Commun. Comput. Phys.
doi: 10.4208/cicp.090513.040414a

Vol. **16**, No. 4, pp. 1010-1030
October 2014

Conditional Simulation of Flow in Heterogeneous Porous Media with the Probabilistic Collocation Method

Heng Li*

Department of Energy and Resources Engineering, College of Engineering, Peking University, Beijing 100871, China.

Received 9 May 2013; Accepted (in revised version) 4 April 2014

Communicated by Jan S. Hesthaven

Available online 8 August 2014

Abstract. A stochastic approach to conditional simulation of flow in randomly heterogeneous media is proposed with the combination of the Karhunen-Loeve expansion and the probabilistic collocation method (PCM). The conditional log hydraulic conductivity field is represented with the Karhunen-Loeve expansion, in terms of some deterministic functions and a set of independent Gaussian random variables. The propagation of uncertainty in the flow simulations is carried out through the PCM, which relies on the efficient polynomial chaos expansion used to represent the flow responses such as the hydraulic head. With the PCM, existing flow simulators can be employed for uncertainty quantification of flow in heterogeneous porous media when direct measurements of hydraulic conductivity are taken into consideration. With illustration of several numerical examples of groundwater flow, this study reveals that the proposed approach is able to accurately quantify uncertainty of the flow responses conditioning on hydraulic conductivity data, while the computational efforts are significantly reduced in comparison to the Monte Carlo simulations.

AMS subject classifications: 60H35, 65C50, 65M70, 76S05

Key words: Conditional simulation, probabilistic collocation method, Karhunen-Loeve expansion, polynomial chaos expansion.

1 Introduction

It is well recognized that the geological formations normally exhibit spatial heterogeneity to certain degrees. On the other hand, our information about the formations is limited due to insufficient measurements. As such, the properties of geological formations such as the hydraulic conductivity are usually considered as random space functions, and the

*Corresponding author. *Email address:* `liheng@coe.pku.edu.cn` (H. Li)

equations describing flow and transport in the subsurface become stochastic. Extensive studies on flow and transport in random porous media have been conducted in the past and many stochastic approaches have been developed [4,6,7,20,22,29].

Uncertainties of the subsurface flow can be quantified through stochastic simulation of flow in random porous media, in that the statistical moments or probability density functions can be evaluated for the flow responses of interests. When some measurements of the hydraulic conductivity are prescribed in the process of stochastic simulation, it is called the conditional simulation [5,10,17,21]. The conditioning on the measured hydraulic conductivity can reduce the overall uncertainty of the hydraulic conductivity field, and that of responses of the flow and transport.

The most common approach to stochastic simulation of flow in porous media is the Monte Carlo method [1,29]. In this approach, a random field is represented by an ensemble of equally probable realizations. With each of the realizations as input, multiple flow simulations are performed independently, and statistical properties of flow responses can be evaluated. Monte Carlo method is straightforward to implement, either for unconditional or conditional simulation. However, it normally requires a large number of simulations to achieve statistical convergent results, thus it is computationally demanding, which prohibits its applications in large scale problems.

An alterative to the Monte Carlo method is the moment equation method, in which a system of deterministic differential equations governing the statistical moments (usually the first two) of the random variables are derived with the perturbation method or the closure approximation method [9,11,19,27,28,31]. However, the number of resulting deterministic equations in the moment equation method is dependent on the number of grid blocks in the numerical simulation, thus the computational cost of the this method is still high especially for large scale problems [29]. And it is limited to relatively small variance of hydraulic conductivity. The Karhunen-Loeve decomposition based moment equation approach (KLME) was developed for unconditional simulation of single phase flow in porous media [17], which was applied for groundwater flow and transport problems [2,3,16,26]. Lu and Zhang [17] extended the KLME to conditional simulation of flow in heterogeneous media, to incorporate the existing measurements of hydraulic conductivity.

The probabilistic collocation method (PCM) is another efficient stochastic approach [13,24]. It is based on the polynomial chaos expansions of random variables or fields, and a collocation technique is used to solve for the coefficients of the polynomial chaos expansions, which leads to uncoupled deterministic differential equations, similar to the governing flow equations. Li and Zhang [13] explored the PCM for single phase flow and showed its superiority compared to other stochastic approaches. With the PCM, a small number of flow simulations are performed independently, and the existing simulators can be employed. The method has been applied for efficient uncertainty quantification of unconfined groundwater flow [23], unsaturated and multiphase flows [14,15]. However in these studies, the PCM was only used for unconditional simulation.

In this study, an approach for conditional simulation of flow in randomly heterogeneous porous media is proposed, with the aid of the Karhunen-Loeve expansion and the probabilistic collocation method. The accuracy, efficiency, and applicability of the approach will be discussed by comparing with the Monte Carlo method through numerical examples.

2 Governing equations

Consider single phase water flow in porous media, which is governed by the following continuity equation and Darcy's law [29]:

$$S_s \frac{\partial h(\mathbf{x},t)}{\partial t} + \nabla \cdot \mathbf{q}(\mathbf{x},t) = g(\mathbf{x},t), \tag{2.1}$$

$$\mathbf{q}(\mathbf{x},t) = -K(\mathbf{x}) \nabla h(\mathbf{x},t), \tag{2.2}$$

subject to initial and boundary conditions

$$h(\mathbf{x},0) = H_0(\mathbf{x}), \quad x \in D, \tag{2.3}$$

$$h(\mathbf{x},t) = H(\mathbf{x},t), \quad x \in \Gamma_D, \tag{2.4}$$

$$\mathbf{q}(\mathbf{x},t) \cdot \mathbf{n}(\mathbf{x}) = Q(\mathbf{x},t), \quad x \in \Gamma_N, \tag{2.5}$$

where $K(\mathbf{x})$ is the hydraulic conductivity, $h(\mathbf{x},t)$ is hydraulic head, $\mathbf{q}(\mathbf{x},t)$ is the specific discharge (flux), and $g(\mathbf{x},t)$ is the source (or sink) term. $H_0(\mathbf{x})$ is the initial head in the domain D, $H(\mathbf{x},t)$ is the prescribed head on Dirichlet boundary segment Γ_D, $Q(\mathbf{x},t)$ is the prescribed flux across Neumann boundary segments Γ_N, $\mathbf{n}(\mathbf{x}) = (n_1,\cdots,n_d)^{\mathrm{T}}$ is an outward unit vector normal to the boundary $\Gamma = \Gamma_D \cup \Gamma_N$, and S_S is the specific storage.

