Research Article

CPT-Based Probabilistic Characterization of Undrained Shear Strength of Clay

Mi Tian¹ and Xiaotao Sheng²

¹School of Civil Engineering, Architecture and Environment, Hubei University of Technology, 28 Nanli Road, Wuhan 430068, China
²Key Laboratory of Geotechnical Mechanics and Engineering of Ministry of Water Resources, Yangtze River Scientific Research Institute, 23 Huangpu Street, Wuhan 430010, China

Correspondence should be addressed to Mi Tian; mitian0525@gmail.com

Received 14 November 2019; Accepted 9 January 2020; Published 7 February 2020

Academic Editor: Arnaud Perrot

Copyright © 2020 Mi Tian and Xiaotao Sheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Applying random field theory involves two important issues: the statistical homogeneity (or stationarity) and determination of random field parameters and correlation function. However, the profiles of soil properties are typically assumed to be statistically homogeneous or stationary without rigorous statistical verification. It is also a challenging task to simultaneously determine random field parameters and the correlation function due to a limited amount of direct test data and various uncertainties (e.g., transformation uncertainties) arising during site investigation. This paper presents Bayesian approaches for probabilistic characterization of undrained shear strength using cone penetration test (CPT) data and prior information. Homogeneous soil units are first identified using CPT data and subsequently assessed for weak stationarity by the modified Bartlett test to reject the null hypothesis of stationarity. Then, Bayesian approaches are developed to determine the random field parameters and simultaneously select the most probable correlation function among a pool of candidate correlation functions within the identified statistically homogeneous layers. The proposed approaches are illustrated using CPT data at a clay site in Shanghai, China. It is shown that Bayesian approaches provide a rational tool for proper determination of random field model for probabilistic characterization of undrained shear strength with consideration of transformation uncertainty.

1. Introduction

Soil materials are nature materials, and their properties inherently vary from one location to another due to their natural geologic process, which is known as “inherent spatial variability (ISV).” ISV of design soil properties (e.g., undrained shear strength, $S_u$) can be probabilistically characterized using random field theory in probability-based geotechnical analyses and designs. In the context of random field theory, a design soil property within a statistically homogeneous layer is described by a series of random variables. These random variables are with the same mean, $\mu$, and standard deviation, $\sigma$; and the autocorrelation among these random variables depends on scale of fluctuation, $\lambda$, and the correlation function (such as those shown in Figure 1) (e.g., [2–6]). For a statistically homogeneous soil layer, the random field model of design soil properties can be uniquely defined by a given set of random field parameters (i.e., $\mu$, $\sigma$, and $\lambda$) and correlation function. Note that applying random field model to describe ISV of design soil properties involves two important issues: the statistical homogeneity (or stationarity) and determination of random field parameters and correlation function.

Statistical homogeneity (or stationarity) means that the mean and standard deviation of design soil properties are assumed to be spatially constant, and the correlation function only depends on their separation distance between two different locations (e.g., [2, 7, 8]). Statistical homogeneity is an important prerequisite for statistical analyses (e.g., determining random field parameters) on a set of observation data (e.g., [2, 7, 9, 10]). If the stationarity of observation data is not properly verified, the random field
model may be biased estimate. However, the profiles of soil properties are typically assumed to be statistically homogeneous or stationary without rigorous statistical verification in the previous studies (e.g., [11, 12]).

For observation data sets that satisfy the statistically homogeneous or stationary conditions, the inherent soil variability can be evaluated using random field model. However, it is a challenging task to identify random field parameters and select the proper correlation function of design soil properties at a given site, because only a small portion of soil materials are examined during site investigation and the amount of observation data is usually too limited for meaningful estimates of random field parameters. Furthermore, design soil properties (e.g., undrained shear strength, \( S_o \)) used in geotechnical designs are often estimated indirectly through transformation models (e.g., the empirical regression between the cone penetration test (CPT) data and undrained shear strength, \( S_o \)). The transformation model is not a perfect relationship but is associated with uncertainties/dispersion about a mean trend, namely, “transformation uncertainty” (e.g., [13]), which shall be rationally considered to describe ISV of design soil properties. This can be addressed as an inverse analysis problem within a Bayesian framework, which can provide a rational vehicle to determine the random field parameters and correlation function by combining the project-specific data with prior information (e.g., [14]). Bayesian approaches have been increasingly used for probabilistic characterization of soil properties, such as probabilistic characterization of sand effective friction angle (e.g., [11, 12, 15]) and normalized cone tip resistance (e.g., [16]).

In this study, Bayesian approaches are adapted to characterize the ISV of undrained shear strength of clay using CPT data. The homogeneous soil units are first identified based on the site-specific CPT data and subsequently assessed for stationarity using the modified Bartlett test. Then, Bayesian approaches are developed to determine the random field parameters (i.e., \( \mu, \sigma, \) and \( \lambda \)) and simultaneously select the most probable correlation function among the candidate correlation functions (as shown in Figure 1) of undrained shear strength within a statistically homogeneous layer. Prior knowledge and project-specific CPT data are incorporated into the proposed Bayesian approaches to describe the ISV of undrained shear strength. Finally, the Bayesian approaches are illustrated using a set of real-life CPT data.

