Structural Safety 49 (2014) 10–17

Contents lists available at ScienceDirect

Structural Safety

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

Bayesian model comparison and selection of spatial correlation functions for soil parameters

Zijun Cao [a,1], Yu Wang [b,*]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, No.8 Donghu South Road, Wuhan, PR China
[b] Department of Civil and Architectural Engineering, Shenzhen Research Institute, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Available online 10 January 2014

Keywords:
Inherent spatial variability
Random fields
Correlation function
Bayesian model comparison

ABSTRACT

The inherent spatial variability of soils is one of the major sources of uncertainties in soil properties, and it can be characterized explicitly using random field theory. In the context of random fields, the spatial correlation between the values of a soil property concerned at different locations is represented by its correlation structure (i.e., correlation functions). How to select a proper correlation function for a particular site has been a challenging task, particularly when only a limited number of project-specific test results are obtained during geotechnical site characterization. This paper develops a Bayesian model comparison approach for selection of the most probable correlation function among a pool of candidates (e.g., single exponential correlation function, binary noise correlation function, second-order Markov correlation function, and squared exponential correlation function) for a particular site using project-specific test results and site information available prior to the project (i.e., prior knowledge, such as engineering experience and judgments). Equations are derived for the proposed Bayesian model comparison approach, in which the inherent spatial variability is modeled explicitly using random field theory. Then, the proposed method is illustrated and validated through simulated cone penetration test (CPT) data and four sets of real CPT data obtained from the sand site of the US National Geotechnical Experimentation Sites (NGES) at Texas A&M University. In addition, sensitivity studies are performed to explore the effects of prior knowledge, the measurement resolution (i.e., sampling interval), and data quantity (i.e., sampling depth) on selection of the most probable correlation function for soil properties. It is found that the proposed approach properly selects the most probable correlation function and is applicable for general choices of prior knowledge. The performance of the method is improved as the measurement resolution improves and the data quantity increases.

1. Introduction

Geotechnical materials are natural materials, and their properties are affected by various spatially variable factors during their formation processes, such as parent materials, weathering and erosion processes, transportation agents, sedimentation conditions, etc. [1]. Geotechnical properties, therefore, vary spatially, which is usually known as "inherent spatial variability" [2,3]. The inherent spatial variability has been considered as one of the major sources of uncertainties in geotechnical properties [4–8]. Such spatial variability can be modeled explicitly in probabilistic analysis (e.g., [2,3,9–11]) and reliability-based design (RBD) (e.g., [12,13]) of geotechnical structures using random field theory. In the context of random fields, the spatial correlation between values of a geotechnical property at different locations is represented by its correlation structure (i.e., correlation functions) (e.g., [2,3,14,15]). Proper selection of the correlation function for the geotechnical property is a prerequisite for random field modeling of inherent spatial variability in probabilistic analysis and RBD of geotechnical structures.

Several theoretical correlation functions have been used in literature to analyze geotechnical data and/or to model inherent spatial variability of soil properties in probabilistic analysis and RBD of geotechnical structures, such as single exponential correlation function (SECF), binary noise correlation function (BNCF), second-order Markov correlation function (SMCF), and squared exponential correlation function (SQECF) (e.g., [16–21]). A suitable correlation function can be selected by fitting these theoretical correlation functions to the sample autocorrelation function (SACF) estimated from site observation data and comparing the goodness-of-fitting of different correlation functions (e.g., [18–21]). To enable a meaningful estimate of the SACF, a large number of site observation data is usually required. This poses a challenge in the selection of the correlation function for a given site because, generally speaking, only a limited number of project-specific test results are obtained

* Corresponding author. Tel.: +852 3442 7605; fax: +852 3442 0427.
E-mail addresses: zijuncao3@gmail.com (Z. Cao), yuwang@cityu.edu.hk (Y. Wang).
[1] Tel.: +86 27 6877 4036.

0167-4730/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.strusafe.2013.06.003

during geotechnical site characterization, particularly for projects with medium or relatively small sizes.

