Engineering Geology 210 (2016) 197–211

Contents lists available at ScienceDirect

Engineering Geology

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

Bayesian identification of random field model using indirect test data

Mi Tian [a], Dian-Qing Li [a], Zi-Jun Cao [a,*], Kok-Kwang Phoon [b], Yu Wang [c]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China
[b] Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07-03, 1 Engineering Drive 2, Singapore 117576, Singapore
[c] Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 27 September 2015
Received in revised form 23 May 2016
Accepted 26 May 2016
Available online 28 May 2016

Keywords:
Inherent spatial variability
Effective friction angle
Cone penetration test
Bayesian approaches
Markov Chain Monte Carlo Simulation
Gaussian copula

ABSTRACT

Inherent spatial variability (ISV) of design soil properties (e.g., effective friction angle φ') can be incorporated into probability-based geotechnical analyses and designs using random field models. Defining a random field model includes determination of random field parameters (i.e., mean μ, standard deviation σ, and scale of fluctuation λ) and the correlation function that specifies the spatial correlation of the concerned design soil property (e.g., φ') at different locations. This is, however, a challenging task at a given site due to a lack of direct test data of design soil properties and various uncertainties (e.g., transformation uncertainty) arising during site investigation. This paper develops Bayesian approaches for probabilistic characterization of the ISV of φ' using indirect test data (i.e., cone penetration test (CPT) data) and prior knowledge, which identify random field parameters of φ' through Markov Chain Monte Carlo Simulation (MCMCS) and, simultaneously, make use of Gaussian copula to select the most probable correlation function M^* among a pool of candidate correlation functions based on MCMCS samples. The proposed Bayesian approaches account, rationally and transparently, for the transformation uncertainty associated with the transformation model between φ' and CPT data. The proposed approaches are illustrated and validated using real-life and simulated CPT data. Results show that the proposed approaches properly identify the random field model (including μ, σ, λ, and M^*) of φ' using project-specific CPT data, and the random field parameters of φ' depend on the correlation function used to interpret CPT data. In addition, the suitability of MCMCS in Bayesian probabilistic characterization of soil properties is highlighted, particularly for the cases with a limited number of test data.

1. Introduction

Inherent spatial variability (ISV) of soils is one of major sources of uncertainties in soil properties (e.g., Phoon and Kulhawy, 1999a; Baecher and Christian, 2003; Wang et al., 2016). It can be incorporated into probability-based geotechnical analyses and designs through random field theory (e.g., Fenton and Griffiths, 2008; Vanmarcke, 2010; Gong et al., 2014; Jamshidi Chenari and Alaie, 2015; Li et al., 2015a, 2016a, 2016b). A random field model probabilistically characterizes the ISV through a set of random field parameters (i.e., mean μ, standard deviation σ, and scale of fluctuation λ) and a correlation function (such as those shown in Fig. 1) (e.g., Fenton, 1999; Fenton and Griffiths, 2008; Lloret-Cabot et al., 2014; Kasama and Whittle, 2016). Determining the random field parameters and the correlation function of design soil properties, which are directly used in geotechnical designs (e.g., effective friction angle φ'), at a site is, therefore, a necessary prerequisite for probabilistic characterization of ISV of soil properties at the site. This is, however, a challenging task in geotechnical practice.

Consider, for example, probabilistic characterization of the ISV of the effective friction angle φ'. Values of φ' in a soil layer can be directly measured from laboratory tests (e.g., triaxial tests) on soil samples retrieved from boreholes in a discrete manner. The number of direct measurements of φ' in a soil layer is usually too sparse to generate meaningful statistics and correlation function because a large number of laboratory tests are costly. On the other hand, φ' can be indirectly estimated using fast and economical in-situ tests (e.g., cone penetration test (CPT)) through transformation models (e.g., the empirical regression between normalized cone tip resistance q from CPT and φ', as shown in Fig. 2). The transformation model is not a perfect relationship but is associated with uncertainties/dispersion about a mean trend, namely "transformation uncertainty" (e.g., Phoon and Kulhawy, 1999b), which shall be rationally considered when using indirect test data (e.g., CPT data) to characterize the ISV of φ'. This can be formulated as an inverse analysis problem under a Bayesian framework (Wang et al., 2016).