2. Identification of Statistical Homogeneity or Stationarity

2.1. Identification of Homogeneous Soil Units. An important prerequisite for the determination of random field model is the condition of observation data, if necessary, to achieve statistical homogeneity (or stationarity). The mean and standard deviation tend to be specifically related to the soil type, especially for the mean. The correlation function of soil properties also depends on the soil type (e.g., [17]). Investigating uniform soil units in composition or behavior is of great significance for the assumption of statistical homogeneity, which should be satisfied in the subsequent statistical analyses (e.g., [1, 7, 10, 17–21]). Homogeneous soil units can be examined using CPT data, which provides direct information regarding the soil behavior type by measuring both the cone tip resistance \( q_c \) and sleeve friction \( f_s \). Based on the measured \( q_c \) and \( f_s \), the soil behavior type can be identified using the soil behavior classification index, \( I_\text{c} \) (e.g., [19, 20, 22]). Consider, for example, the soil behavior classification index shown in Table 1. The soil behavior type is divided into nine different types corresponding to the ranges of \( I_\text{c} \). Based on which zones the measured \( I_\text{c} \) is located in, the corresponding soil type is determined accordingly. \( I_\text{c} \) is calculated by the normalized cone resistance, \( Q_t \), and the normalized friction ratio, \( F_R \):

\[
I_\text{c} = \left[ \left( \frac{3.47 - \log Q_t}{\log F_R + 1.22} \right)^2 \right]^{0.5}.
\]

The normalized cone resistance \( Q_t \) is proposed by Robertson [23] and Robertson [24] and it is written as

\[
Q_t = \frac{q_t - \sigma_{o0}}{\sigma_{o0}}
\]

where \( q_t = q_c + (1 - a)u \) is the corrected cone tip resistance, in which \( u \) is pore pressure measured between the cone tip and the friction sleeve, and \( a \) is net area ratio; \( \sigma_{o0} \) is the total vertical stress; \( \sigma_{o0} \) is the vertical effective stress. The normalized friction ratio \( F_R \) is given by (e.g., [23, 24])

\[
F_R = 100 \frac{f_s}{q_t - \sigma_{o0}}.
\]

Homogeneous soil units are identified from the CPT data by delineating the data into sections where coefficient of
variation (COV) of \( I_c \) is less than 0.1 (e.g., [19, 25]) and \( I_c \) does not exhibit gradual changes with depth. These conditions are imposed to ensure that there are sufficient data points for stationarity assessment and the sampling size is large enough for meaningful statistical analyses.

2.2. Assessment of Stationarity. The homogeneous soil units are further screened to reject the null hypothesis of stationarity. Due to the limitation of sampling size in geotechnical practice, stationarity generally can only be assessed in a weak sense. Numerous classical statistical tests have been applied to assess weak stationarity, such as statistical runs test, Kendall’s tau test, and Spearman’s rank coefficient, to correlated data (e.g., [20, 26]), which may result in biased assessments. The modified Bartlett statistic (MBS) procedure proposed by Phoon et al. [1] provides a more rational basis for rejecting the null hypothesis of weak stationarity in the correlated case. In the MBS procedure, a continuous Bartlett statistic profile, \( B_{\text{stat}} \), is first generated by moving a sampling window over the soil profile. The peak value of the Bartlett statistic profile, \( B_{\text{max}} \), is compared with the critical value, \( B_{\text{crit}} \), for rejection or acceptance of the weak stationarity hypothesis. Phoon et al. [1] has proposed the rejection criterion corresponding to the commonly used correlation functions (shown in Table 2) (e.g., [1]).

Bartlett’s test is used to test the equality of multiple sample variances for independent data sets. For the case of two sample variances, the sampling window is divided into two equal segments \( s_1^2 \) and \( s_2^2 \). The Bartlett test statistic is written as

\[
B_{\text{stat}} = \frac{2.30259(w-1)}{A} \left[ 2 \log s^2 - \left( \log s_1^2 + \log s_2^2 \right) \right],
\]

in which \( w \) is the number of data points used to evaluate \( s_1^2 \) or \( s_2^2 \); \( s^2 = (s_1^2 + s_2^2)/2 \) is the total variance; \( A = 1 + 1/(2w-2) \) is a constant. The Bartlett statistic profile can be generated by moving a sampling window over the soil profiles. Then, the peak value \( B_{\text{max}} \) is compared with the critical value, \( B_{\text{crit}} \), to assess the stationarity. The null hypothesis of stationarity is rejected at the 5% level of significance if \( B_{\text{max}} > B_{\text{crit}} \).

3. Random Field Model of Undrained Shear Strength

Random field model (e.g., [3]) is applied in this paper to describe the ISV of undrained shear strength, \( S_u \), within a single statistically homogeneous layer. Insights from soil mechanics show that \( S_u \) itself is not a fundamental soil property but is related to its corresponding vertical effective stress, \( \sigma_{vo} \) (e.g., [27–30]). Then, the undrained shear strength ratio \( r = S_u/\sigma_{vo} \), where \( S_u \) and \( \sigma_{vo} \) represent the undrained shear strength and vertical effective stress at a given depth, respectively. Normalization by the vertical effective stress is used to remove the trend in the undrained shear strength profile, which generally increases with depth [31]. Other detrending methods can also be applied, such as trend line regression analysis using either linear or polynomial curve fittings [32].

Consider, for example, \( r \) at different depths are modeled by a series of spatially correlated lognormal variables with a mean \( \mu \) and standard deviation \( \sigma \) (i.e., one-dimensional stationary lognormal random field) without explicit consideration of the variation trend at different depths. The random field model adopted in this study is generally reasonable for soft clays (e.g., normal consolidated or slightly overconsolidated clays). The logarithm of \( r \) (i.e., \( y = \ln r \)) is a normal random variable with a mean \( \mu_N = \ln \mu - \sigma_N^2/2 \) and standard deviation \( \sigma_N = \sqrt{\ln[1 + (\sigma/\mu)^2]} \). The spatial correlation between variations of \( y \) at different depths is specified by the scale of fluctuation \( \lambda \) and a correlation function \( M \). Examples of correlation functions include the single exponential correlation function (SECF), binary noise correlation function (BNCF), second-order Markov correlation function (SMCF), and squared exponential correlation function (SQECF), as shown in Figure 1.