In addition, accuracy of the SACF usually deteriorates as the separation distance between a data pair increases because the number of data pairs used to estimate SACF decreases as the separation distance increases (e.g., [16]). In practice, the theoretical correlation function is usually only fitted to the initial part of the SACF, i.e., the part of the SACF with relatively small separation distances (e.g., [18]). Therefore, the information on the spatial correlation provided by data pairs with relatively large separation distances is not taken into account, and the limited number of site observation data obtained during geotechnical site characterization is somehow wasted.

This paper develops a Bayesian model comparison approach for selection of the most probable correlation function of a soil property concerned among a pool of candidates for a given site. The proposed Bayesian approach makes full use of available information about the site, including not only all project-specific test data but also site information available prior to the project (i.e., prior knowledge, such as local engineering experience and sound engineering judgments), and explicitly models the inherent spatial variability of soil properties. It starts with random field modeling of the inherent spatial variability of soil properties, followed by development of the Bayesian model comparison approach and description of its implementation procedure. Then, the proposed Bayesian model comparison approach is illustrated using four sets of cone penetration test (CPT) data obtained from the sand site of the US National Geotechnical Experimentation Sites (NGES) at Texas A&M University. In addition, sensitivity studies are performed using simulated CPT data to validate the effectiveness of the proposed approach and to explore the effects of prior knowledge, measurement resolution (i.e., sampling interval), and data quantity (i.e., sampling depth) on the selection of spatial correlation functions.

2. Random field modeling of inherent spatial variability

Random field theory [2,3] is applied in this study to model the inherent spatial variability of a soil property x within a statistically homogenous soil layer. Consider, for example, a one-dimensional lognormal random field $x(D)$, where D is the depth and x is a lognormal random variable with a mean μ and standard deviation σ. By the definition of lognormal random variables (e.g., [22]), the logarithm (i.e., $\xi(D) = \ln[x(D)]$) of $x(D)$ is a normal random variable with a mean $\mu_N = \ln\mu - \sigma_N^2/2$ and standard deviation $\sigma_N = \sqrt{\ln\left[1+(\sigma/\mu)^2\right]}$. In the context of random fields, the spatial correlation between the values of $\xi(D)$ at different depths is characterized by a correlation function M. For example, Fig. 1 shows four correlation functions that are commonly used in geotechnical literature (e.g., [2,3,16–19]), including single exponential correlation function (SECF, see the solid line), binary noise correlation function (BNCF, see the dashed line), second-order Markov correlation function (SMCF, see the solid line with circles), and squared exponential correlation function (SQECF, see the solid line with triangles). For SECF, BNCF, SMCF, and SQECF, the correlation coefficient $\rho(\Delta D)$ between $\xi(D_i)$ and $\xi(D_j)$ at respective depths of D_i and D_j is calculated, respectively, as (e.g., [2,3,17–19])

$$\text{SECF}: \quad \rho(\Delta D) = \exp(-2|\Delta D|/\lambda) \tag{1}$$

$$\text{BNCF}: \quad \rho(\Delta D) = \begin{cases} 1-|\Delta D|/\lambda & \text{for } |\Delta D| \leqslant \lambda \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$SMCF: \quad \rho(\Delta D) = (1+4|\Delta D|/\lambda)\exp(-4|\Delta D|/\lambda) \tag{3}$$