Wang et al. (2010) and Cao and Wang (2013) proposed Bayesian approaches to inversely infer the random field model parameters (i.e., μ, σ,

* Corresponding author.
E-mail address: zijuncao@whu.edu.cn (Z.-J. Cao).

http://dx.doi.org/10.1016/j.enggeo.2016.05.013

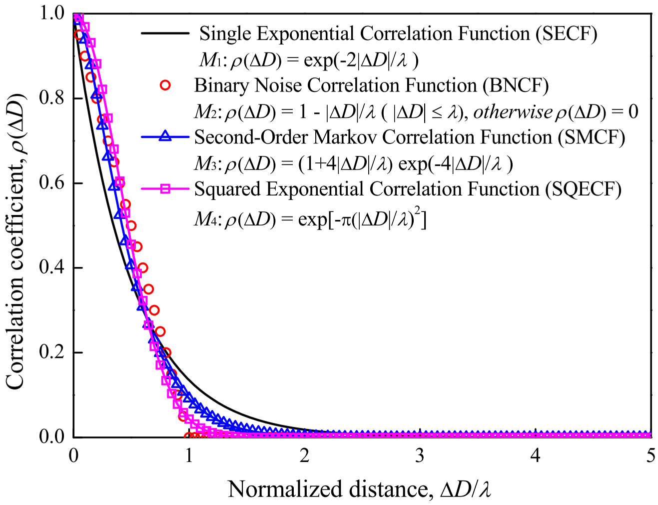

Fig. 1. Four commonly-used correlation functions in geotechnical engineering (After Phoon et al., 2003).

and λ) of φ' using CPT data, which account for the transformation uncertainty in an explicit and rational manner. However, these Bayesian approaches need to prescribe a correlation function before the analysis, which is unknown prior to site investigation. Cao and Wang (2014a) proposed a Bayesian approach to select a proper correlation function of q among a pool of candidate correlation functions, in which the random field parameters of q are treated as nuisance parameters and the transformation uncertainty is not involved. How to use a limited number of indirect test data (e.g., CPT data) to identify the random field parameters and simultaneously select an appropriate correlation function for probabilistic characterization of ISV of design soil properties (e.g., φ') at a specific site remains an outstanding challenge. In addition, effects of the correlation function on the random field parameters are also not clear.

This paper develops Bayesian approaches that identify the random field parameters (i.e., μ, σ, and λ) of φ' and simultaneously select the most probable correlation function of φ' among a pool of candidates

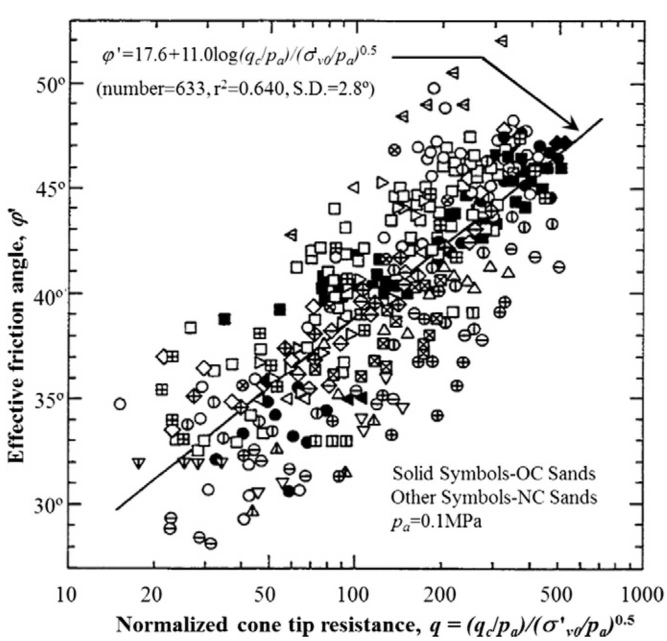

Fig. 2. Regression between effective friction angle and normalized cone tip resistance (After Kulhawy and Mayne, 1990; Phoon and Kulhawy, 1999b; Wang et al., 2010).

