ARTICLE IN PRESS

Computers and Geotechnics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers and Geotechnics

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

Research Paper

Efficient method for probabilistic estimation of spatially varied hydraulic properties in a soil slope based on field responses: A Bayesian approach

Hao-Qing Yang[a,b], Lulu Zhang[a,b,*], Dian-Qing Li[c]

[a] State Key Laboratory of Ocean Engineering, Department of Civil Engineering, Shanghai Jiaotong University, Shanghai 200240, PR China
[b] Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration (CISSE), Shanghai 200240, PR China
[c] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China

ARTICLE INFO

Keywords:
Spatial variability
Back analysis
Slope
Karhunen–Loève, polynomial chaos
Markov Chain Monte Carlo

ABSTRACT

An efficient probabilistic back estimation method for characterization of spatial variability is proposed by integration of the Karhunen–Loève (K-L) expansion method, the Polynomial Chaos Expansion (PCE) method and the Markov Chain Monte Carlo (MCMC) method. To reduce the dimension of back estimation, the spatially varied soil property is simulated using the K-L expansion method and the basic random variables of K-L terms are parameters to be estimated. To further reduce computation load, a PCE surrogate model is constructed to substitute the original model. The proposed method is applied on an example where a randomly heterogeneous soil slope is subject to surface infiltration. The pressure responses are used to estimate the spatial variability of the saturated coefficient permeability. The results show that the spatial variability can be satisfactorily estimated. The coefficient of variation of the estimation is less than 5%.

1. Introduction

For an unsaturated soil slope, rainfall infiltration can diminish soil suction in the unsaturated zone, raise the ground water level, reduce soil shear strength and may consequently induce a slope failure. The soil hydraulic properties are the most important soil properties to affect the infiltration in a soil slope [1]. Accurate prediction of the slope performance under a rainfall requires reliable estimation of soil hydraulic properties in the slope.

Significant natural spatial variability is observed in field for soil properties [2–5]. This may induce the uncertainty of pore water pressure distribution and groundwater table in a slope [6,7]. Traditional methods to estimate spatial varied soil properties require large amount of samples from site investigation [8–10]. The accuracy of the estimation depends on both the total number and the spacing of samples. Field hydraulic response data, i.e., pore water pressure, ground water level, reflect the actual overall slope performance under a rainfall event and may provide more representative estimation of insitu soil hydraulic properties [11–13]. Therefore, estimation of spatial varied soil properties based on field responses using inverse methods might be an alternative approach to characterize field spatial variability.

Numerous studies have been conducted to estimate soil properties using deterministic or probabilistic inverse methods [13–18]. Some researchers utilized measured data from site investigation to characterize inherent spatial variability [19–23]. Limited studies focus on the characterization of spatial variability using field responses [24,25]. There are two reasons for this. First, spatial discretization of a randomly heterogeneous field can produce an extremely high-dimensional variable to be estimated, which will be a daunting task for inverse estimation. Secondly, most of inverse methods use iteration or sampling methods to pursue optimal solutions or stationary posterior distributions [12,13,26–28]. As numerical models are usually adopted for infiltration analysis in a soil slope, the computation cost of inverse estimation could be prohibitively large. Thus, inverse estimation of spatial varied soil hydraulic properties is a challenging task.

The objective of this study is to propose an efficient probabilistic estimation method for characterizing spatial variability of soil hydraulic properties based on field responses in a soil slope. The probabilistic back estimation is posed within a Bayesian framework and solved using MCMC simulation. To reduce the dimension of random variables in back estimation, the spatially varied soil hydraulic property is simulated using the Karhunen–Loève (K-L) expansion method. To reduce the computation load of probabilistic back analysis, the polynomial chaos expansion (PCE) is used as a surrogate model to approximate the deterministic numerical model. The procedures and the performance of the proposed method are illustrated using an example problem with a random heterogeneous slope which is subject to a rainfall along the slope surface.

