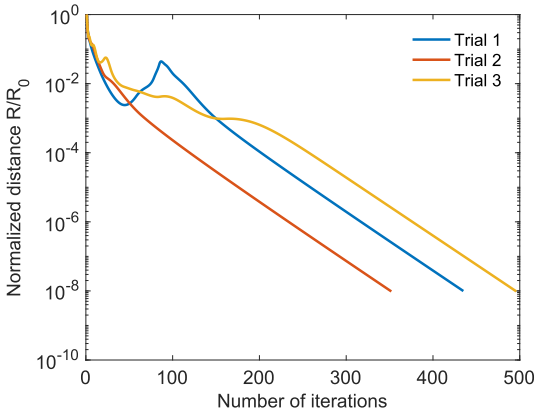


Fig. 4. Convergence profiles of the Lloyd algorithm for the CVTs with 15 seeds.

simulation. In this work, the prior PDF of the RF mean is assumed to follow a normal distribution, and the prior PDFs of the other RF statistical parameters (i.e., the standard deviation and the scales of fluctuation in the x and y directions) are assumed to follow lognormal distributions. However, in a practical implementation, a logarithmic form for these RF statistical parameters is used. The means of the prior PDFs for the RF mean and RF standard deviation are specified as

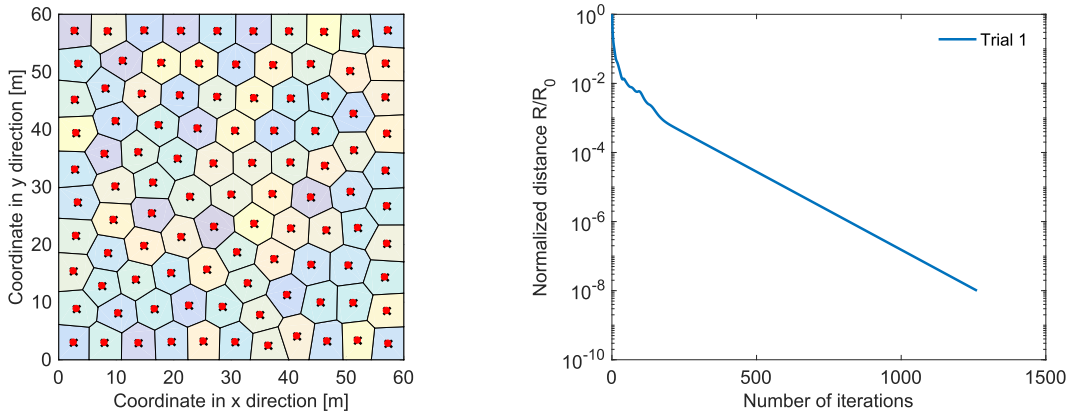
$$\mu_{\mu} = \frac{1}{n_p}\sum_{i=1}^{n_p}\phi_i \tag{17}$$

$$\mu_{\ln\sigma} = \ln\left(\frac{1}{n_p - 1}\sum_{i=1}^{n_p}(\phi_i - \mu_{\mu})^2\right) \tag{18}$$

where μ_{μ} and $\mu_{\ln\sigma}$ represent the means of the prior PDFs for the RF mean μ and RF standard deviation $\ln\sigma$, respectively, ϕ is the friction angle of the investigated site location, and n_p represents the number of investigated locations. For the scales of fluctuation in the x and y directions, the means of their prior PDFs are taken as half of the site dimension (i.e., 30 m), as suggested by [24]. The standard deviations of the prior PDFs for the RF statistics are calculated as their corresponding means multiplied by a scale of 0.1. The settings of the Bayesian inference and MCMC simulation are summarized in Table 1. Note that some studies have also used uniform distributions for the prior PDFs of RF statistics [38]. As suggested by [24], the prior PDFs of RF statistics could be determined based on empirical knowledge or experience obtained from the particular site of interest. The specific settings used in this work are adopted for illustration purposes.

The jump functions in the MCMC simulation take the same forms as their corresponding prior PDFs. First, the results of an MCMC simulation for the sampling of RF statistics are presented (Fig. 7). In these figures, 10,000 samples are accepted from the MCMC simulation. The accepted samples are widely spread throughout the sampling space, indicating that the MCMC sampling scheme is rather effective. The MCMC requires an initial burn-in stage to stabilize. Herein, the first 5,000 samples are considered burn-in samples and are discarded. In addition, for the later 5,000 samples, one sample is extracted from every 5 samples as real samples of the RF statistics following a bootstrap sampling scheme [46,38]. Eventually, 1,000 samples of the RF statistics are obtained from the MCMC simulation process.

Histograms of the RF statistic samples are shown in Fig. 8. The results show that the samples of the RF mean can be effectively fitted by a normal distribution, and the samples of the RF standard deviation and scales of fluctuation can be effectively fitted by lognormal distributions. The distribution types coincide with the prior specified distribution types, while the variances of these RF statistical parameters are reduced as a result of the posterior constraints from known site investigation data. It should be noted that the reduction in the variance of the scales of fluctuation is not as notable as those of the RF mean or RF standard deviation. This phenomenon is consistent with the findings in the literature that the scales of fluctuation are more difficult to characterize



(a) CVT (b) Convergence profile

Fig. 5. Results of the Lloyd algorithm-based CVT with 100 seeds: (a) the final CVT and (b) the convergence profile.

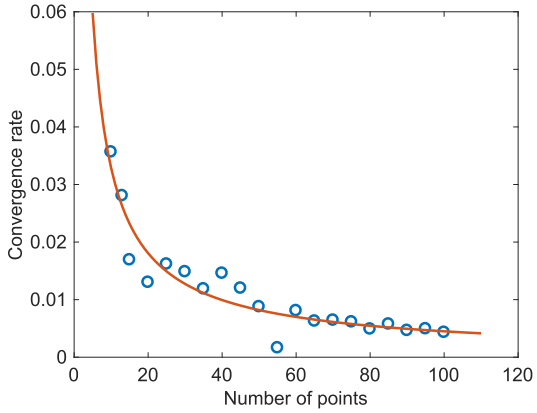


Fig. 6. Relationship between the convergence rate (in terms of the slope of the convergence profile) and the number of seeds for the Lloyd algorithm.

Table 1
Settings of the Bayesian inference and MCMC simulation.