based on a limited number of project-specific CPT data and site information available prior to the project, namely "prior knowledge". The information from CPT data and prior knowledge is systematically integrated as the posterior knowledge on μ, σ, and λ under a Bayesian framework, which is quantitatively reflected by their posterior distribution. The computational complexity in solving the posterior distribution has been considered as one key limitation of Bayesian methods (e.g., Zhang et al., 2009). To bypass the computational complexity, the posterior distribution can be solved by Laplace asymptotic approximation method (LAAM) when it is well approximated by a Gaussian distribution (e.g., Wang et al., 2010; Cao and Wang, 2013; Ching et al., 2016). When there are only a limited number of test data, which is often the case in geotechnical engineering, the Gaussian approximation might not be valid. In such a case, Markov Chain Monte Carlo Simulation (MCMCS) provides a more appropriate tool to obtain the posterior knowledge in Bayesian analysis by generating random samples of model parameters concerned (e.g., μ, σ, and λ) from the posterior distribution (e.g., Zhang et al., 2010, 2012; Wang and Cao, 2013, Juang et al., 2013; Peng et al., 2014; Cao and Wang, 2014b; Kelly and Huang, 2015; Huang et al., 2016; Ching et al., 2016). Among various MCMCS algorithms, Metropolis–Hastings (M-H) algorithm (e.g., Metropolis et al., 1953; Hastings, 1970) is widely used for its simplicity. However, it has a key limitation that M-H algorithm does not give the likelihood of test data for a given model, which is often referred to as the "evidence" on the given model provided by test data in Bayesian model selection problems. This limitation makes the M-H algorithm infeasible in model selection problems (e.g., Bayesian selection of the most probable correlation function).

This paper removes the abovementioned limitations of M-H algorithm using Gaussian copula and selects the most probable correlation function among a pool of candidates (e.g., those shown in Fig. 1) based on MCMCS samples generated by M-H algorithm for candidate correlation functions. In addition, the proposed approaches also provide insights into effects of correlation functions on random field parameters. The paper starts with the development of the proposed Bayesian approaches, followed by a brief description of their implementation. Finally, the proposed approaches are illustrated and validated using real-life and simulated CPT data.

2. Bayesian identification of random field parameters

Random field theory (Vanmarcke, 2010) is used to explicitly model the ISV of φ' within a statistically homogenous sand layer in this study, by which φ' at different depths are modeled by a series of spatially correlated normal variables with a mean μ and standard deviation σ (i.e., a one-dimensional stationary normal random field). The spatial correlation between variations of φ' at different depths is then specified by the scale of fluctuation λ 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 Fig. 1. Note that the correlation function M is assumed to take a specific form (e.g., one of the correlation functions shown in Fig. 1) in this section, but its corresponding λ value is unknown herein. A proper form of the correlation function will be determined among a pool of candidates (e.g., those shown in Fig.1) by a Bayesian model selection approach in Section 3 entitled "Bayesian selection of spatial correlation function using MCMCS samples".

For a given correlation function M, the stationary normal random field of φ' is uniquely represented by the random field parameters $\boldsymbol{X}$, i.e., [μ, σ, λ]. For a given set of prior knowledge and CPT data $\underline{\hat{\xi}}$, there are various possible values of random field parameters, and their respective plausibility can be quantified by the posterior distribution $P(\boldsymbol{X}|\underline{\hat{\xi}}, M)$ under a Bayesian framework, where the condition on M indicates that the correlation function is assumed to take a specific form.