Let \( y = [y(D_1), y(D_2), \ldots, y(D_n)] \) be a vector of \( y(D) \) at different depths \( D_1, D_2, \ldots, D_n \). Because \( y \) is considered to be normally distributed with a mean \( \mu_N \) and standard deviation \( \sigma_N \), \( y \) is a Gaussian vector with a mean vector \( \mu_N \) and

<table>
<thead>
<tr>
<th>Table 1: Soil behavior type index ( I_c ) (after [20]).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil behavior zone</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Critical modified Bartlett test statistic at 5% level of significance for various common autocorrelation models (( I_2 = 1 )) (after [1]).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation function</td>
</tr>
<tr>
<td>SECF</td>
</tr>
<tr>
<td>BNCF</td>
</tr>
<tr>
<td>SMCF</td>
</tr>
<tr>
<td>SQEFC</td>
</tr>
</tbody>
</table>

Note: \( I_1 = N_0/\nu \) is the normalized sampling length, where \( N_0 \) is the total number of points in entire soil record, \( \nu = \lambda/\Delta \) is the number of points in one scale of fluctuation, and \( \Delta \) is the sampling interval; \( I_2 = w/\nu \) is the normalized segment length.
covariance matrix $\sigma_N^2 R$, in which $l$ is a vector with $n$ components that are all equal to one and $R$ is the correlation matrix of $\gamma(D)$. Then, $\gamma$ can be written as

$$\gamma = \mu_N l + \sigma_N L^T Z,$$

in which $Z$ is a standard normal random variable vector; $L$ is an $n$ by $n$ upper-triangular matrix obtained by Cholesky decomposition of $R$. The $(i, j)$th entry of $R$ represents the correlation coefficient between $\gamma(D_i)$ and $\gamma(D_j)$ at respective depths of $D_i$ and $D_j$, and it is given by the scale of fluctuation $\lambda$ and a correlation function $M$ as shown in Figure 1.

The undrained shear strength $S_u$ of clay can be estimated from in situ test, such as CPT through empirical regression functions. Consider, for example, the following empirical regression function (e.g., [33]):

$$S_u = 0.0789 (q_t - \sigma_{ob}').$$

Equation (6) can be rewritten as

$$r = \frac{S_u}{\sigma_{ob}} = 0.0789 \left(\frac{q_t}{\sigma_{ob}'}\right),$$

where the effects of vertical effective stress, $\sigma_{ob}'$, are taken into account. In a log scale, the logarithm of $r$ is rewritten as

$$\ln\left(\frac{S_u}{\sigma_{ob}}\right) = \ln\left(\frac{q_t}{\sigma_{ob}'}\right) - 2.540 = \ln Q_t - 2.540.\quad (7)$$

A simple Bayesian argument can transform the above equation into

$$\ln Q_t = a \ln r + b + \epsilon,$$  

where $a = 1; b = 2.540; \epsilon$ is a Gaussian random variable with a mean of zero and a standard deviation $\sigma_\epsilon$ of 0.34. $\epsilon$ represents the transformation uncertainty associated with the empirical regression function.

Let $\xi = [\xi(D_1), \xi(D_2), \ldots, \xi(D_n)]^T$ be a vector of ln$Q_t$ data at different depths $D_1, D_2, \ldots, D_n$. Using equations (5) and (8) gives

$$\xi = (a\mu_N + b) l + (a\sigma_N L^T Z + \sigma_\epsilon \xi).\quad (9)$$

When the spatial variability is assumed to be independent of transformation uncertainty, that is, $Z$ and $\xi$ are independent, $\xi$ is a Gaussian random vector with a mean vector $(a\mu_N + b) L$ and covariance matrix $C = a^2 \sigma_N^2 R + \sigma_\epsilon^2 L$, where $L$ is an $n \times n$ identity matrix.

Knowledge on $\mu$, $\sigma$, and $\lambda$ and correlation function $M$ are required to define the probability distribution of $\xi$. The next section presents Bayesian approaches for identifying the random field parameters (i.e., $\mu, \sigma, \lambda$) and select the most probable correlation function among a pool of candidates (e.g., those shown in Figure 1) based on both project-specific data (e.g., CPT data $\xi$) and prior knowledge.

4. Bayesian Identification of Random Field Model

4.1. Bayesian Identification of Random Field Parameters.

For a given correlation function $M$, the stationary normal random field of $r$ within a single statistically homogeneous layer is uniquely represented by the random field parameters $X = [\mu, \sigma, \lambda]$. Note that the correlation function $M$ is assumed to take a specific form (e.g., one of the correlation functions shown in Figure 1) in this section, but its corresponding $\lambda$ value is unknown herein. A proper form of the correlation function will be determined among a pool of the candidates (e.g., those shown in Figure 1) by a Bayesian model class selection approach in Section 4.2 entitled “Bayesian Selection of Correlation Function.”