	Prior PDF model	Mean of prior PDF	Standard deviation of prior PDF
μ	Normal	Eq. (17)	Mean of prior PDF × 0.1
$\ln\sigma$	Normal	Eq. (18)	Mean of prior PDF × 0.1
$\ln\theta_x$	Normal	ln30	Mean of prior PDF × 0.1
$\ln\theta_y$	Normal	ln30	Mean of prior PDF × 0.1

than are the mean or standard deviation [22,38]. Nevertheless, the reduction in the variance will increase if additional data are available from site investigation.

4.3. Results of the site investigation and characterization

In this section, the performance of each CVT-based site investigation program is analyzed. For comparison purposes, three site investigation schemes are considered, including (1) a random layout, (2) a structural grid, and (3) a CVT. For the random layout, the locations to be investigated are randomly placed throughout the site domain. The structural grid adopts an $a \times b$ layout, with a and b being the number of investigation locations in the x and y directions, respectively. Examples of the three different site investigation schemes with 15 investigation locations are displayed in Fig. 9. The background contour map represents the hypothetical field of the friction angle generated by RF theory.

First, the robustness of the site investigation in terms of the characterized RF statistics is analyzed. Fig. 10 shows the SNR plots of the RF mean, standard deviation, and scales of fluctuation. All SNR plots gradually stabilize within 1,000 samples. For all three site investigation schemes, the characterized RF mean presents the highest robustness (i.e., the largest SNR), followed by the standard deviation. In contrast, the scales of fluctuation in the x and y directions similarly present the lowest robustness. There is no notable difference among the robustness of the three different site investigation schemes, while the robustness of the site investigation schemes with either a structural grid or a CVT is slightly higher than that with a random layout. These results indicate that the robustness (in terms of the characterized RF statistics) of a site investigation program may not be sufficiently sensitive to the site investigation scheme. Instead, other metrics (e.g., the SNR of the point site property proposed in this work) are necessary to provide a more thorough evaluation of the site investigation robustness.

Next, the robustness of the site investigation in terms of the characterized point site property is analyzed. Fig. 11 shows the SNR contours of the point site property and the corresponding histograms. In these contours, it is observed that the areas surrounding the site investigation locations present high robustness. This is expected because these locations should have less variance (i.e., uncertainty) due to the strong correlation with the site investigation locations. For the case of a random layout, the bottom right corner of the contour plot presents a much lower robustness (see Fig. 11a) than the other contour plots due to the sparseness of investigated locations. This phenomenon can be clearly observed in the histogram of the point site property SNR, which displays a considerably large portion of positions with an SNR smaller than 18 (see Fig. 11d). For the cases with a structural grid and CVT, similar contours of the robustness of the point site property are observed (see Fig. 11b and c). A detailed investigation of these histograms indicates that the CVT-based site investigation program results in a more concentrated robustness distribution (the histogram profile in Fig. 11f is narrower than that in Fig. 11e). These results also indicate that the SNR of the point site property can reflect the local characterization robustness of a site property better than the SNRs of the RF statistics.

To more quantitatively investigate the performance of these three site investigation schemes, the overall SNR of the RF statistics and the SNR of the point site property for each scheme are summarized in Table 2. The structural grid and CVT schemes exhibit similar robustness in terms of the RF statistics and outperform the random layout scheme. In addition, the CVT scheme exhibits the best performance in terms of the mean and standard deviation among all point site property SNRs. These findings are consistent with the claims in the literature that site investigation programs with systematically ordered investigation locations perform better overall than those without systematically ordered investigation locations [30,23,25].

Because there are multiple CVT solutions for a certain number of

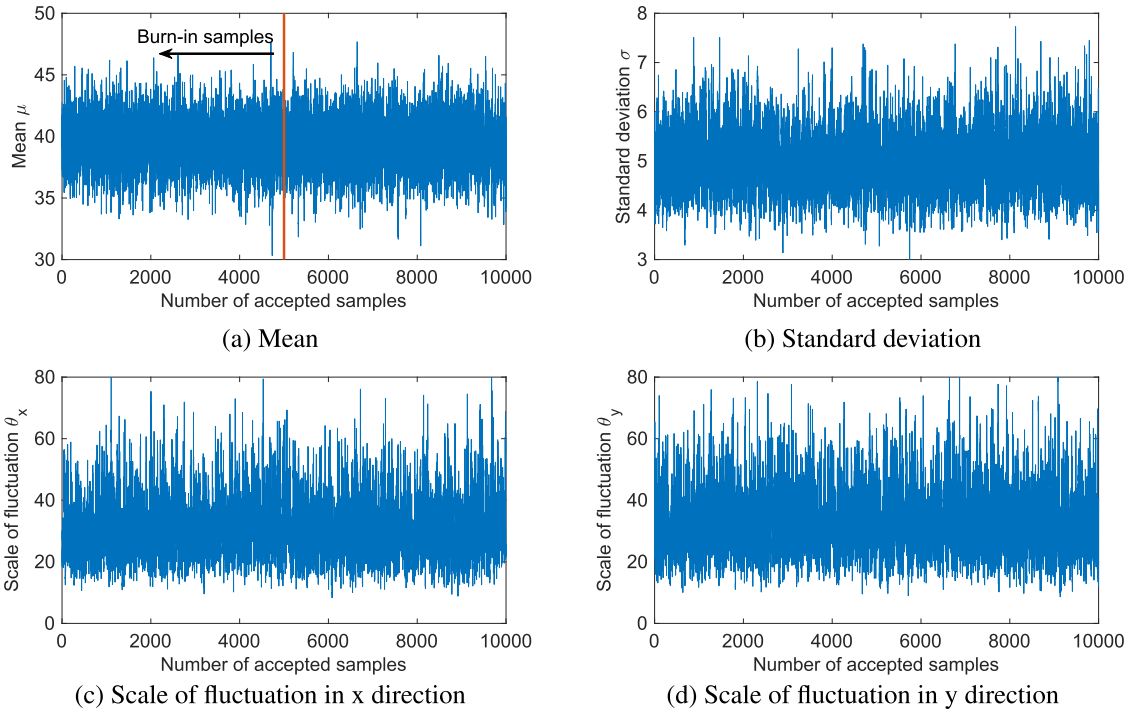


Fig. 7. Results of the MCMC simulation for the sampling of RF statistics: (a) mean, (b) standard deviation, (c) scale of fluctuation in the x direction, and (d) scale of fluctuation in the y direction.