The hydraulic conductivity $K(\mathbf{x})$ is spatially heterogeneous and it is treated as a spatial random field, whose covariance function and measurements at some locations are prescribed. As a result of the governing flow equations, the hydraulic head is no longer deterministic and its statistical properties need to be evaluated.

3 KL expansion of log hydraulic conductivity

3.1 Unconditional KL expansion

Suppose the hydraulic conductivity is a random field, and it can be written as a random space function $K(\mathbf{x},\theta)$, where $\mathbf{x} \in D$ and $\theta \in \Theta$ (a probability space). Assume that $Y(\mathbf{x},\theta) = \ln K(\mathbf{x},\theta)$ is a Gaussian random field. It can be decomposed to:

$$Y(\mathbf{x},\theta) = \langle Y(\mathbf{x}) \rangle + Y'(\mathbf{x},\theta), \tag{3.1}$$

where $\langle Y(\mathbf{x}) \rangle$ is the mean and $Y'(\mathbf{x},\theta)$ is the fluctuation. The covariance function $C_Y(\mathbf{x},\mathbf{y})$ is used to describe the spatial structure of the random field $Y(\mathbf{x},\theta)$:

$$C_Y(\mathbf{x},\mathbf{y}) = \langle Y'(\mathbf{x},\theta) Y'(\mathbf{y},\theta) \rangle. \tag{3.2}$$

Since the covariance is bounded, symmetric and positive-definite, it can be decomposed as [8]:

$$C_Y(\mathbf{x},\mathbf{y})=\sum_{n=1}^{\infty}\lambda_n f_n(\mathbf{x})f_n(\mathbf{y}), \tag{3.3}$$

where λ_n and $f_n(\mathbf{x})$ are eigenvalues and deterministic eigenfunctions, respectively, and can be solved from the following Fredholm equation:

$$\int_D C_Y(\mathbf{x},\mathbf{y})f(\mathbf{x})d\mathbf{x}=\lambda f(\mathbf{y}). \tag{3.4}$$

Then the random field $Y(\mathbf{x},\theta)$ can be expressed as

$$Y(\mathbf{x},\theta)=\langle Y(\mathbf{x})\rangle+\sum_{n=1}^{\infty}\sqrt{\lambda_n}f_n(\mathbf{x})\xi_n(\theta), \tag{3.5}$$

where $\xi_n(\theta)$ are orthogonal Gaussian random variables with zero mean and unit variance. The expansion (3.5) is called the Karhunen-Loeve (KL) expansion. The KL expansion, which is a spectral expansion, is optimal with mean square convergence when the underlying random field is Gaussian [8].

Generally, the eigenvalue problem of (3.4) needs to be solved numerically. However, analytical or semi-analytical solutions can be obtained under certain conditions. For a one-dimensional random field with a covariance function $C_Y(x_1,y_1)=\sigma_Y^2\exp(-|x_1-y_1|/\eta)$, where σ_Y^2 and η are the variance and the correlation length of the random field, respectively, the eigenvalues and their corresponding eigenfunctions can be expressed as [30]:

$$\lambda_n=\frac{2\eta\sigma_Y^2}{\eta^2\omega_n^2+1}, \tag{3.6}$$

and

$$f_n(x)=\frac{1}{\sqrt{(\eta^2\omega_n^2+1)L/2+\eta}}\left[\eta\omega_n\cos(\omega_n x)+\sin(\omega_n x)\right], \tag{3.7}$$

where ω_n are positive roots of the following equation:

$$(\eta^2\omega^2-1)\sin(\omega L)=2\eta\omega\cos(\omega L). \tag{3.8}$$

For problems in multi-dimension, if we assume that the covariance function $C_Y(\mathbf{x},\mathbf{y})$ is separable, for example $C_Y(\mathbf{x},\mathbf{y})=\sigma_Y^2\exp(-|x_1-y_1|/\eta_1-|x_2-y_2|/\eta_2)$ in a rectangular domain $D=\{(x_1,x_2):0\le x_1\le L_1;0\le x_2\le L_2\}$, the eigenvalues and eigenfunctions can be obtained by combining those in each dimension. For a non-separate covariance in a domain of arbitrary shape, the eigenvalue problem of (3.4) has to be solved numerically. Furthermore, the KL expansion is not limited to stationary random fields [18].

On the basis of the KL expansion (3.5), one can have $\sum_{n=1}^{\infty}\lambda_n=D\sigma_Y^2$, where D is the domain size. It means that the variance σ_Y^2 is decomposed by an infinite series of eigenvalues λ_n. If the roots ω_n of Eq. (3.8) are sorted in an increasing order, one can have the monotonically decreasing λ_n. Then the KL expansion can be truncated to finite terms by excluding the small eigenvalues. The number of terms retained in the KL expansion determines the random dimensionality of the problem, and it depends on the decay rate of the eigenvalues.

3.2 Conditional KL expansion

Conditional simulation of a Gaussian random field is based on the kriging technique [12]. For the Gaussian random field $Y(\mathbf{x})$. The measurements of $Y(\mathbf{x})$ at some locations $\mathbf{x}_i$ ($i=1,2,\cdots,M$) are known. The objective is to simulate $Y(\mathbf{x})$ at other locations without measurements. Let $Y^k(\mathbf{x})$ denote the kriging estimate of $Y(\mathbf{x})$. The simulated field of $Y(\mathbf{x})$ is expressed as follows,

$$Y^c(\mathbf{x})=Y^k(\mathbf{x})+\epsilon(\mathbf{x}), \tag{3.9}$$

where $\epsilon(\mathbf{x})$ is the estimation error at location $\mathbf{x}$, and the superscript c denotes "conditional".