For a given set of prior knowledge and site observation data $\xi$, there are various possible values of random field parameters, and their respective plausibility can be quantified by the posterior distribution $P(X|\xi, M)$ under a Bayesian framework, such as the condition on $M$ indicates that the correlation function $M$ is assumed to take a specific form. Using Bayes’ theorem, the posterior distribution of random field parameters $P(X|\xi, M)$ is written as (e.g., [14, 15, 34–36])

$$P(X|\xi, M) = K^{-1}P(\xi|X, M)P(X|M),\quad (10)$$

in which $\xi = [\xi(D_1), \xi(D_2), \ldots, \xi(D_n)]^T$ is a set of values of $\xi (D)$ measured at different depths $D_1, D_2, \ldots, D_n$ in a single clay layer, respectively; $K = P(\xi|X, M)$ is a normalizing constant that is independent of $X$ for a given $M$, and it is calculated as $\int P(\xi|X, M)P(X|M)dX; P(\xi|X, M)$ is the likelihood function that reflects the model fit with observation data $\xi$ for a given $M$ and random field parameters $X$; and $P(X|M)$ is the prior distribution of random field parameters $X$ reflecting the available prior knowledge on $X$ in the absence of data. When there is no prevailing knowledge on $X$, a relatively noninformative prior distribution can be used to reflect prior knowledge, for example, a joint uniform distribution that is equal to a constant (e.g., [36]). When local information (e.g., local experience and engineering judgments) is available at a site, the prior knowledge becomes more informative. More sophisticated types of prior distributions (e.g., an arbitrary histogram type of prior distribution, [37]) can be obtained. This may complicate the posterior distribution in equation (10) and lead to difficulties in expressing it analytically and explicitly.

As mentioned before, $\xi$ is a Gaussian random vector with a mean vector $(a\mu_N + b)L$ and covariance matrix $C$. The likelihood function is thus given by

$$P(\xi|X, M) = (2\pi)^{-n/2}|det C|^{-1/2} \exp\left\{-\frac{1}{2} (\xi - (a\mu_N + b)L)^T C^{-1} (\xi - (a\mu_N + b)L)\right\}.\quad (11)$$

The prior distribution and the likelihood function are then substituted into equation (10) to obtain the posterior distribution $P(X|\xi, M)$. The posterior distribution involves high dimension integration, which is one key limitation of Bayesian methods. To avoid the computational complexity, Markov Chain Monte Carlo Simulation (MCMCS) is adapted in this study to generate a large number of samples of random field parameters $X$ from the posterior distribution (i.e., equation (10)). Note that MCMCS provides an
appropriate tool to obtain the posterior knowledge in Bayesian analysis (e.g., [38]). Metropolis-Hastings (M-H) algorithm is used to generate a large number of \( X \) samples to describe the posterior distribution (i.e., equation (10), e.g., [15]). Samples of \( X \) are subsequently used to estimate the statistics of random field parameters through conventional statistical analyses.

### 4.2. Bayesian Selection of Correlation Function

The correlation function \( M \) within a single statistically homogeneous clay layer is considered as a prescribed one with unknown \( \lambda \) in the previous section. This section presents a Bayesian model class selection method to determine the most probable correlation function \( M^* \) among a pool of candidate correlation functions for a given set of \( \xi \). Consider a number (i.e., \( N_{\text{CF}} \)) of candidate correlation functions \( M_k, \ k = 1, 2, \ldots, N_{\text{CF}} \) within the clay layer, such as SECF, BNCF, SMCF, and SQECF shown in Figure 1. For a given set of \( \xi \), the plausibility of the \( k \)-th candidate correlation function \( M_k \) in the clay layer is quantified by its occurrence probability \( P(M_k \mid \xi) \). Using Bayes’ theorem, \( P(M_k \mid \xi) \) is given by (e.g., [15, 36])

\[
P(M_k \mid \xi) = \frac{P(\xi \mid M_k) P(M_k)}{P(\xi)}, \quad k = 1, 2, \ldots, N_{\text{CF}}, \tag{12}
\]

where \( P(\xi) \) is a normalizing constant that is independent of \( M_k \) and it is calculated as \( P(\xi) = \sum_{k=1}^{N_{\text{CF}}} P(\xi \mid M_k) P(M_k) \); \( P(M_k) \) is the prior probability of \( M_k \); and \( P(\xi \mid M_k) \) is the conditional probability density function of \( \xi \) for a given correlation function \( M_k \), and it quantifies the information on \( M_k \) provided by CPT data and is referred to as “evidence” for \( M_k \) provided by \( \xi \).

In the case of no prevailing prior knowledge, the \( N_{\text{CF}} \) candidate correlation functions in the clay layer are considered to have the same prior probability \( P(M_k) = 1/N_{\text{CF}} \). Since \( P(\xi) \) and \( P(M_k) \) in equation (12) are constants, \( P(M_k \mid \xi) \) is proportional to \( P(\xi \mid M_k) \). The most probable correlation function \( M^* \) is the one that has the maximum value of \( P(M_k \mid \xi) \), and it also has the maximum value of \( P(\xi \mid M_k) \). Note that \( P(\xi \mid M_k) \) is actually the normalizing constant \( K \) in equation (10) by setting \( M = M_k \), which involves a multidimensional integration. As discussed in the previous section, there is no need to calculate \( K \) (i.e., \( P(\xi \mid M_k) \)) when MCMCS is adapted to the posterior distribution in equation (10) for \( M_k \). However, it is needed herein for Bayesian selection of correlation function. MCMCS samples of \( X \) generated from equation (10) for \( M_k \) and Gaussian copula are used to calculate the evidence \( P(\xi \mid M_k) \) (e.g., [15]), maintaining the advantage of no need to directly calculate the multidimensional integration (i.e., \( K \)).