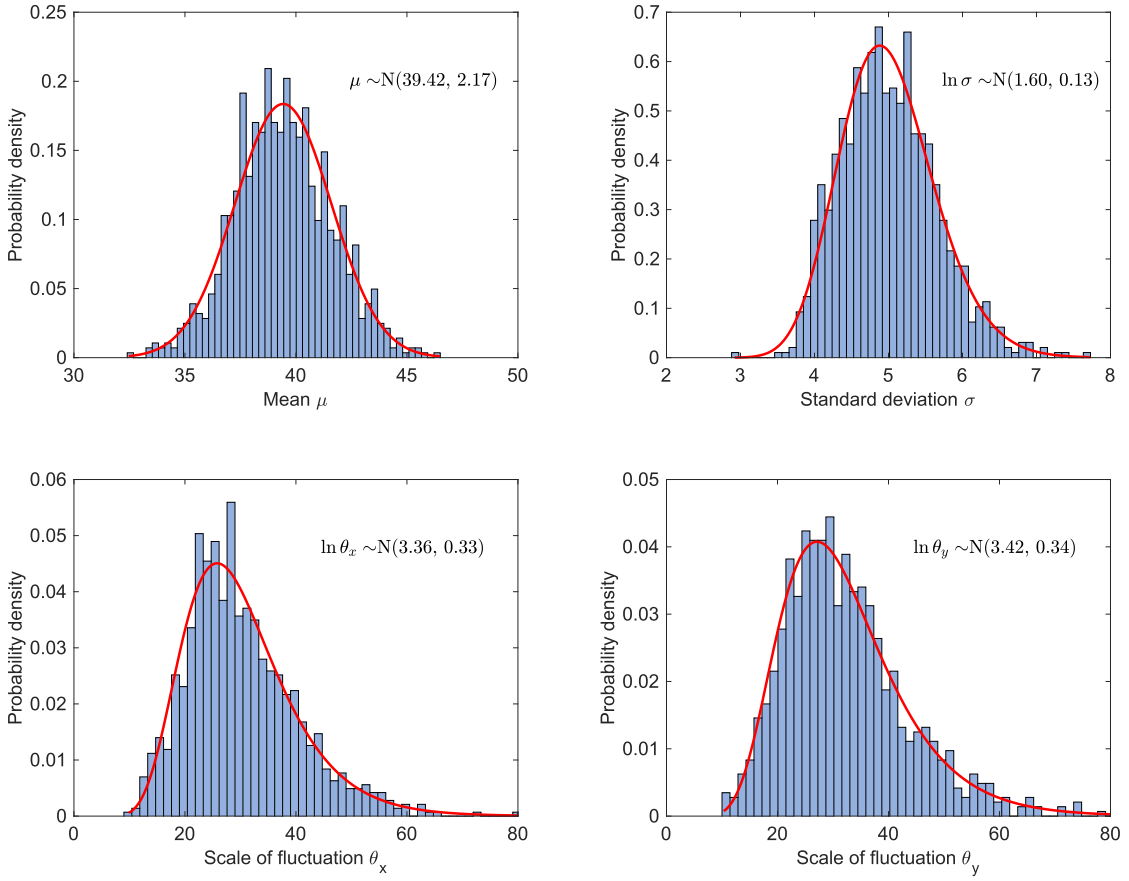


Fig. 8. Histograms of the RF statistic samples obtained from the MCMC simulation: (a) mean, (b) standard deviation, (c) scale of fluctuation in the x direction, and (d) scale of fluctuation in the y direction. The red curves represent the fitted normal (for the mean) or lognormal (for the other three statistics) distributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

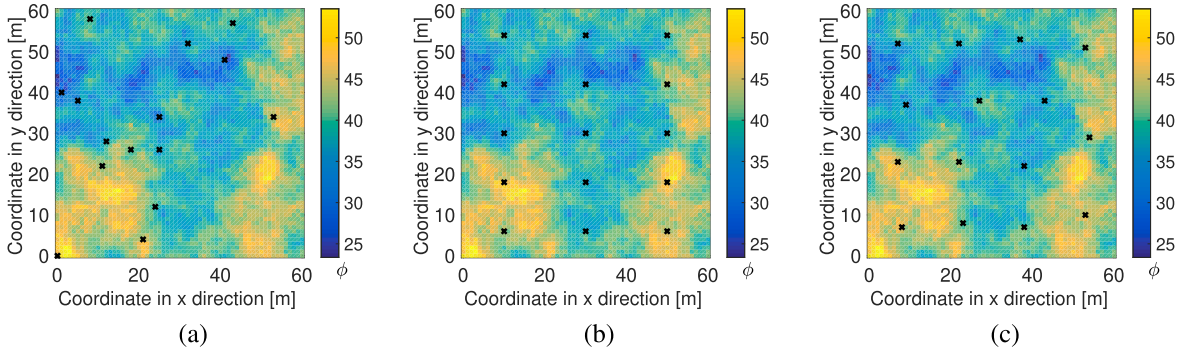


Fig. 9. Three different site investigation schemes: (a) random layout, (b) structural grid, and (c) CVT.

seeds, different CVT-based site investigation programs may present a different robustness. In this regard, the CVTs with 15 seeds and 100 random initial configurations are solved, and the corresponding site characterization results are analyzed. The SNR of the RF statistics and the SNR mean and SNR standard deviation of the point site property are plotted in Fig. 12. In this figure, although 100 random initial configurations are considered, there are only 14 different solutions because some random initial configurations converge to the same CVT solution. The SNR of the RF statistics ranges from 60 to 63. The SNR mean of the point site property ranges from 23 to 27, while the SNR standard deviation of the point site property is much more stable and fluctuates around 2.85. The SNR mean of the point site property presents a positive linear correlation with the robustness of the RF statistics, while the standard deviation presents no notable correlation. These results indicate that the robustness, especially that for the SNR standard deviation of the point site property, of a CVT-based site investigation is not considerably sensitive to the CVT example layout. For the following analyses, only one example CVT with random initial seed configurations will be used.

4.4. Site investigation program optimization

With the performance of different site investigation programs being evaluated, the optimization of site investigation programs can be conducted. To this end, Gong et al. [24] developed a biobjective optimization framework for improving site investigation programs by considering the SNRs of characterized RF statistics and the number of site investigation locations. In this work, the same biobjective optimization framework is adopted, but CVT-based site investigation programs are used instead of structural grids. In addition, two types of robustness are considered. The CVT-based site investigation programs with different numbers of investigation locations ranging from 10 to 300 are considered, and their robustness in terms of the SNR of the RF statistics and the SNR of the point site property are calculated. Fig. 13 plots the SNR of the RF statistics for the CVT-based site investigation programs with different numbers of seeds. When the number of investigation locations changes from 10 to approximately 150, the SNR of the RF statistics increases rapidly (i.e., from approximately 61 to approximately 71). However, it only increases from 71 to approximately 73 when the number of investigation locations changes from 150 to 300. These results indicate that it becomes inefficient to further increase the number of investigation locations when it reaches a certain threshold (hereafter referred to as the knee point).