The kriging estimate $Y^k(\mathbf{x})$ can be expressed as follows,

$$Y^k(\mathbf{x})=\langle Y(\mathbf{x})\rangle+\sum_{i=1}^{N}\alpha_i(\mathbf{x})(Y(\mathbf{x}_i)-\langle Y(\mathbf{x})\rangle), \tag{3.10}$$

where $\alpha_i(\mathbf{x})$ are the weighting functions for the kriging estimate, and can be obtained from the following equation,

$$\sum_{j=1}^{N}\alpha_j(\mathbf{x})C(\mathbf{x}_i,\mathbf{x}_j)=C(\mathbf{x}_i,\mathbf{x}),\quad i=1,2,\cdots,M. \tag{3.11}$$

The conditional mean and covariance of $Y^c(\mathbf{x})$ for the simple kriging can be derived as follows,

$$\langle Y^c(\mathbf{x})\rangle=\langle Y(\mathbf{x})\rangle+\sum_{i=1}^{M}\alpha_i(\mathbf{x})\left[(Y(\mathbf{x}_i)-\langle Y(\mathbf{x})\rangle\right], \tag{3.12}$$

$$C_Y^c(\mathbf{x},\mathbf{y})=C_Y(\mathbf{x},\mathbf{y})-\sum_{i,j=1}^{M}\alpha_i(\mathbf{x})\alpha_j(\mathbf{y})C(\mathbf{x}_i,\mathbf{y}_j). \tag{3.13}$$

The covariance function from Eq. (3.13) is nonstationary and the eigenvalues and eigenfunctions need to be solved from the following Fredholm equation (3.14) numerically,

$$\int_D C_Y^c(\mathbf{x},\mathbf{y})f^c(\mathbf{x})d\mathbf{x}=\lambda^c f^c(\mathbf{y}). \tag{3.14}$$

When the unconditional covariance function is of the separable exponential form, the conditional eigenvalues and eigenfunctions can be related to the corresponding unconditional ones [17]. Here we follow the algorithm of [17] for the special case with an unconditional separable exponential covariance function. Since the set of eigenfunctions is complete, one can expand $\alpha_i(\mathbf{x})$ in terms of the basis $\{f_n\}$, $\alpha_i(\mathbf{x})=\Sigma_{k=1}^{\infty}\alpha_{ik}f_k(\mathbf{x})$, where α_{ik} are coefficients to be determined. Substituting this expansion into Eq. (3.13), multiplying $f_m(\mathbf{x})$ on both sides, and integrating the derived equation with respect to $\mathbf{x}$ over D yields,

$$\sum_{i=1}^{M} C_Y(\mathbf{x}_i,\mathbf{x}_j)\alpha_{im}=\lambda_m f_m(\mathbf{x}_j), \quad j=1,2,\cdots,M; \quad m=1,2,\cdots. \tag{3.15}$$

If only N terms are retained in the KL expansion, all α_{im} in Eq. (3.15) will be obtained by just solving the $M\times M$ linear algebraic equations for N times. The computational cost is small since the number of conditioning points (M) is usually small.

The conditional eigenfunctions $f^c(\mathbf{x})$ can also be expanded with the unconditional eigenfunctions $f_n(\mathbf{x})$, $f^c(\mathbf{x})=\sum_{i=1}^{N} d_i f_i(\mathbf{x})$. Substituting this expansion and Eq. (3.13) into Eq. (3.14), multiplying $f_m(\mathbf{y})$ on the derived equation, and integrating it with respect to $\mathbf{y}$ over domain D, one obtains:

$$\lambda_m d_m - \sum_{k=1}^{N}\left(\sum_{i,j=1}^{M} C_Y(\mathbf{x}_i,\mathbf{x}_j)\alpha_{ik}\alpha_{jm}\right)d_k=\lambda^c d_m, \quad m=1,2,\cdots,N, \tag{3.16}$$

or in the matrix form:

$$(\mathbf{A}-\lambda^c\mathbf{E})\mathbf{d}=0, \tag{3.17}$$

where the components of $\mathbf{A}$ are $(a_{km})_{N\times N}$, $a_{km}=\lambda_m\delta_{km}-\sum_{i,j=1}^{M} C_Y(\mathbf{x}_i,\mathbf{x}_j)\alpha_{ik}\alpha_{jm}$, and $\mathbf{E}$ is an $N\times N$ identical matrix. The matrix $\mathbf{A}$ is symmetric since $C_Y(\mathbf{x}_i,\mathbf{x}_j)$ is symmetric. After solving the eigenvalue problem of an $N\times N$ matrix $\mathbf{A}$ through Eq. (3.17), one can obtain the conditional eigenvalues and eigenfunctions. The conditional eigenfunction corresponding to each conditional eigenvalue λ_n^c is constructed with the eigenvector $\mathbf{d}_n$,

$$f_n^c(\mathbf{x})=\sum_{i=1}^{N} d_{ni} f_i(\mathbf{x}). \tag{3.18}$$

In this way, the computational cost of finding the conditional eigenvalues and eigenfunctions is reduced in comparison to directly solving the conditional covariance function numerically.

Once the conditional eigenvalues and their corresponding eigenfunctions are obtained, the conditional log hydraulic conductivity field can be represented by the KL expansion,

$$Y^c(\mathbf{x},\theta)=\langle Y^c(\mathbf{x})\rangle+\sum_{n=1}^{N}\sqrt{\lambda_n^c} f_n^c(\mathbf{x})\xi_n(\theta). \tag{3.19}$$

4 Probabilistic collocation method

In this section, the probabilistic collocation method (PCM) used for uncertainty propagation of the flow equations is introduced. The outputs of flow simulations, e.g., the hydraulic head, are dependent on medium properties, i.e. the random hydraulic conductivity. While the covariance of the dependent random fields are yet to be found, the KL expansion cannot be used to represent their random structures. Instead, the PCE, introduced by Wiener (1938) and now widely used for uncertainty quantification, can be used to represent the dependent random fields. For example, we express the output random field $h(\mathbf{x},t)$ with the polynomial chaos expansion,

$$\begin{aligned} h(\mathbf{x},t,\theta) =& a_0(\mathbf{x},t) + \sum_{i_1=1}^{\infty} a_{i_1}(\mathbf{x},t)\Gamma_1(\xi_{i_1}(\theta)) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}(\mathbf{x},t)\Gamma_2(\xi_{i_1}(\theta),\xi_{i_2}(\theta)) \\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}(\mathbf{x},t)\Gamma_3(\xi_{i_1}(\theta),\xi_{i_2}(\theta),\xi_{i_3}(\theta)) + \cdots, \end{aligned} \tag{4.1}$$

where the coefficients $a_0(\mathbf{x},t)$ and $a_{i_1 i_2\cdots i_d}(\mathbf{x},t)$ are deterministic functions of $\mathbf{x}$ and t, and $\Gamma_d(\xi_{i_1},\cdots,\xi_{i_d})$ are orthogonal polynomial chaos of order d with respect to the random variables $(\xi_{i_1},\cdots,\xi_{i_d})$. For independent standard Gaussian random variables $(\xi_{i_1},\cdots,\xi_{i_d})$,

$$\Gamma_d(\xi_{i_1},\cdots,\xi_{i_d}) = (-1)^d e^{\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}} \frac{\partial^d}{\partial\xi_{i_1}\cdots\partial\xi_{i_d}}[e^{-\frac{1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi}}], \tag{4.2}$$

where $\boldsymbol{\xi}$ is a vector denoting $(\xi_{i_1},\cdots,\xi_{i_d})^T$. Hermite polynomials form the best orthogonal basis for Gaussian random variables [8]. In case of other random distributions, generalized polynomial chaos expansions [25] can be used to represent the random field.