* Corresponding author at: Department of Civil Engineering, Shanghai Jiaotong University, 800 Dongchuan Road, Shanghai 200240, PR China.
E-mail address: lulu_zhang@sjtu.edu.cn (L. Zhang).

https://doi.org/10.1016/j.compgeo.2017.11.012
Received 7 April 2017; Received in revised form 25 September 2017; Accepted 29 November 2017
0266-352X/

2. Probabilistic back estimation method of spatial varied soil property

2.1. Parameter estimation based on Bayesian theory

Consider a deterministic prediction model is established to simulate the performance of a soil slope under rainfall infiltration. The random variables of input soil parameters are represented by a random variable vector $\boldsymbol{\theta}$. The calculated hydraulic response of the model (e.g. pore pressure at a location) is $P(\boldsymbol{\theta})$. The field data obtained from the soil slope is the measured response at the same location, $\widehat{P}$. If the observations are at multiple locations or at different times, the observed data can be expressed as a vector, $\widehat{\mathbf{P}} = \{\widehat{P}_1,...,\widehat{P}_J\}$ in which J is the number of observations. The difference between the model output vector $\mathbf{P}(\boldsymbol{\theta})$ and the field observation vector is the residual error vector $\boldsymbol{\varepsilon} = \{\varepsilon_1,...,\varepsilon_i...,\varepsilon_J\}$ with

$$\varepsilon_i(\theta|\widehat{\mathbf{P}}) = P_i(\theta) - \widehat{P}_i \quad (1)$$

Due to inadequacies of the prediction model, errors of the initial and boundary conditions, parameter uncertainties and measurement errors, the residual values of the prediction model are not expected to be zero. Assuming the residuals are mutually independent and Gaussian-distributed with a constant variance σ_e^2, the likelihood function is [29]:

$$l(\theta|\widehat{\mathbf{P}}) = \prod_{i=1}^{J} \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{(P_i(\theta)-\widehat{P}_i)^2}{2\sigma_e^2}\right) \quad (2)$$

It should be noted that the residual errors can be non-Gaussian and biased with non-zero means. The likelihood formulation in Eq. (2) should be changed when different types of errors are defined. To keep the likelihood formulation simple and reasonably acceptable, the residual errors are assumed to be Gaussian and independently distributed with zero mean, just as in many previous studies [13,25,29–31]. Alternatively, discrepancy between the model prediction and that observation can be represented with a multiplicative model bias factor [32,33]. For prediction models with biased and non-Gaussian residual errors, more complicated formulations of model errors should be used [34,35].

Based on the Bayes' theorem [36], the posterior probability density function of $\boldsymbol{\theta}$ is proportional to the product of the likelihood function and the prior distribution function, and can be written as follows:

$$g(\theta|\widehat{\mathbf{P}}) = C\cdot g(\theta)\cdot l(\theta|\widehat{\mathbf{P}}) \quad (3)$$

where C is a normalizing constant, $g(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$.

For most geotechnical engineering problems, the posterior distribution function in Eq. (3) cannot be derived through analytical means or analytical approximation. Random sampling methods such as Monte Carlo simulation are therefore needed to generate samples from the posterior distribution function. In this study, the random sample generation from the posterior distribution is efficiently done using the Markov Chain Monte Carlo (MCMC) simulation and the differential evolution adaptive metropolis (DREAM) algorithm [37] is adopted. The convergence of the algorithm is monitored with the R criterion of Gelman and Rubin based on the within and between chain variance of each parameter. The convergence diagnostic R_{stat} value of less than 1.2 for each random variable is required to declare convergence to a stationary distribution. Details of the algorithm and the convergence criterion can be found in [37].

2.2. Dimensionality reduction with Karhunen–Loève expansion method

To estimate spatial varied soil properties using the probabilistic back analysis method in the previous section, the slope must be discretized spatially and each random variable of the random vector $\boldsymbol{\theta}$ represents the soil property at a given location. For instance, if the entire domain is 100 m × 40 m and be discretized into 50 × 20 cells. Therefore, the dimension of the random vector $\boldsymbol{\theta}$ is 1000. The extreme high dimensionality of the estimated parameter makes probabilistic back estimation of spatial varied soil properties a daunting task.