Similarly, the SNR mean and standard deviation of the point site property for the CVT-based site investigation programs with different numbers of seeds are plotted in Fig. 14. The SNR mean increases with an increasing number of investigation locations, whereas the SNR standard deviation decreases with an increasing number of investigation locations and reaches a plateau when the number of investigation locations reaches approximately 150, indicating that the SNR standard deviation cannot be further reduced simply by increasing the number of investigation locations. Based on these results, the optimum site investigation programs can be determined according to the knee point method [47,22]. In this example, the CVT-based site investigation program with approximately 150 investigation locations is regarded as the optimum design.

4.5. Application to a site with an irregular domain geometry

To demonstrate the performance of the CVT-based site investigation programs on a site with an irregular domain geometry, the irregular site example shown in Fig. 1b is analyzed in this section. Similar to the square site example, this irregular site example also considers the friction angle and adopts a hypothetical field of the friction angle for illustration purposes. The settings of the Lloyd algorithm and MCMC simulation are also kept the same as those described for the square site example.

First, some basic results regarding the CVT, SNR of the RF statistics, and SNR of the point site property are presented for an example involving a CVT-based site investigation program with 100 investigation locations. The CVT-based site investigation program with 100 investigation locations is displayed in Fig. 15a. Also shown in Fig. 15a is a

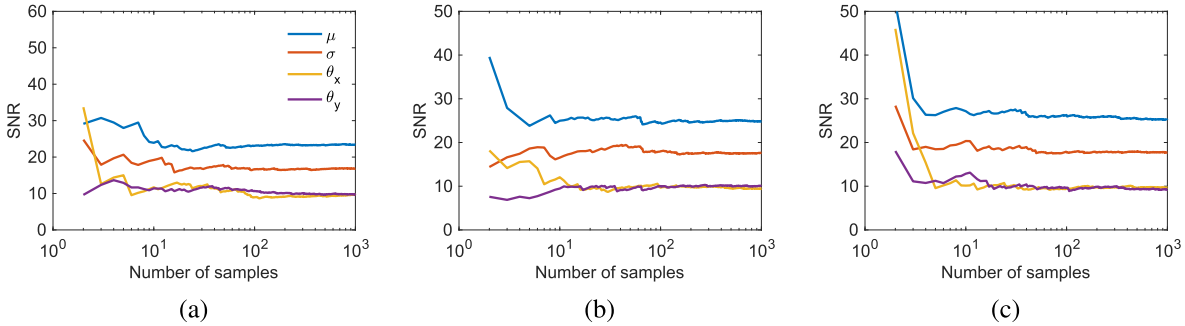


Fig. 10. Robustness of the RF statistics for the three different site investigation schemes: (a) random layout, (b) structural grid, and (c) CVT.

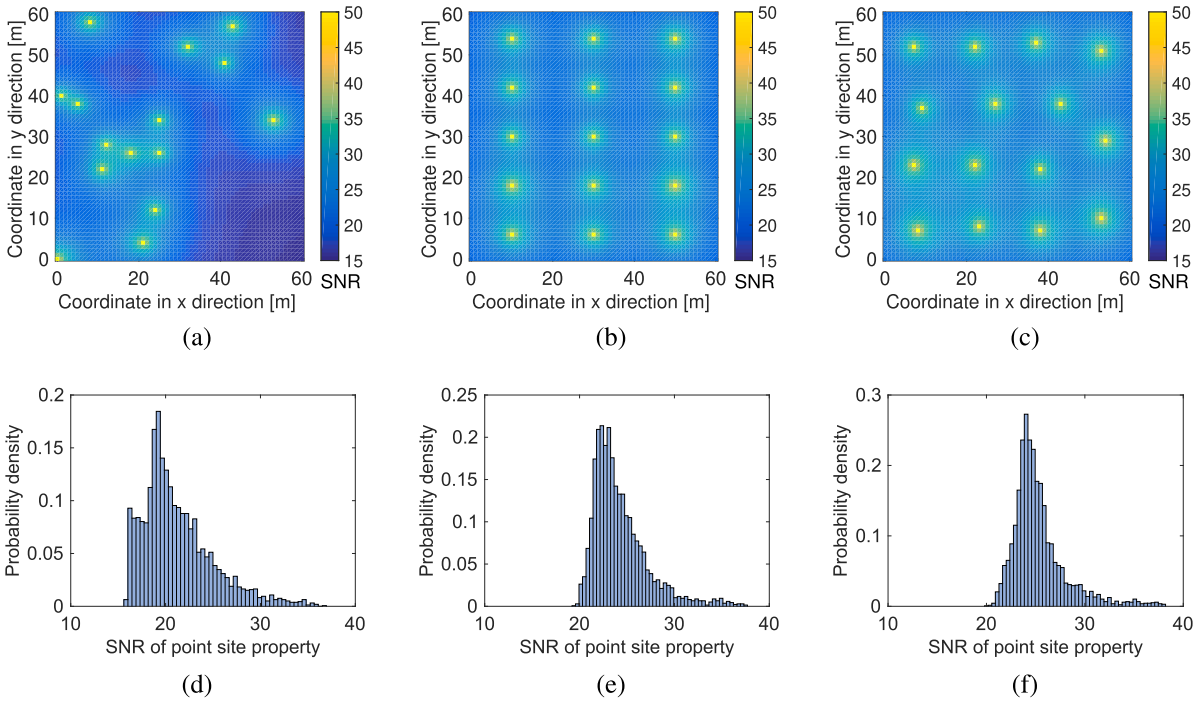


Fig. 11. Robustness of the point site property for the three different site investigation schemes: (a, d) random layout, (b, e) structural grid, and (c, f) CVT. The top row presents the SNR contours of the characterized site property at each location, and the bottom row presents the corresponding histograms.

Table 2
Robustness of the different site investigation schemes in terms of the overall SNR of the RF statistics and the SNR of the point site property. $a \pm b$ indicates the *mean* $\pm$ *standard deviation* of the SNR of the point site property.