In practice, Eq. (4.1) is usually truncated by finite terms, and the approximation $\hat{h}(\mathbf{x},t)$ can be written as

$$\hat{h}(\mathbf{x},t,\theta) = \sum_{j=1}^{P} c_j(\mathbf{x},t)\Psi_j(\boldsymbol{\xi}), \tag{4.3}$$

where $\boldsymbol{\xi}$ is a vector of dimension N. There is a one-to-one correspondence between the terms in Eqs. (4.1) and (4.3). The total number of terms $P = (N+d)!/N!d!$, where N is the random dimensionality and d is the degree of the polynomial chaos expansion. The P terms of coefficients are to be determined.

The general form of a stochastic differential equation can be expressed as:

$$\mathcal{L}h(\mathbf{x},t,\theta) = f(\mathbf{x},t), \tag{4.4}$$

where $h(\mathbf{x},t,\theta)$ is the unknown random space function and $f(\mathbf{x},t)$ is the source term. The operator $\mathcal{L}$ involves differentiations in space ($\mathbf{x}$) and time (t). In our problem, Eq. (4.4) represents the governing flow equations. Since $h(\mathbf{x},t,\theta)$ is approximated by the polynomial chaos expansion, with its approximation denoted as $\hat{h}(\mathbf{x},t,\theta)$, one can define the residual R as

$$R(\{c_i\},\boldsymbol{\xi}) = \mathcal{L}\hat{y} - f. \tag{4.5}$$

Relying on the weighted residual method in the random space, one can have

$$\int_{\xi} R(\{c_i\},\boldsymbol{\xi})w_j(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi}=0, \tag{4.6}$$

where $w_j(\boldsymbol{\xi})$ is the weighting function, $j=1,\cdots,P$, and $p(\boldsymbol{\xi})$ is the joint probability density function of $\boldsymbol{\xi}$.

In the probabilistic collocation method, the weighting function is chosen as the Dirac delta function,

$$w_j(\boldsymbol{\xi})=\delta(\boldsymbol{\xi}-\boldsymbol{\xi}_j), \tag{4.7}$$

where $\boldsymbol{\xi}_j$ is a particular set selected with a proper algorithm out of the random vector $\boldsymbol{\xi}$. The elements in $\boldsymbol{\xi}_j$ are called the collocation points. Then Eq. (4.6) becomes,

$$R\left(\{c_i\},\boldsymbol{\xi}_j\right)=0, \tag{4.8}$$

which results in a set of independent equations, evaluated at the given sets of collocation points, $\boldsymbol{\xi}_j$, where $j=1,2,\cdots,P$. It means that P sets of collocation points are required to solve for the P terms of coefficients $\{c_i\}$, where $i=1,2,\cdots,P$. The collocation points at a given order of polynomial chaos expansion can be selected from the roots of the next higher order orthogonal polynomial for each uncertain parameter. One can refer to [13] for the details regarding to the algorithm for the selection of the collocation points.

For our problem, the governing flow equations (2.1) and (2.2) combining with the KL expansion (3.19) can be expressed as,

$$S_s\frac{\partial h(\mathbf{x},t)}{\partial t}-\nabla\cdot\left\{\exp\left[\langle Y^c(\mathbf{x})\rangle+\sum_{n=1}^{N}\sqrt{\lambda_n^c}f_n^c(\mathbf{x})\xi_n\right]\nabla h(\mathbf{x},t)\right\}=g(\mathbf{x},t). \tag{4.9}$$

With the PCM, we only need to choose certain sets of collocation points $\{\boldsymbol{\xi}_j=(\xi_1,\xi_2,\cdots,\xi_N)_j,\ j=1,2,\cdots,P\}$, and solve Eq. (4.9) for the hydraulic head independently at each set of $\boldsymbol{\xi}_j$. Then the coefficients of the polynomial chaos expansion of the hydraulic head can be evaluated by solving a liner system of equations based on Eq. (4.3). The PCM is non-intrusive to the flow models since Eq. (4.9) has the same form as the original deterministic governing equations, and it can be solved independently, similar to the Monte Carlo method. As such, existing codes or flow simulators can be employed directly.

Once the coefficients of the polynomial chaos expansion (4.3) are obtained, the statistical quantities of the hydraulic head can be easily evaluated by sampling the random variables in the expansion. Since the expansion is already in an explicit form, the evaluation is computational efficient. Alternatively, the statistical moments of $h(\mathbf{x},t)$ such as the mean and variance can be directly derived from Eq. (4.3),

$$\langle h(\mathbf{x},t)\rangle=c_1(\mathbf{x},t), \tag{4.10}$$

$$\sigma_h^2=\sum_{j=2}^{P}c_j(\mathbf{x},t)^2\langle\Psi_j^2\rangle. \tag{4.11}$$

The procedure of the KL-based PCM for conditional simulation is straightforward and the main steps are summarized as follows: 1) Representing the conditional log hydraulic conductivity field using the KL expansion (3.19) in terms of a set of independent Gaussian random variables; 2) approximating the hydraulic head with the polynomial chaos expansion in terms of the set of independent Gaussian random variables with Eq. (4.3); 3) determining the set deterministic coefficients of the polynomial chaos expansion using the PCM technique; 4) evaluating the statistical properties of the hydraulic head based on the constructed polynomial chaos expansion.

5 Case studies

This and the following sections show some examples to illustrate the KL-based PCM for conditional simulation and compare it with the Monte Carlo method. Different cases of one-dimensional flow in random porous media are used to analyze the effect of various factors, such as the effect of conditioning, correlation length, and spatial variability. Finally a two-dimensional case is conducted to demonstrate its applicability for large scale problems.

5.1 Effect of conditioning

We first consider steady state groundwater flow in a one-dimensional heterogeneous medium of length $L=10$ [L] (where [L] denotes any consistent length unit) and assume the forcing term to be zero. The boundary conditions are prescribed hydraulic heads at the two ends, $H_0=7$ [L] and $H_L=5$ [L]. The log hydraulic conductivity, $Y(x)=\ln K(x)$, is a Gaussian random field. The unconditional mean of the log hydraulic conductivity is given as $\langle Y\rangle=0$. The covariance of the unconditional $Y(x)$ is assumed to have the exponential form, $C_Y(x_1,y_1)=\sigma_Y^2\exp(-|x_1-y_1|/\eta)$.