Using the Theorem of Total Probability, the evidence \( P(\xi \mid M_k) \) for \( M_k \) is expressed as

\[
P(\xi \mid M_k) = \int P(\xi \mid X, M_k) P(X \mid M_k) dX, \tag{13}
\]

where \( P(\xi \mid X, M_k) \) is the likelihood function by setting \( M = M_k \); and \( P(X \mid M_k) \) is the prior distribution of \( X \). Then, the logarithm of evidence (i.e., \( \ln[P(\xi \mid M_k)] \)) can be written as (e.g., [36])

\[
\ln[P(\xi \mid M_k)] = \int \ln[P(\xi \mid X, M_k)] P(X \mid M_k) dX - \int \ln \left[ \frac{P(X \mid \xi, M_k)}{P(X \mid M_k)} \right] P(X \mid \xi, M_k) dX.
\]

(14)

Based on the \( N_p \) random samples of \( X \) estimated from \( P(X \mid \xi, M_k) \), \( \ln[P(\xi \mid M_k)] \) can be estimated as

\[
\ln[P(\xi \mid M_k)] = \frac{\sum_{i=1}^{N_p} \ln\left\{ P(\xi_i \mid M_k) \right\}}{N_p} - \frac{\sum_{i=1}^{N_p} \ln\left\{ \frac{P(x_i \mid \xi, M_k)}{P(x_i \mid M_k)} \right\}}{N_p}.
\]

(15)

in which \( P(x_i \mid M_k) \) is the prior distribution evaluated at \( i \)-th sample \( x_i = [\mu, \sigma, \lambda] \) of \( X \), which is assumed to be a constant; \( P(\xi_i \mid x_i, M_k) \) is the likelihood function calculated using equation (11) for each \( x_i \). \( P(x_i \mid \xi, M_k) \) is posterior distribution evaluated using equation (10) at \( i \)-th sample \( x_i \).

Evaluating \( P(x_i \mid \xi, M_k) \) needs the posterior distribution \( P(X \mid \xi, M_k) \) in equation (10), which is unknown and is numerically represented by \( N_p \) MCMCS samples of \( X \). To evaluate \( P(X \mid \xi, M_k) \), copula theory is used to reconstruct \( P(X \mid \xi, M_k) \) based on \( N_p \) samples of \( X \) in this paper. Details of calculation of the evidence using MCMCS and Gaussian copula can be found in Tian et al. [15].

For a given \( M_k \), the MCMCS-based Bayesian approach in the previous subsection is applied to generate random samples of \( X \) from the posterior distribution. These samples are subsequently used in equation (15) to evaluate \( \ln[P(\xi \mid M_k)] \). The procedure is repetitively performed \( N_{\text{CF}} \) times for \( M_k \), \( k = 1, 2, \ldots, N_{\text{CF}} \) to obtain their respective values of \( \ln[P(\xi \mid M_k)] \). Then, the most probable correlation function \( M^* \) is determined by comparing the values of \( \ln[P(\xi \mid M_k)] \) in each candidate correlation function. The one with the maximum value of \( \ln[P(\xi \mid M_k)] \) is taken as \( M^* \). Its corresponding posterior statistics (e.g., posterior mean values \( \mu', \sigma', \text{ and } \lambda' \)) estimated from MCMCS samples constitute the posterior knowledge on random field parameters. In other words, the most probable correlation function \( M^* \) and posterior estimates (e.g., \( \mu', \sigma' \), and \( \lambda' \)) of random field parameters are determined simultaneously for probabilistic characterization of the ISV of undrained shear strength based on both prior knowledge and CPT data.

### 5. Implementation Procedure

The Bayesian approach for probabilistic characterization comprises ten steps. Details of each step and its associated equations are summarized as follows:

1. Obtain a set of cone tip resistance \( q_c \) and sleeve friction \( f_s \) versus depth data for a clay layer from...
CPT, and convert them to the values of normalized cone resistance $Q_t$ and normalized friction ratio $F_R$ at different depths using equations (2) and (3), respectively.

(2) Calculate the soil behavior classification index $I_c$ using equation (1) and identify homogeneous soil layers by examining whether the Cov of $I_c$ is less than 0.1.

(3) Convert the values of $Q_t$ to $\xi = [\xi(D_1), \xi(D_2), \ldots, \xi(D_n)]$ at different depths $D_1, D_2, \ldots, D_n$ and perform conventional statistical analyses on the CPT data $\xi$ (i.e., ln $Q_t$) to estimate its scale of fluctuation by fitting a candidate correlation function (e.g., SECF, BNCF, SMCF, and SQECF shown in Figure 1) to the sample autocorrelation function, calculate Bartlett statistic profiles using

\[ \sigma'_{v0} = 8D + 14.7 \]
Figure 4: Modified Bartlett statistic profiles of CPT data for the medium clay layer. (a) $M_1$. (b) $M_2$. (c) $M_3$. (d) $M_4$. 

Figure 5: Continued.
equation (4) where estimated scale of fluctuation \( \ln Q_t \) is used, and compare the peak values of Bartlett statistic profiles \( B_{\text{stat}} \) with the critical values \( B_{\text{crit}} \) to assess the stationarity of CPT data \( \xi \).

(4) Choose the soil layers that satisfy both statistical homogeneity and weak stationarity to probabilistically characterize the ISV of undrained shear strength using random field model.

(5) Determine \( N_{\text{CF}} \) candidate correlation functions \( M_k \), \( k = 1, 2, \ldots, N_{\text{CF}} \), of a statistically homogeneous clay layer, such as SECF, BNCF, SMCF, and SQECF shown in Figure 1.

(6) Use M-H algorithm to generate \( N_p \) samples of \( \mu, \sigma, \) and \( \lambda \) based on equation (10) for a given \( M_k \).

(7) Estimate the respective posterior statistics (e.g., \( \mu', \sigma', \) and \( \lambda' \)) based on their MCMCS samples through conventional statistical analyses.

(8) Calculate \( \ln[P(\xi | M_k)] \) of \( M_k \) using equation (15), and the evidence \( P(\xi | M_k) \) is then obtained by taking the exponential of \( \ln[P(\xi | M_k)] \).