Site investigation program	SNR of the RF statistics	SNR of the point site property
Random	59.96	21.26 ± 3.72
Structural grid	61.28	24.06 ± 2.89
CVT	61.92	25.38 ± 2.75

contour visualization of the hypothetical field of the friction angle. It can be observed that the 100 investigation locations are scattered throughout the site domain in a spatially ordered, systematic order; i.e., the density of investigation locations is uniform everywhere across the site domain. Even the concave corners of the site can be effectively accommodated by the Lloyd algorithm and the CVT-based site investigation scheme. With this CVT-based site investigation program, the friction angles at the investigated locations are obtained. Then, the RF

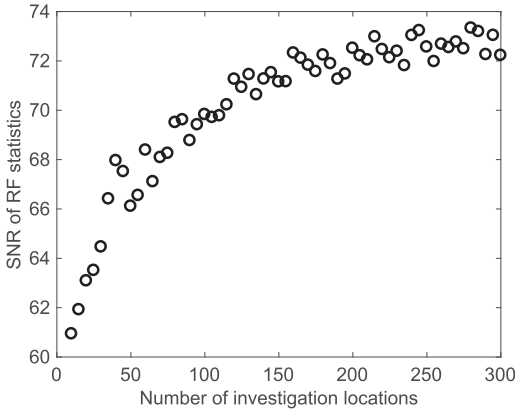


Fig. 13. SNR of the RF statistics for the CVT-based site investigation programs with different numbers of seeds.

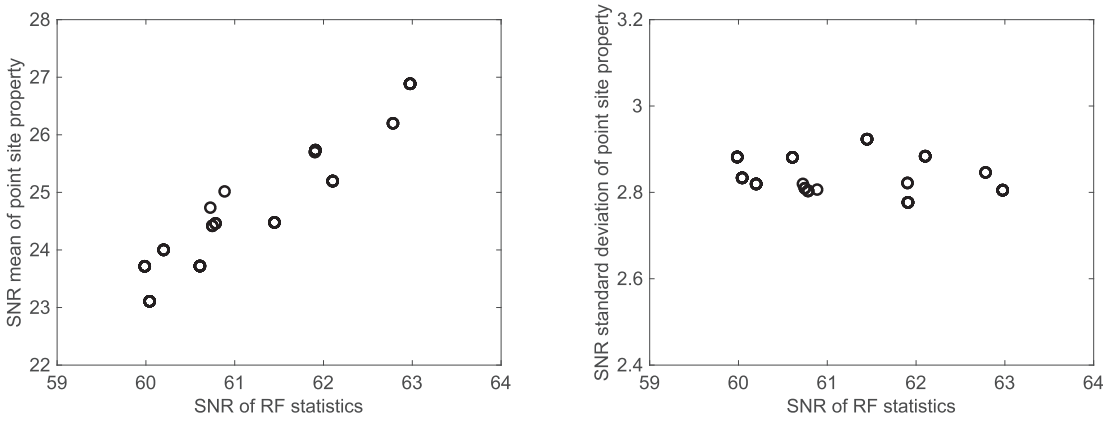


Fig. 12. Results of the investigation robustness of the CVT-based investigation programs with 100 different initial configurations. Note that the 100 different initial configurations only result in 14 different solutions (i.e., the 14 points in these two plots).

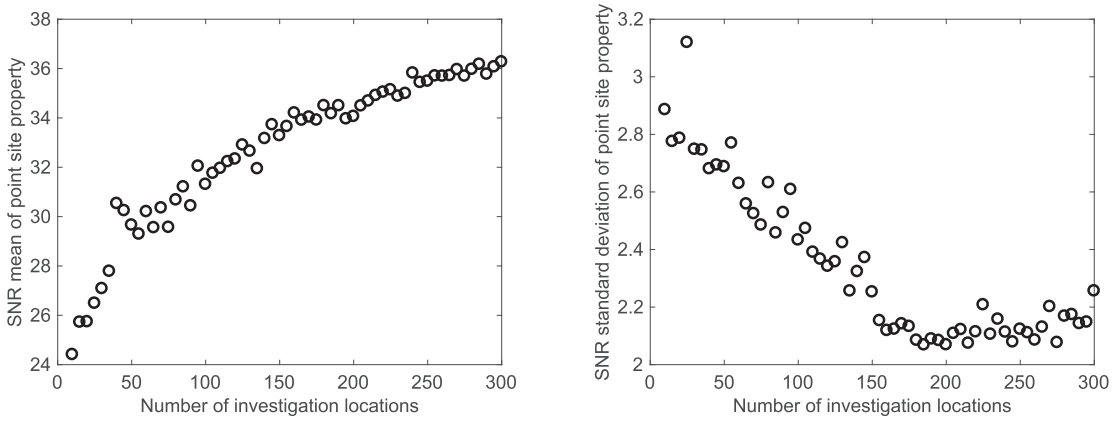


Fig. 14. Robustness of the point site property for the CVT-based site investigation programs with different numbers of seeds: (a) mean and (b) standard deviation.

statistical parameters can be characterized based on Bayesian inference (i.e., Eq. (9)) and an MCMC simulation. The SNR of the RF statistics is plotted in Fig. 15b. Similar to the square site example, the SNR of the RF statistics becomes stable when the number of MCMC samples exceeds approximately 100 (see the plateaus of the plots in Fig. 15b). The RF mean presents the highest SNR (i.e., the lowest uncertainty), followed by the RF standard deviation. Again, the scales of fluctuation in the x and y directions similarly present the lowest SNR.

A contour map of the point site property SNR at every location is displayed in Fig. 15c. Again, the areas surrounding the investigation locations present high SNRs (i.e., low uncertainty). A histogram of all the point site property SNRs is presented in Fig. 15d. The SNR of the point site property varies from 25 to 40 with a mean value of approximately 32. As the friction angle field is generated hypothetically based on RF theory, the SNR of the point site property at each location for the case without any prior site investigation data is theoretically calculated to be approximately 16.5 (i.e. $10\log_{10}(35^2/5.25^2)$). The effect of a prior site investigation can be clearly identified by comparing the SNR of the point site property before a site investigation (i.e., theoretically calculated as 16.5) with that after a site investigation (e.g., 32).