Case 1 is for unconditional simulation, where no conditional data are used for simulation. Here the results of Case 1 are used to compare with conditional simulations and demonstrate the effect of conditioning. In Case 1, the correlation length $\eta=4$, and the variance of log conductivity is $\sigma_Y^2=1.0$ (corresponding to the coefficient of variation of hydraulic conductivity CV_K as 131%). Fig. 1 shows the variance of hydraulic head at different locations. Both the PCM and Monte Carlo (MC) method are performed. Due to rapid decay of the eigenvalues, 6 terms are retained in the KL expansion used to represent the unconditional log hydraulic conductivity field, and there are only 28 simulations involved for the 2nd order PCM. It can be seen from Fig. 2 that the preserved energy of retained eigenvalues $\Sigma_{n=1}^{N}\lambda_n/\Sigma_{n=1}^{\infty}\lambda_n$ is about 92% when 6 terms of eigenvalues are retained for the unconditional case. Fig. 1 shows that the results from the PCM (with 28 simulations) agree very well with those from MC (with 10,000 simulations).

Next, conditional data are added to the flow configuration of Case 1 to conduct conditional simulation. There are three available measurements at three spatial points: $x=1$, $x=3$, and $x=7$, respectively. The effect of different combinations of the conditioning data

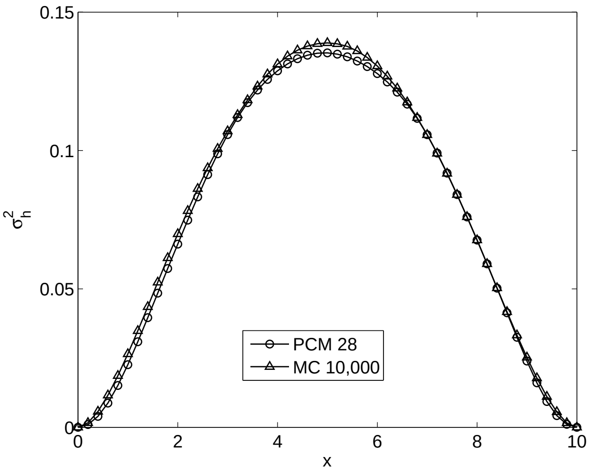


Figure 1: Variance of hydraulic head for the unconditional case with $\sigma_Y^2=1$.

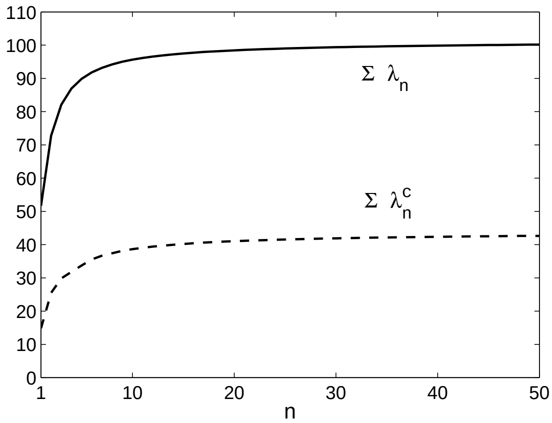


Figure 2: Summation of unconditional and conditional eigenvalues for $\eta=4$.

is explored in the following cases. Case 2 (a) is for conditioning points at $x=1$ and $x=7$, Case 2 (b) is for $x=1$ and $x=3$, Case 2 (c) is for $x=3$ and $x=7$, and Case 2 (d) is for $x=1$, $x=3$, and $x=7$.

For the conditional PCM, we first solve for the unconditional eigenvalues and eigenfunctions. Then the conditional eigenvalues and eigenfunctions are obtained with the algorithm described in Section 3.2. Fig. 2 shows the summation of unconditional eigenvalues and that of the conditional eigenvalues for η=4, where the conditional eigenvalues are for Case 2(a). It can be seen that the summation of all conditional eigenvalues is much less than that of the unconditional case, indicating a smaller variability of the conditional $Y(x)$. When 6 terms are retained in the KL expansion of the conditional $Y(x)$, the pre-