(9) Repeat steps (6)–(8) \( N_{\text{CF}} \) times to generate \( N_p \) MCMCS samples of \( \mu, \sigma, \) and \( \lambda \) and calculate \( \ln[P(\xi | M_k)] \) and \( P(\xi | M_k) \) for each candidate correlation function in the statistically homogeneous clay layer.

(10) Compare the values of \( \ln[P(\xi | M_k)] \) (or \( P(\xi | M_k) \)) of the \( N_{\text{CF}} \) candidate correlation functions and select the one with the maximum value of \( \ln[P(\xi | M_k)] \) as \( M^* \) in the statistically homogeneous clay layer. Its corresponding posterior statistics and distributions reflect the posterior knowledge on \( \mu, \sigma, \) and \( \lambda \) and probabilistically characterize the ISV of undrained shear strength of the statistically homogeneous clay layer.

The ten steps can be readily implemented in commercial software packages, such as MATLAB (e.g., [39]). For geotechnical practitioners, the user function can be used as a “black box.” They need to provide project-specific test results (e.g., CPT data), prior knowledge (e.g., reasonable ranges of soil properties), transformation model, candidate correlation functions, and the number of MCMCS samples as input. The user function returns the most probable correlation function and its corresponding posterior estimates (e.g., mean values and standard deviations) of random field parameters within the statistically homogeneous layers at the

| Soil type                  | Correlation function | \( P(\xi | M_k) \) | \( \mu' \) | \( \sigma' \) | \( \lambda' \) (m) |
|----------------------------|----------------------|-------------------|----------|-----------|-----------------|
| Medium clay layer          | \( M_1 \)            | 0.251             | 0.21     | 0.01      | 3.41            |
|                            | \( M_2 \)            | 0.189             | 0.21     | 0.01      | 3.33            |
|                            | \( M_3 \)            | 0.267             | 0.21     | 0.01      | 3.58            |
|                            | \( M_4^a \)          | 0.293             | 0.21     | 0.02      | 3.44            |
| Medium silty clay layer    | \( M_1 \)            | 0.105             | 0.23     | 0.03      | 4.41            |
|                            | \( M_2 \)            | 0.173             | 0.23     | 0.03      | 4.04            |
|                            | \( M_3 \)            | 0.281             | 0.23     | 0.03      | 4.46            |
|                            | \( M_4^a \)          | 0.441             | 0.24     | 0.04      | 4.30            |

*Most probable correlation function.
specific site as output. This significantly improves the practicality of the Bayesian approaches. Identification of statistically homogeneous units and the Bayesian approach and its implementation procedure described above are illustrated through a set of real-life CPT data in the next section.

6. Illustrative Example

The Bayesian approaches are illustrated for probabilistic characterization of the ISV of undrained shear strength of soft clay at Yili station in Shanghai, China. Yili station serves as an interchange station for two orthogonal crossing subway lines in Shanghai (e.g., [40]). The site is comprised of six layers below the ground surface at a depth of about 49.2 m. The uppermost layer is a thin fill soil extended from the ground surface to about 0.8 m. The second layer, termed medium-soft clay, is slightly over-consolidated owing to desiccation. The third layer, the fourth layer, and the fifth layer are generally normally consolidated clay, with low permeability, high compressibility, and low strength. The sixth layer consists of medium compressible and slightly overconsolidated clay. The thicknesses of the six layers are 0.8 m, 2.7 m, 4.5 m, 9 m, 10.3 m, and 21.9 m, respectively. The groundwater table is located at about 1.5 m below the ground surface, and the total unit weight of the six layers is about 17.8 kN/m^3.

Figure 2 shows a set of cone tip resistance $q_c$, sleeve friction $f_s$, and pore water pressure $u$ versus depth, respectively, obtained from CPT at this clay site. The average measurement interval is about 0.1 m. The profile of vertical effective stress, $\sigma_{v0}'$, is shown in Figure 2. The relationship between $\sigma_{v0}'$ and depth $D$ is a linear equation: $\sigma_{v0}' = 8D + 14.7$ (kPa), where $D$ is the depth below the ground surface varying from 1.5 m to 50 m. Figure 2 also shows the $S_a$ values indirectly obtained by equation (6) using CPT data and $r$ values using the $S_a$ values and their corresponding $\sigma_{v0}'$. The CPT data is assessed for its statistical homogeneity and weak stationarity. Then, it is applied to identify random field model of $r$ using the proposed Bayesian approaches.

Table 4: Results of random field parameters and correlation function of $r$ of the medium clay layer and medium silty clay layer obtained from conventional statistical analyses.

<table>
<thead>
<tr>
<th>Soil type</th>
<th>Correlation function</th>
<th>Goodness-of-fitting $R^2$</th>
<th>$\mu'$</th>
<th>$\sigma'$</th>
<th>$\lambda'$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium clay layer</td>
<td>$M_1$</td>
<td>0.848</td>
<td>0.21</td>
<td>0.02</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>$M_2$</td>
<td>0.722</td>
<td>0.21</td>
<td>0.02</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>$M_3$</td>
<td>0.767</td>
<td>0.21</td>
<td>0.02</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>$M_4$</td>
<td>0.722</td>
<td>0.21</td>
<td>0.02</td>
<td>1.08</td>
</tr>
<tr>
<td>Medium silty clay layer</td>
<td>$M_1$</td>
<td>0.906</td>
<td>0.23</td>
<td>0.05</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>$M_2$</td>
<td>0.822</td>
<td>0.23</td>
<td>0.05</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>$M_3$</td>
<td>0.844</td>
<td>0.23</td>
<td>0.05</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>$M_4$</td>
<td>0.810</td>
<td>0.23</td>
<td>0.05</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*Best-fit correlation function.