Then, the site investigation program is optimized. Again, a series of CVT-based site investigation programs with numbers of investigation

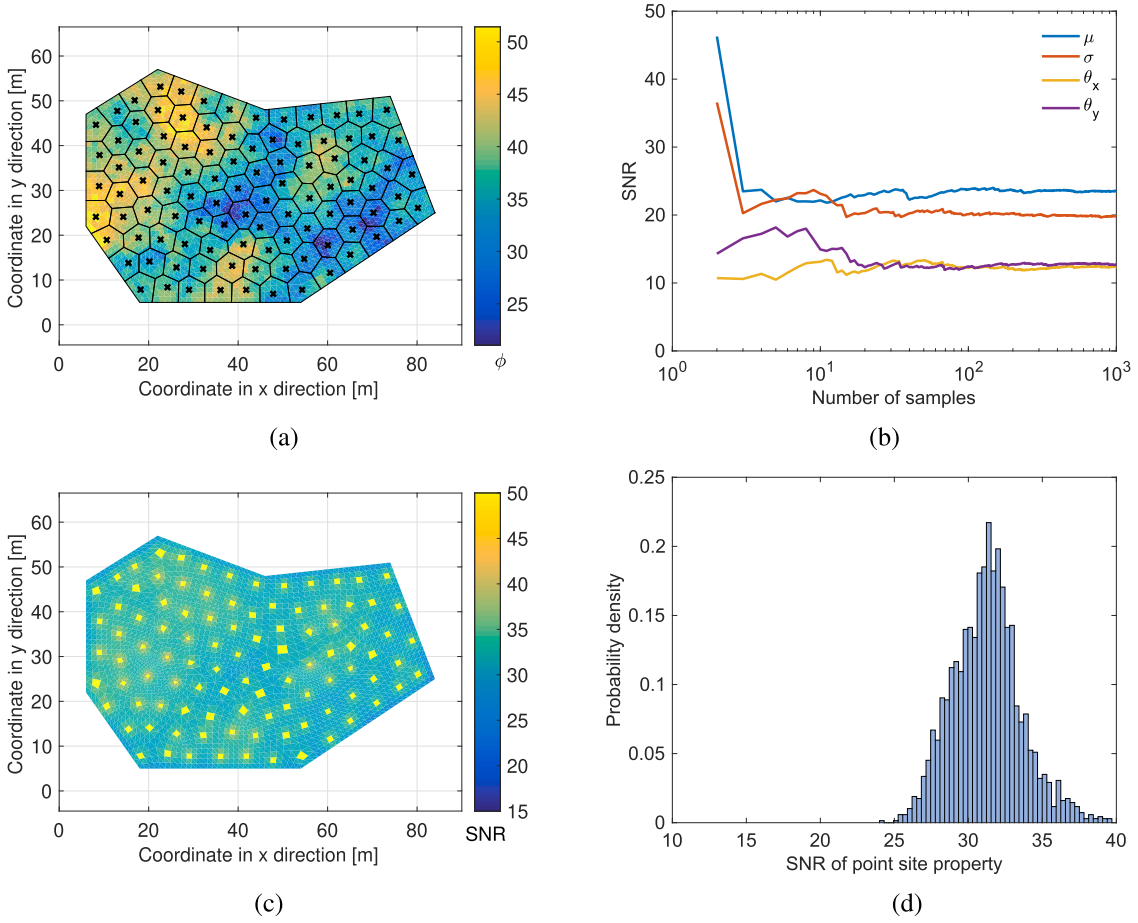


Fig. 15. Results of the CVT-based site investigation and characterization for a site with an irregular domain geometry: (a) a CVT-based site investigation program with 100 investigation locations, (b) the characterization robustness of the RF statistics, (c) the characterization robustness of the point site property, and (d) a histogram of the robustness of the point site property.

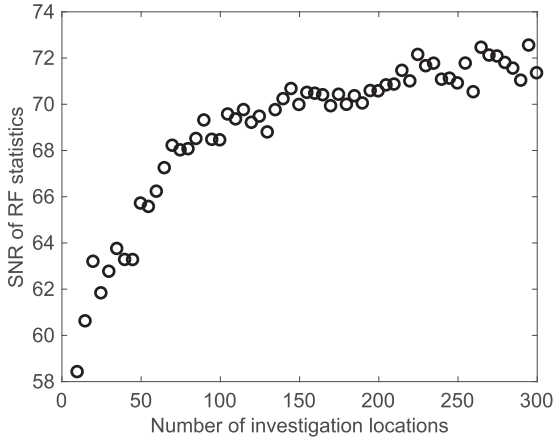


Fig. 16. Robustness of the RF statistics for the CVT-based site investigation programs with different numbers of investigation locations.

locations ranging from 10 to 300 is considered, and the robustness of each program is calculated. The results of the SNR of the RF statistics are presented in Fig. 16. Similar to the square site example, the SNR of the RF statistics presents a positive correlation with the number of instigation locations. The rate of increase is rather high when the number of investigation locations is less than approximately 120 and then reduces significantly after that (i.e., when the number of investigation locations varies from 120 to 300).

Finally, the results of the point site property SNR mean and standard deviation are shown in Fig. 17. The SNR mean presents a positive correlation with the number of investigation locations, while the SNR standard deviation presents a negative correlation. Similar to the SNR of the RF statistics, as well as the results of the square site example, there exists an knee point after which the change in the SNR results becomes relatively mild. For this irregular site example, the CVT-based site investigation program with approximately 120 investigation locations is adopted as the optimum program characterized by a compromise between the investigation robustness and investigation effort.

5. Discussion

As the first effort of a new CVT-based site investigation scheme, the focus of this work is to develop the methodology and demonstrate its potential for arbitrary numbers of investigation locations and irregularly shaped sites. There are several limitations in this work and need for further study. First, in practice, usually the three-dimensional spatial distribution of a soil property is required to be characterized for a site investigation project. This fact alone causes that the optimization of site investigation programs concerns not only the number and spacing of locations, but also the depth of investigation. While this aspect is not considered in this work, the current two-dimensional CVT-based site investigation schemes and optimization framework is applicable to the two-dimensional problems, such as the characterization of soil liquefaction potential, compaction effectiveness, soil pH values, groundwater level, soil contamination conditions, or the GIS-based automatic generation of geological maps, et al. In addition, it should be noted that in this work optimum site investigation programs are determined by sorely considering the trade-off between the investigation robustness and investigation effort. For the cases where there are specific restrictions on the investigation robustness or investigation efforts (e.g., required level of site investigation robustness, commitment of investigation budgets, indicatory sampling density based on code or standards), the intended objectives of the optimization problem should be modified to accommodate these specific restrictions.