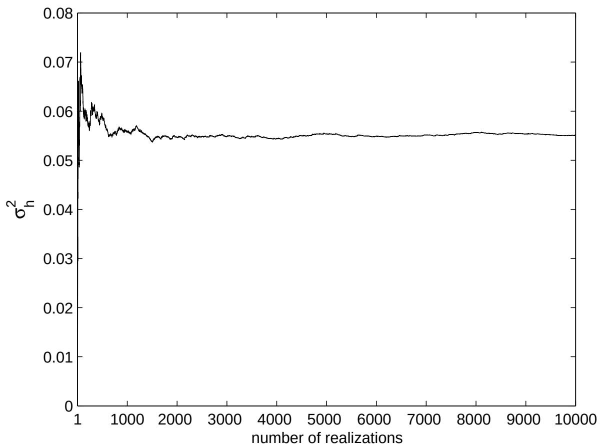


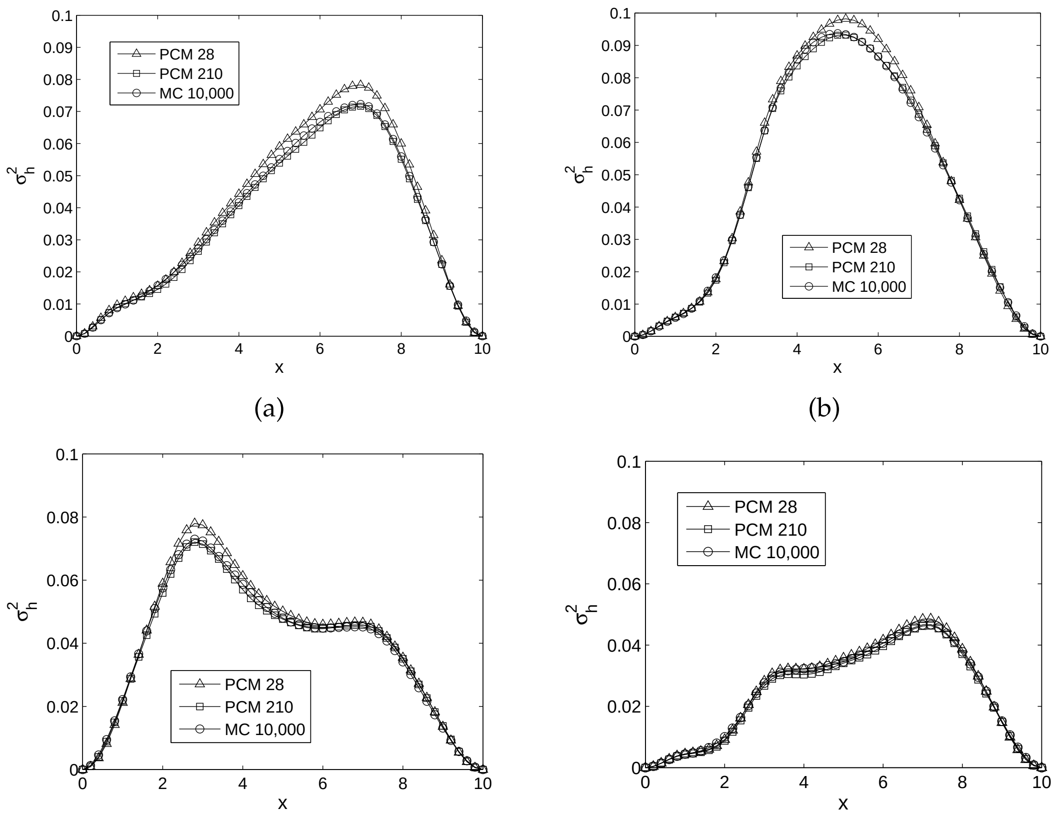
Figure 3: The variance of hydraulic head at $x=5$ as a function of the number of realizations for Monte Carlo simulations for Case 2 (a).

served energy of retained eigenvalues is about 83%. After representing the conditional hydraulic conductivity with KL expansion, the governing flow equations are solved at the specified collocation points through the PCM. Both the 2nd order PCM and 4th order PCM are performed and compared with the MC to investigate the accuracy. The 2nd order PCM requires 28 simulations, and the 4th order PCM requires 210 simulations.

Monte Carlo (MC) simulations are performed for comparison. Since the Monte Carlo simulations are used as benchmark, it is necessary to check the convergence rate of the Monte Carlo simulations. Fig. 3 shows the variance of hydraulic head at some specific location ($x=5$) as a function of the number of realizations used for Monte Carlo simulations in Case 2 (a). It shows that 10,000 realizations are enough to ensure statistical convergence, thus MC (10,000) is used as benchmark. In our MC simulation, the KL expansion is implemented to generate the realizations of the conditional log hydraulic conductivity $Y(x)$, based on Eq. (3.19). To ensure enough accuracy for reproducing the ensemble statistics of $Y(x)$, 50 terms are retained in Eq. (3.19) for the MC simulations. It should be noted that the number of realizations for the MC simulations is independent of the number of terms retained in the KL expansion.

Fig. 4 (a)-(d) show the variance of hydraulic head for Cases 2 (a)-(d), respectively. In Fig. 4 (a)-(c), where only 2 conditioning points are used, the 2nd order PCM (with 28 simulations) has slight deviations from the Monte Carlo solutions. However, the results from the 4th order PCM (210 simulations) are identical to the MC solutions. After all three conditioning points are taken into consideration, the overall variance of hydraulic head is reduced in Fig. 4 (d), compared to 4 (a)-(c). While the uncertainty is reduced in Fig. 4 (d), both the 2nd order PCM (28 simulations) and the 4th order PCM (210 simulations) agree well with the MC solutions.

The comparison of computational cost and solution accuracy can be made for the



(a) (b) (c) (d)

Figure 4: Variance of hydraulic head for the case with $\sigma_Y^2 = 1.0$ and $\eta = 4$. The conditioning points are at (a) $x = 1$ and $x = 7$; (b) $x = 1$ and $x = 3$; (c) $x = 3$ and $x = 7$; (d) $x = 1$, $x = 3$, and $x = 7$.

PCM and MC. Both the PCM and MC involve solving the governing flow equations independently, and the computational cost for each PCM and MC simulation is almost the same, regardless of the number of terms retained in the KL expansion. Therefore, the difference of computational efficiency for the PCM and MC can be determined by the number of simulations involved. Fig. 5 shows the results of three sets of MC (1,000 simulations) for Case 2 (a). It can be seen that the three sets of MC solutions diverge, thus 1,000 realizations are not enough for this case. That is, different MC simulations yield different solutions when the number for realizations are not enough and the statistical convergence is not ensured. However, the PCM (with 28 or 210 simulations) yields robust results close to the benchmark solutions. For this case with moderate spatial variability, the 2nd order PCM with only 28 simulations can obtain satisfactory results. As the order is increased, the results are more accurate. For the PCM, how to choose a certain order for a specific problem is an undergoing research topic, and it is out of the scope of this paper.

The effect of conditioning can be found by comparing Fig. 1 with Fig. 4. As shown

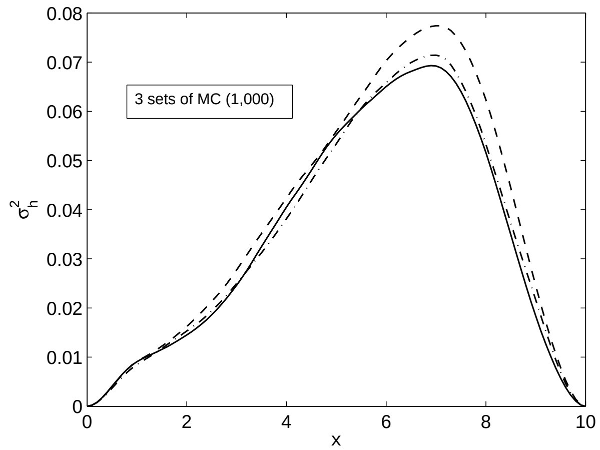


Figure 5: Variance of hydraulic head obtained from 3 sets of MC (with 1,000 simulations) for Case 2 (a).

in Fig. 1, for the unconditional simulation in Case 1, the variance of the hydraulic head is symmetric about the center of the domain due to the specific boundary conditions. However, when conditioning points are considered, the symmetry of head variance is changed. The changes of head variance are most prominent around the conditioning points, which would have significant impact on the entire variance profile. This is due to the change of uncertainty in the conditional hydraulic conductivity, especially the abrupt change around the conditioning points.