In this study, the Karhunen-Loève (K-L) expansion method [38,39] is used to simulate the field with spatially random soil properties and reduce the dimension of the random variables. Consider the spatial varied soil hydraulic property (e.g., saturated coefficient of permeability k_s), can be represented by a spatial random variable $U(\mathbf{x})$ with mean $\mu(\mathbf{x})$ and covariance function $C(\mathbf{x}_1,\mathbf{x}_2)$, where $\mathbf{x} \in \mathbf{D}$ is the coordinates in the physical domain $\mathbf{D}$. The covariance function $C(\mathbf{x}_1,\mathbf{x}_2)$ is symmetric and positive definite, where $\mathbf{x}_1 = (x_1, z_1)$ and $\mathbf{x}_2 = (x_2, z_2)$ represent the coordinates of the two points in domain $\mathbf{D}$.

The covariance function $C(\mathbf{x}_1,\mathbf{x}_2)$ can be decomposed as [40,41]

$$C(\mathbf{x}_1,\mathbf{x}_2) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(\mathbf{x}_1)\varphi_i(\mathbf{x}_2) \quad (4)$$

where λ_i and $\varphi_i(\mathbf{x})$ are the ith eigenvalues and eigenfunctions (eigenvectors) of the covariance function, respectively. The eigenvalues and eigenfunctions can be solved using the homogeneous Fredholm integral equation of the second kind [42].

Based the K-L expansion, the spatial random variable $U(\mathbf{x})$ can be expressed as [43]

$$U(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\,\theta_i\varphi_i(\mathbf{x}) \quad (5)$$

where $\mu(\mathbf{x})$ is the mean of $U(\mathbf{x})$; θ_i is the ith basic independent Gaussian random variable with $\theta_i \sim N(0,1)$.

The physical meaning of the K-L expansion is to separate the spatial variability on different spatial scales. Consider a two dimensional domain with size of 100 m × 40 m as an example. The domain is discretized into 50 × 20 square cells with a size of 2 m × 2 m. In this study, the covariance exponential function of $U(\mathbf{x})$ is adopted:

$$C(\mathbf{x}_1,\mathbf{x}_2) = \sigma^2 \exp\left\{-\left[\frac{(x_1-x_2)^2}{l_x^2} + \frac{(z_1-z_2)^2}{l_z^2}\right]^{\frac{1}{2}}\right\} \quad (6)$$

where σ^2 is the variance of $U(\mathbf{x})$; and l_x and l_z are the horizontal and vertical correlation length, respectively. Assume σ^2 is 1.0 and l_x is 50 m and l_z is 10 m. Fig. 1 shows the first 8 eigenfunctions φ_i of the covariance function. The eigenfunctions not only decrease in magnitude but also reduce in scale. It should be noted that the proposed method in this study is not restricted to one specific spatial correlation structure because the K-L expansion method can be applied to different covariance functions [20,41]. In this study, the exponential correlation structure is adopted because it is commonly adopted in many previous studies [6,44–47]. Other forms covariance functions can also be used. When enough field data are available, the field covariance function can be estimated using the least square method.

According to Eq. (5), any realization of a spatially varied property is a summation of an infinite number of eigenfunctions $\varphi_i(\mathbf{x})$ weighted by the product of $\sqrt{\lambda_i}$ and independently Gaussian random variable θ_i. Different K-L terms in Eq. (5) reflect the variability on different spatial scales. When a random sampling technique is used to generate realizations of the field, the randomness of any realization depends on the generated samples of Gaussian random variable θ_i. The total variability of $U(\mathbf{x})$ over the whole domain is distributed to all K-L terms with the weight of λ_i. Hence, the sum of all eigenvalues is related to the total variability of the random field [48].

The approximation of the actual random field can be obtained by truncating the ordered series in Eq. (5):

$$\widehat{U}(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^{n} \sqrt{\lambda_i}\,\theta_i\varphi_i(\mathbf{x}) \quad (7)$$

where n is the truncation level. Truncation of K-L expansion means ignoring the small-scale variation of the field. The truncation level is