Second, it is worth noting that geotechnical investigations are usually conducted in stages (e.g., feasibility study, preliminary design, detailed design, etc.). At each stage, additional investigations are conducted in the locations characterized by the highest uncertainty. This staged approach allows for graduate increase in the knowledge of ground conditions and better relation to the expected soil-structure complexity when more knowledge is available about the designed structure at later stages of design. The staged approach is not considered in this work. Also, the current development of the proposed method has not considered the constraints from multiple criteria, such as the knowledge of ground conditions or soil-structure complexity, that come with each following site investigation stage. It emphasizes on the stationary spatial variability of the soil parameters, and is generally applicable to the stages of ground investigation when the other criteria (e.g., ground conditions or soil-structure complexity) are less influential to the investigation planning.

Finally, usually various parameters, for structure design purposes, are under investigation with a number of different testing methods (boreholes, cone penetration test, flat plate dilatometer test, pressuremeter test, etc.). Different parameters may need different site investigation plans due to at least three reasons: (1) different parameters could have different spatial variability, (2) different testing methods may have different levels of measurement or transformation errors, and (3) different parameters require different levels of investigation robustness. Therefore, the site investigation program optimized for one specific parameter or testing method may not be optimal for another. In this work, spatial variability is described with factors only from soil inherent heterogeneity, which refers to the natural fluctuations due to geological and environmental effects (e.g., the variable soil constituents, preferred alignments during the transport and deposition processes of the soil grains). It has been assumed that the friction angle

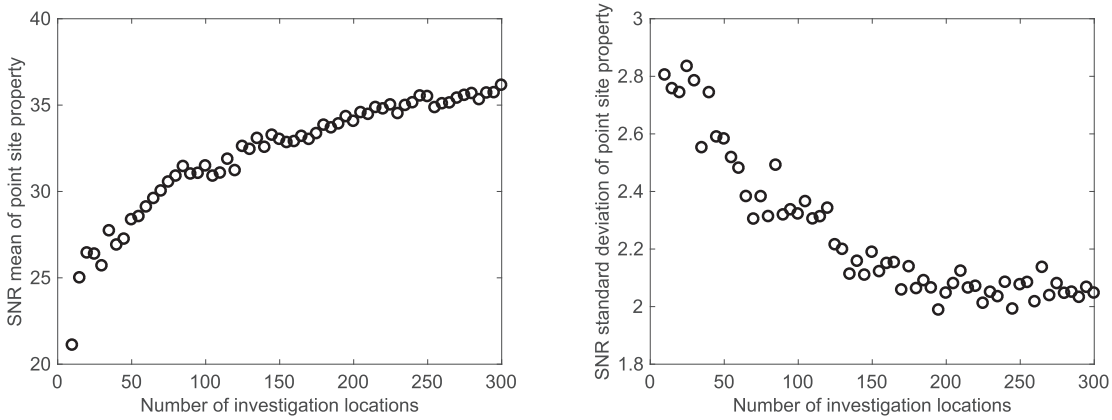


Fig. 17. Robustness of the point site property for the CVT-based site investigation programs with different numbers of investigation locations: (a) mean and (b) standard deviation.

is readily available without any measurement or transformation errors (i.e., regardless of the testing methods). Also, while this work has focused on the friction angle as an illustrative example, the proposed method is applicable to other parameters. In the cases where multiple parameters are to be characterized, it becomes a multiple-objective optimization problem with regard to the investigation robustness (or reliability) of multiple parameters and investigation effort. The issue of testing method effects and multiple soil parameters of interest is crucial in practical applications and merits future study.

6. Conclusion

This paper presented a CVT-based approach for the design and optimization of site investigation programs. This approach is applicable to an arbitrary number of investigation locations and to sites with arbitrary geometries. A modified Lloyd algorithm for the generation of CVTs has been presented. An MCMC-based Bayesian inference approach for estimating RF statistics and a Kriging interpolation-based approach for estimating the site properties at a particular location have also been briefly described. The SNR of the RF statistics and the SNR of the point site property are adopted to evaluate the robustness of a site investigation program. The optimization of the site investigation program is approached using a biobjective framework considering the investigation robustness and the investigation effort. Based on the results of two illustrative examples, the following main conclusions can be made:

1. The Lloyd algorithm for generating CVTs exhibits a linear convergence profile in a semilog diagram. The convergence rate decreases with an increasing number of CVT seeds and presents an exponential relation.
2. The robustness of a site investigation program can be better evaluated using both the SNR of the RF statistics and the SNR of the point site property. The SNR of the RF statistics reflects the global characterization accuracy, while the SNR of the point site property emphasizes the local accuracy.
3. The CVT-based site investigation programs overall perform better than those with a random layout or structural grid. Thus, it is recommended to apply the proposed method to sites with an irregular geometry and to the optimization of site investigation programs.
4. The SNR of the RF statistics, as well as the SNR mean of the point site property, increases with an increasing number of investigation locations, while the rate of increase reduces significantly once the number of investigation locations becomes sufficiently large. The SNR standard deviation of the point site property decreases with an increasing number of investigation locations and roughly reaches a plateau once the number of investigation locations becomes sufficiently large.
5. There exist knee points in the robustness plots (i.e., the SNR of the RF statistics and the SNR mean and standard deviation of the point site property vs. the number of investigation locations), corresponding to which a CVT-based site investigation program can be adopted as the optimum design considering both the investigation robustness and the investigation effort.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Project No. 51678578), the Natural Science Foundation of Guangdong Province (Project No. 2016A030313233), and the Guangzhou Municipal Science & Technology Program (Project No. 201704020139). These financial support sources are gratefully acknowledged.