5.2 Effect of correlation length

To further test the effect of correlation length on the PCM, Case 3 (a)-(d) are performed for a smaller correlation length $\eta=2$. The conditions of Case 3 (a)-(d) are the same as Case 2 (a)-(d), respectively, except for the correlation length η. Note that the prescribed values of hydraulic conductivity at the conditioning points are different from those in Case 2. They are generated from one unconditional realization of $Y(x)$ with variance $\eta=2$. Fig. 6 shows the summation of unconditional eigenvalues and that of the conditional eigenvalues for $\eta=2$, where the conditional eigenvalues are for Case 3(a). Comparing Fig. 6 with Fig. 2, one can find that the rate of decay in the eigenvalues is dependent on the correlation length η relative to the domain length L. In Fig. 6, the preserved energy of retained eigenvalues $\Sigma_{n=1}^{N}\lambda_n/\Sigma_{n=1}^{\infty}\lambda_n=83\%$ and 78% for the unconditional and conditional case, respectively. We still choose 6 terms retained in the KL expansion for $\eta=2$, the same number as in the cases of $\eta=4$. The head variances obtained from the 2nd order PCM, 4th order PCM, and the MC are presented in Fig. 7. In terms of accuracy, both the 2nd order PCM (with 28 simulations) and 4th order PCM (with 210 simulations) have satisfactory results compared to MC (with 10,000 simulations), while the 4th order PCM is more accurate than the 2nd order PCM. By comparing Fig. 4 with Fig. 7, one can also

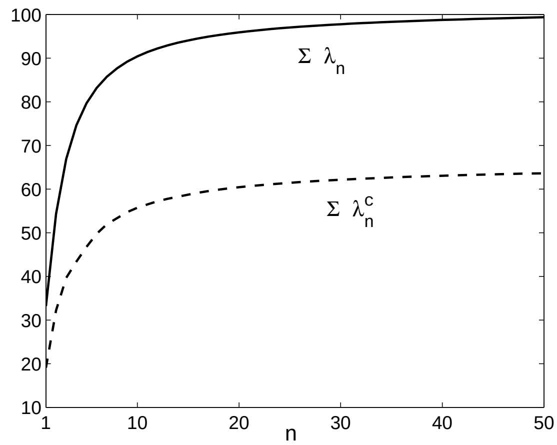


Figure 6: Summation of unconditional and conditional eigenvalues for $\eta = 2$.

observe that the profiles of head variance are different when correlation length of the log hydraulic conductivity field is different.

5.3 Effect of spatial variability

The effect of the spatial variability of hydraulic conductivity is investigated in Case 4 (a)-(d) and Case 5 (a)-(d). We modify Case 2 (a)-(d) by increasing the spatial variability of $Y(x)$ to $\sigma_Y^2 = 2.0$, or $\sigma_Y^2 = 4.0$, corresponding to the coefficient of variation of hydraulic conductivity (CV_K) as 253% or 732%, respectively. $\sigma_Y^2 = 2.0$ is considered in Case 4 (a)-(d), which correspond to the same combinations of conditioning points in Case 2 (a)-(d), respectively. Fig. 8 (a)-(d) show the variance of hydraulic head for Case 4 (a)-(d), respectively. Note that the values of hydraulic conductivity at the conditioning points are different from those in Case 2. They are generated from one unconditional realization of $Y(x)$ with variance $\sigma_Y^2 = 2.0$. Both the PCM and MC are performed. The results of MC (with 10,000 simulations) are also used as benchmark solutions. The 4th order PCM (with 210 simulations) still agrees well with the MC for each case, while the 2nd order PCM (with 28 simulations) has some deviations. When spatial variability of $Y(x)$ becomes larger, so does the resulting uncertainty of hydraulic head (σ_h^2).

Case 5 (a)-(d) are conducted for $\sigma_Y^2 = 4.0$. New values of Y at the three conditioning points are generated from one unconditional realization of $Y(x)$ with $\sigma_Y^2 = 4.0$, and used for conditional simulation. Fig. 9 (a)-(d) show the variance of hydraulic head for Case 5 (a)-(d), respectively. The magnitude of σ_h^2 is increasing, compared to Case 2 (a)-(d) and Case 4 (a)-(d). For these cases where the spatial variability of $Y(x)$ is huge ($CV_K =$ 732%), the 2nd order PCM (28) yields larger deviations against the MC (10,000) solutions. However, the 4th order PCM (210) can still obtain good results compared to the MC (10,000) solutions, with smaller computational efforts.