$$\text{SQECF}: \quad \rho(\Delta D) = \exp[-\pi(\Delta D/\lambda)^2] \tag{4}$$

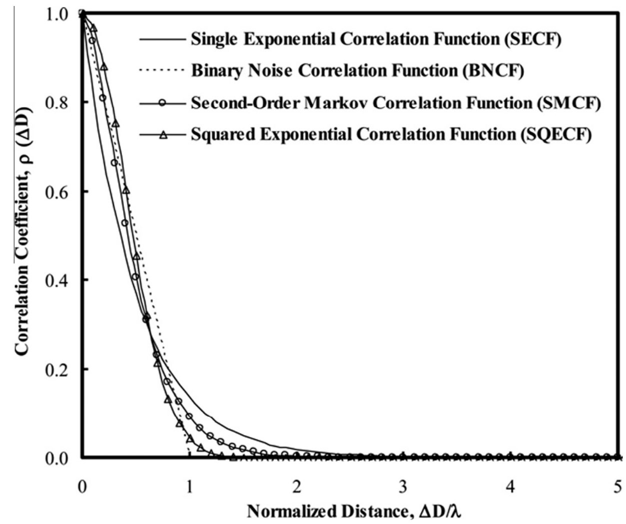

Fig. 1. Four correlation functions commonly used in geotechnical engineering (After [17]).

in which λ = correlation length, also known as "scale of fluctuation" (e.g., [2,3]); $\Delta D = |D_i - D_j|$ = separation distance between D_i and D_j.

Let $\underline{\xi} = [\xi(D_1), \xi(D_2), \ldots, \xi(D_k)]^T$ be a vector of $\xi(D)$ at k different depths. Because $\xi(D)$ is considered as normally distributed with a mean μ_N and standard deviation σ_N, $\underline{\xi}$ is a Gaussian vector with a mean vector $\mu_N\underline{l}$ and covariance matrix $\underline{C} = \sigma_N^2\underline{R}$, in which $\underline{l}$ is a vector with k components that are all equal to one and $\underline{R}$ is the correlation matrix of $\xi(D)$. Then, $\underline{\xi}$ can be written as (e.g., [8,10,11,23])

$$\underline{\xi} = \mu_N\underline{l} + \sigma_N\underline{L}^T\underline{Z} \tag{5}$$

in which $\underline{Z}$ is a standard Gaussian random variable vector; $\underline{L}$ is a k-by-k upper-triangular matrix obtained by Cholesky decomposition of $\underline{R}$. The (i, j)th entry of $\underline{R}$ represents the correlation coefficient between $\xi(D_i)$ and $\xi(D_j)$ at respective depths of D_i and D_j, and it is given by a correlation function M, such as Eqs. (1)–(4). Selection of the correlation function of a soil property concerned is, therefore, a prerequisite for characterization of the inherent spatial variability of soil property. The next section presents a Bayesian model comparison approach to select the most probable correlation function among a pool of candidates (e.g., Eqs. (1)–(4)) for a given site using both project-specific test results and prior knowledge.

3. Bayesian model comparison

Let $\hat{\underline{\xi}} = \left[\hat{\xi}(D_1), \hat{\xi}(D_2), \ldots, \hat{\xi}(D_k)\right]^T$ be a set of $\xi(D)$ values measured at different depths $D_1, D_2, \ldots, D_k$, respectively. $\hat{\underline{\xi}}$ can be considered as a realization (or sample) of $\underline{\xi}$. As indicated by Eq. (5), $\hat{\underline{\xi}}$ contains the information on the possible correlation function (e.g., Eqs. (1)–(4)) of $\xi(D)$. Consider, for example, a number N_{CF} of possible correlation functions M_J, $J = 1, 2, \ldots, N_{CF}$ (e.g., $N_{CF} = 4$ for Eqs. (1)–(4)). For a given set of project-specific test results $\hat{\underline{\xi}}$, the plausibility of a correlation function M_J is defined by its occurrence probability $P(M_J|\hat{\underline{\xi}})$, $J = 1, 2, \ldots, N_{CF}$, conditional on $\hat{\underline{\xi}}$. The most probable correlation function M_J^* for a given $\hat{\underline{\xi}}$ has the maximum occurrence probability $P(M_J|\hat{\underline{\xi}})$. Hence, M_J^* can be determined by comparing the values of $P(M_J|\hat{\underline{\xi}})$ for the candidate correlation functions and selecting the one with the maximum value of $P(M_J|\hat{\underline{\xi}})$.

Using the Bayes' Theorem, $P(M_J|\hat{\underline{\xi}})$ is written as (e.g., [8,22,24–28])