Figure 6: Comparison of correlation functions estimated from the Bayesian approaches and the sample autocorrelation function. (a) Medium clay layer. (b) Medium silty clay layer.
Consider, for example, four candidate correlation functions shown in Figure 1, that is, SECF, BNCF, SMCF, and SQECF, which are denoted as $M_1$, $M_2$, $M_3$, and $M_4$ in this study, respectively. The prior knowledge on $\mu$, $\sigma$, and $\lambda$ is taken as a joint uniform distribution, which is equal to a constant (e.g., [36]). Based on the set of CPT data shown in Figure 2 and the prior knowledge, an MCMCS run is performed to generate 100,000 samples from the posterior distribution for each candidate correlation function. Note that the number of MCMCS samples is sufficient for MCMCS to reach the stationary condition of Markov Chain and to generate reasonably accurate probabilistic estimates of $\mu$, $\sigma$, and $\lambda$ in this study.

6.1. Results of Identification of Statistically Homogeneous Layers. Figure 3 shows the normalized cone resistance $Q_t$, normalized friction ratio $F_R$, and soil behavior classification index $I_c$ at Yili station. It can be seen that $I_c$ ranges from 2.167 to 3.712, and the soil types cover a relatively wide range (zones 3 to 5), that is, from clay to sand mixtures. Figure 3 also shows that $I_c$ of the medium clay layer and medium silty clay layer do not exhibit gradual obvious changes with depth. The mean values of $I_c$ of the medium clay layer and medium silty clay layer are 0.02 and 0.03, respectively, which in no case exceed 0.1. Therefore, the medium clay layer and medium silty clay layer are essentially identified as homogeneous soil units and selected for subsequent assessment of weak stationarity and statistical analyses.

The homogeneous soil units (i.e., medium clay layer and medium silty clay layer) at this site are further screened using the modified Bartlett test, which is capable of rejecting the null hypothesis of weak stationarity for spatially correlated data. The MBS procedure proposed by Phoon et al. [1] is used in this study to assess the weak stationarity of CPT data $\xi$. According to equation (7), there is a linear empirical relationship between $\ln r$ and CPT data $\xi$. If the profiles of $\xi$ have weak stationarity, $r$ is also assumed as weak stationarity. This assumption would be validated using the Bayesian
identified random field model of $r$. The scale of fluctuation of $\xi$ is estimated by fitting a candidate correlation function, as shown in Figure 1, to the initial part of the sample autocorrelation function. Then, the Bartlett statistic profiles of the homogeneous soil units (i.e., medium clay layer and medium silty clay layer) are calculated by the estimated scale of fluctuation.

Figures 4 and 5 show the Bartlett statistic profiles of the medium clay layer and medium silty clay layer. It is shown that the peak values of the Bartlett statistic profiles, $B_{\text{max}}$, of these two clay layers are less than the critical values, $B_{\text{crit}}$, respectively, which proves that the CPT data profiles of the medium clay layer and medium silty clay layer have weak stationarity. The CPT data $\xi$ obtained from these two clay layers can be used as input in the subsequent statistical analyses for probabilistic characterization of undrained shear strength. Then, under the Bayesian framework, CPT data and prior information are integrated to determine the random field parameters and select the most proper correlation function among a pool of candidates to characterize the ISV of $r$. It should be noted that the Bartlett statistic profiles of the medium clay layer and medium silty clay layer are calculated by equation (4), in which the scale of fluctuation of $\xi$ is robustly estimated by conventional statistical analyses.

6.2. Results of Random Field Model of Undrained Shear Strength Ratio. Table 3 summarizes the results of random field parameters and correlation function of $r$ of the medium clay layer and medium silty clay layer identified from the Bayesian approaches using the given set of CPT data (as shown in Figure 2) and prior knowledge. It is shown that, for the medium clay layer, $M_4$ (i.e., SQECF) has the largest value of $P(M_k | \xi)$ (i.e., 0.293) among $M_1$–$M_4$. Therefore, it is taken as $\hat{M}^*$ for probabilistically characterizing $r$ of the medium clay layer at Yili station. Similar observation is also obtained for the medium silty clay layer. $M_4$ (i.e., SQECF) is also taken as $\hat{M}^*$ for probabilistically characterizing $r$ of the medium silty clay layer. Table 3 also includes the respective

Figure 8: Validation of stationarity of CPT data for the medium silty clay layer. (a) $M_1$. (b) $M_2$. (c) $M_3$. (d) $M_4$. 
posterior estimates (e.g., $\mu'$, $\sigma'$, and $\lambda'$) of $\mu$, $\sigma$, and $\lambda$ of the medium clay layer and medium silty clay layer for $M_1$–$M_4$. The posterior estimates of $\mu$ corresponding to $M_4$ (i.e., $M^*$) of the medium clay layer and medium silty clay layer are 0.21 and 0.24, respectively. These values fall within the reported values (e.g., [10, 16, 41]). The posterior Cov of $r$ corresponding to $M_4$ (i.e., $M^*$) of the two clay layers are 9.5% and 16.7%, respectively. The estimated values of Cov also essentially fall within the reported values (i.e., 10% to 35%).

### 6.3. Comparison with Conventional Statistical Analyses

For comparison, conventional statistical analyses are performed on the $r$ profiles (as shown in Figure 2) estimated from CPT data using the empirical regression function shown in equation (6). $\mu$, $\sigma$, and sample autocorrelation function are directly estimated from the $r$ profiles. $\lambda$ and $M^*$ are obtained by fitting a candidate correlation function (e.g., $M_1$–$M_4$) to the sample autocorrelation function and comparing the goodness-of-fitting in different candidate correlation functions (e.g., [20, 21]). The results of estimates of $\mu$, $\sigma$, $\lambda$, and $M^*$ are compared with those obtained from the Bayesian approaches.