References

[1] Phoon KK, Kulhawy FH. Evaluation of geotechnical property variability. Can Geotech J 1999;36:625–39.
[2] Elkateb T, Chalaturnyk R, Robertson PK. An overview of soil heterogeneity: quantification and implications on geotechnical field problems. Can Geotech J 2003;40:1–15.
[3] Chen Q, Wang C, Juang CH. Probabilistic and spatial assessment of liquefaction-induced settlements through multiscale random field models. Eng Geol 2016;211:135–49.
[4] Liu W, Chen Q, Wang C, Juang CH, Chen G. Spatially correlated multiscale Vs30 mapping and a case study of the Suzhou site. Eng Geol 2017;220:110–22.
[5] Shen M, Juang CH, Chen Q. Mitigation of liquefaction hazard by dynamic compaction – a random field perspective. Can Geotech J 2019.
[6] Samui P, Sitharam TG. Site characterization model using artificial neural network and Kriging. Int J Geomech 2010;10:171–80.
[7] Fenton GA. Random field modeling of CPT data. J Geotech Geoenviron Eng 1999;125:486–98.
[8] Goovaerts P. Geostatistical modelling of uncertainty in soil science. Geoderma 2001;103:3–26.
[9] Cho SE. Effects of spatial variability of soil properties on slope stability. Eng Geol 2007;92:97–109.
[10] Griffiths DV, Huang J, Fenton GA. Influence of spatial variability on slope reliability using 2-D random fields. J Geotech Geoenviron Eng 2009;135:1367–78.
[11] Chen Q, Wang C, Juang CH. CPT-based evaluation of liquefaction potential accounting for soil spatial variability at multiple scales. J Geotech Geoenviron Eng 2015;142:04015077.
[12] Leung YF, Liu W, Li JS, Wang L, Tsang DCW, Lo CY, et al. Three-dimensional spatial variability of arsenic-containing soil from geogenic source in Hong Kong: Implications on sampling strategies. Sci Total Environ 2018;633:836–47.
[13] Lo MK, Leung YF. Probabilistic analyses of slopes and footings with spatially variable soils considering cross-correlation and conditioned random field. J Geotech Geoenviron Eng 2017;143:04017044.
[14] Zhang L, Cheng Y, Li J, Zhou X, Jeng DS, Peng X. Wave-induced oscillatory response in a randomly heterogeneous porous seabed. Ocean Eng 2016;111:116–27.
[15] Hicks MA, Onisiphorou C. Stochastic evaluation of static liquefaction in a predominantly dilative sand fill. Géotechnique 2005;55:123–33.
[16] Wang Y, Cao Z, Li D. Bayesian perspective on geotechnical variability and site characterization. Eng Geol 2016;203:117–25.
[17] Barnes RJ. Bounding the required sample size for geologic site characterization. Math Geol 1988;20:477–90.
[18] Parsons RL, Frost JD. Evaluating site investigation quality using GIS and geostatistics. J Geotech Geoenviron Eng 2002;128:451–61.
[19] Wang Y, Au SK, Cao Z. Bayesian approach for probabilistic characterization of sand friction angles. Eng Geol 2010;114:354–63.
[20] Wang Y, Zhao T. Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. Géotechnique 2016;67:523–36.
[21] Wang X, Wang H, Liang RY, Zhu H, Di H. A hidden Markov random field model based approach for probabilistic site characterization using multiple cone penetration test data. Struct Saf 2018;70:128–38.
[22] Gong W, Luo Z, Juang CH, Huang H, Zhang J, Wang L. Optimization of site exploration program for improved prediction of tunneling-induced ground settlement in clays. Comput Geotech 2014;56:69–79.
[23] Li YJ, Hicks MA, Vardon PJ. Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields. Comput Geotech 2016;79:159–72.
[24] Gong W, Tien YM, Juang CH, Martin JR, Luo Z. Optimization of site investigation program for improved statistical characterization of geotechnical property based on random field theory. Bull Eng Geol Environ 2017;76:1021–35.
[25] Yang H, Zhang L, Xue J, Zhang J, Li X. Unsaturated soil slope characterization with Karhunen-Loève and polynomial chaos via Bayesian approach. Eng Comput 2019:1–14.
[26] BS EN, 1997–2, Eurocode 7 – Geotechnical design – Part 2: Ground investigation and testing, British Standard, British Standard Institution; 2007.
[27] McBratney AB, Webster R, Burgess TM. The design of optimal sampling schemes for local estimation and mapping of regionalized variables – I: Theory and method. Comput Geosci 1981;7:331–4.
[28] Dhakal AS, Amada T, Aniya M. Landslide hazard mapping and its evaluation using GIS: an investigation of sampling schemes for a grid-cell based quantitative method. Photogram Eng Remote Sens 2000;66:981–9.
[29] Jaksa MB, Goldsworthy JS, Fenton GA, Kaggwa WS, Griffiths DV, Kuo YL, et al. Towards reliable and effective site investigations. Géotechnique 2005;55:109–21.
[30] Olea RA. Sampling design optimization for spatial functions. J Int Assoc Math Geol 1984;16:369–92.
[31] Du Q, Faber V, Gunzburger M. Centroidal Voronoi tessellations: applications and algorithms. SIAM Rev 1999;41:637–76.
[32] Der Kiureghian A, Ke JB. The stochastic finite element method in structural reliability. Probab Eng Mech 1988;3:83–91.
[33] Davis MW. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math Geol 1987;19:91–8.
[34] Hoffman Y, Ribak E. Constrained realizations of Gaussian fields-a simple algorithm. Astrophys J 1991;380:L5–8.
[35] Gong W, Juang CH, Martin JR, Tang H, Wang Q, Huang H. Probabilistic analysis of tunnel longitudinal performance based upon conditional random field simulation of soil properties. Tunn Undergr Space Technol 2018;73:1–14.
[36] Gilks WR, Richardson S, Spiegelhalter D. Markov chain Monte Carlo in practice.

London: Chapman & Hall; 1995.
[37] Li XY, Zhang LM, Li JH. Using conditioned random field to characterize the variability of geologic profiles. J Geotech Geoenviron Eng 2015;142:04015096.
[38] Ching J, Wu SS, Phoon KK. Statistical characterization of random field parameters using frequentist and Bayesian approaches. Can Geotech J 2015;53:285–98.
[39] Baecher GB, Christian JT. Reliability and statistics in geotechnical engineering. John Wiley & Sons; 2005.
[40] Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory 1982;28:129–37.
[41] Ju L, Du Q, Gunzburger M. Probabilistic methods for centroidal Voronoi tessellations and their parallel implementations. Parallel Comput 2002;28:1477–500.
[42] Hateley JC, Wei H, Chen L. Fast methods for computing centroidal Voronoi tessellations. J Sci Comput 2015;63:185–212.
[43] Fortune S. A sweepline algorithm for Voronoi diagrams. Algorithmica 1987;2:153–74.
[44] Lloret-Cabot M, Hicks MA, Van Den Eijnden AP. Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging. Géotech Lett 2012;2:123–7.
[45] Phadke MS. Quality engineering using robust design. Prentice Hall PTR; 1995.
[46] Phoon KK, Fenton GA. Estimating sample autocorrelation functions using bootstrap. Proceedings, Ninth ASCE specialty conference on probabilistic mechanics and structural reliability, Albuquerque, New Mexico. 2004. p. 26–8.
[47] Deb K, Gupta S. Understanding knee points in bicriteria problems and their implications as preferred solution principles. Eng Optim 2011;43:1175–204.