Table 4 summarizes the results estimated from the conventional statistical analyses, including the direct estimates of random field parameters and goodness-of-fitting for each candidate correlation function (e.g., $M_1$–$M_4$). Table 4 shows that $M_1$ (i.e., SECF) is selected as $M^*$ among $M_1$–$M_4$ for the medium clay layer and medium silty clay layer, which is obviously different from $M^*$ determined by the Bayesian approaches (see Table 3). Table 4 also includes the direct estimates of random field parameters (see columns 4-5 in Table 4) of the medium clay layer and medium silty clay layer, respectively. Compared with the results obtained from the Bayesian approaches (see columns 5 in Table 3), the differences between $\mu$ (i.e., 0.01) estimated from these two approaches are relatively small. However, the $\sigma$ values (see column 5 in Table 4) estimated from conventional statistical analyses are greater than those from the Bayesian approaches (see column 6 in Table 4).
obtained from conventional statistical analyses are less than the Bayesian approaches (see column 6 in Table 3). Such a difference might be attributed to the reason that the effect of transformation uncertainty is not explicitly considered in conventional statistical analyses (e.g., [15]). Both the ISV of $\gamma$ and the transformation uncertainty propagate into the profiles of $r$ directly estimated from the CPT data using the regression function (i.e., equation (6)). Therefore, the sample autocorrelation function estimated from the $r$ profiles does not reflect the actual spatial correlation. The variability in the $r$ profiles contains not only the ISV but also the transformation uncertainty, and it is, therefore, greater than the actual ISV of $r$, which subsequently leads to overestimation in $\sigma$ and underestimation in $\lambda$. It is also not surprising to see that the correlation function most directly obtained by fitting the sample autocorrelation function of the $r$ profiles is inconsistent with that estimated from the Bayesian approaches. In contrast, the transformation uncertainty is, explicitly and rationally, taken into account in the Bayesian approaches.

6.4. Validation of Bayesian Identification of Random Field Model. For validation, the most probable correlation functions of $r$ of the medium clay layer and medium silty clay layer obtained from the Bayesian approaches are compared with the sample autocorrelation functions, as shown in Figures 6(a) and 6(b). It is shown that the initial parts of circles plot close to the solid lines, indicating that the proposed Bayesian approaches properly identify random field model for probabilistic characterization of ISV of undrained shear strength ratio $r$.

It should be noted that the weak stationarity of CPT data $\xi$ is assessed by the scale of fluctuation estimated by conventional statistical analyses (i.e., estimated by fitting a candidate correlation function, as shown in Figure 1, to the sample autocorrelation function). For illustration, the scale of fluctuation of $\xi$ obtained from the Bayesian approaches is used to validate the weak stationarity of CPT data within the medium clay layer and medium silty clay layer. Figures 7 and 8 show the Bartlett statistic profiles of CPT data $\xi$ in the medium clay layer and medium silty clay layer are calculated.
7. Summary and Conclusions

This paper presented Bayesian approaches for probabilistic characterization of ISV of undrained shear strength within a statistically homogeneous clay layer based on CPT data and prior knowledge. Homogeneous soil units were first identified from the CPT data using the soil behavior classification index. The identified homogeneous soil units were further assessed using modified Bartlett’s test to reject the null hypothesis of weak stationarity. Then, the Bayesian approaches were used to identify random field parameters (i.e., mean $\mu$, standard deviation $\sigma$, and scale of fluctuation $\lambda$) and simultaneously select the most probable correlation function among a pool of candidate correlation functions using identified homogeneous or stationary CPT profiles and prior knowledge. MCMCS and Gaussian copula-based method were used to bypass multidimensional integration involved in the Bayesian approaches.

Equations were derived for identification of statistically homogeneous soil units and the Bayesian approaches. They were illustrated using profiles of CPT data obtained from a clay site in Shanghai, China. The statistically homogeneous clay layers were identified from the site-specific CPT data. Random field parameters and correlation function in a statistically homogeneous clay layer obtained from the Bayesian approaches were compared with those estimated from conventional statistical analyses. It has been shown that the variability in the profiles of undrained shear strength directly estimated from the CPT data using the empirical regression function contains not only the ISV of undrained shear strength but also the transformation uncertainty, which subsequently leads to overestimation in $\sigma$ and underestimation in $\lambda$. The Bayesian approaches provide a rational tool to properly identify random field parameters and correlation function for probabilistic characterization of ISV of undrained shear strength using CPT data with proper consideration of transformation uncertainty.

In addition, the results were validated by comparing the sample autocorrelation function with correlation function obtained from the Bayesian approaches. The assessment of stationarity was also validated based on the random field parameters obtained from Bayesian approaches. It was shown that the most probable correlation function obtained from the Bayesian approaches agrees well with the sample autocorrelation function. The hypothesis of stationarity of CPT data and undrained shear strength ratio profiles is accepted. The random field parameters and correlation function of undrained shear strength ratio, $r$, obtained from the Bayesian approaches can be used to properly characterize the ISV of $r$.

Data Availability

The data in the manuscript are available by contacting the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project no. 51579093), the Open Research Fund of Rock Mechanics in Hydraulic Structural Engineering of Ministry of Education, Wuhan University (Project no. RMHSE1905), and the Research Fund for the Doctoral Program of Hubei University of Technology, China (Project no. BSQD2017034).

References