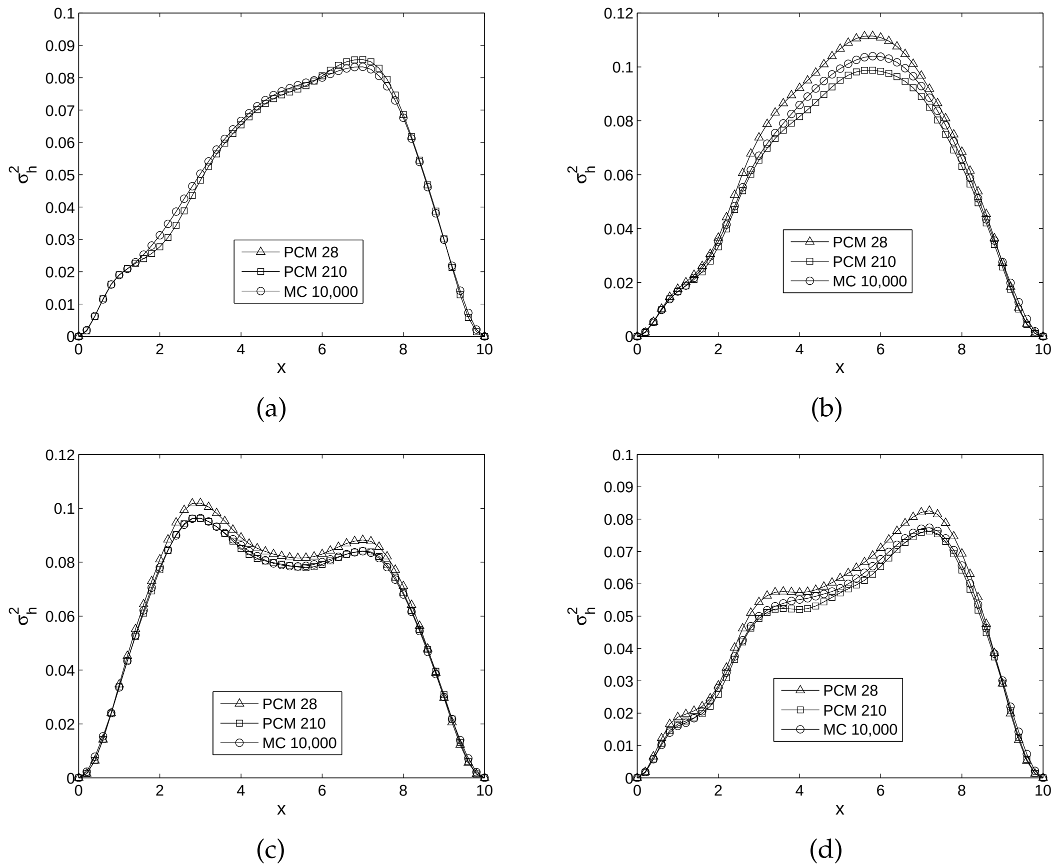


Figure 7: Variance of hydraulic head for the case with $\sigma_Y^2 = 1.0$ and $\eta = 2$. The conditioning points are at (a) $x = 1$ and $x = 7$; (b) $x = 1$ and $x = 3$; (c) $x = 3$ and $x = 7$; (d) $x = 1$, $x = 3$, and $x = 7$.

5.4 Illustrative example in 2D

Case 6 is designed to illustrate the applicability of PCM for two-dimensional flow. The two-dimensional domain of saturated heterogeneous medium is a square of size $L_1 = L_2 = 10$ $[L]$, uniformly discretized into 40×40 square elements. The non-flow conditions are prescribed at two lateral boundaries. The hydraulic head is prescribed at the left and right boundaries as 10.5 [L] and 10.0 [L], respectively. The mean of the log hydraulic conductivity is given as $\langle Y \rangle = 0$. Assume the covariance function of the unconditional log hydraulic conductivity is $C_Y(\mathbf{x},\mathbf{y}) = C_Y(x_1,x_2;y_1,y_2) = \sigma_Y^2 \exp(-|x_1 - y_1|/\eta_1 - |x_2 - y_2|/\eta_2)$, where $\sigma_Y^2 = 1.0$ and $\eta_1 = \eta_2 = 4.0$. There are 9 conditioning points regularly distributed in the square domain. The locations of these conditioning points are shown in Fig. 10. The prescribed log hydraulic conductivity values at these conditioning points are generated from one unconditional realization of the log hydraulic conductivity field.

After conditioning to the 9 points, the conditional mean and variance of the hydraulic

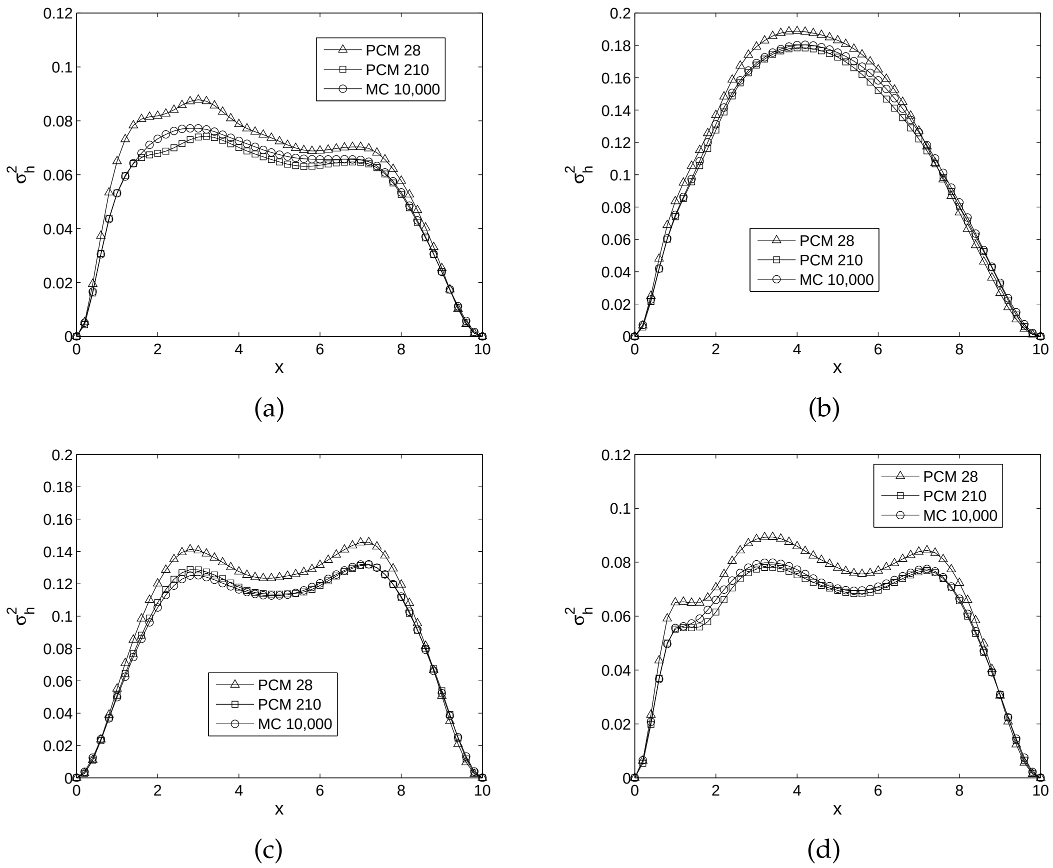


Figure 8: Variance of hydraulic head for the case with $\sigma_Y^2=2.0$. The conditioning points are at (a) $x=1$ and $x=7$; (b) $x=1$ and $x=3$; (c) $x=3$ and $x=7$; (d) $x=1$, $x=3$, and $x=7$.

conductivity are shown in Fig. 11 (a) and (b), respectively. It can be seen obviously that the variance of hydraulic conductivity after conditioning becomes zero at the 9 conditioning points. The overall uncertainty of the hydraulic conductivity field is reduced. Both the PCM and MC are performed for uncertainty quantification of the hydraulic head in the whole domain. For the 2nd order PCM, 20 terms are retained in the KL expansion where the preserved energy of the conditional eigenvalues is about 71%, thus the total number of collocation points is 231. Fig. 12 shows the comparisons of the conditional mean head and head variance in the two-dimensional domain, obtained from PCM (with 231 simulations) and MC (with 10,000 simulations). Fig. 13 shows the comparison of head variance along the profile $x_2=5$. It can be seen that the conditional mean head of the PCM is identical to that of the MC, and the results of head variance for PCM and MC are also in good agreement with each other.

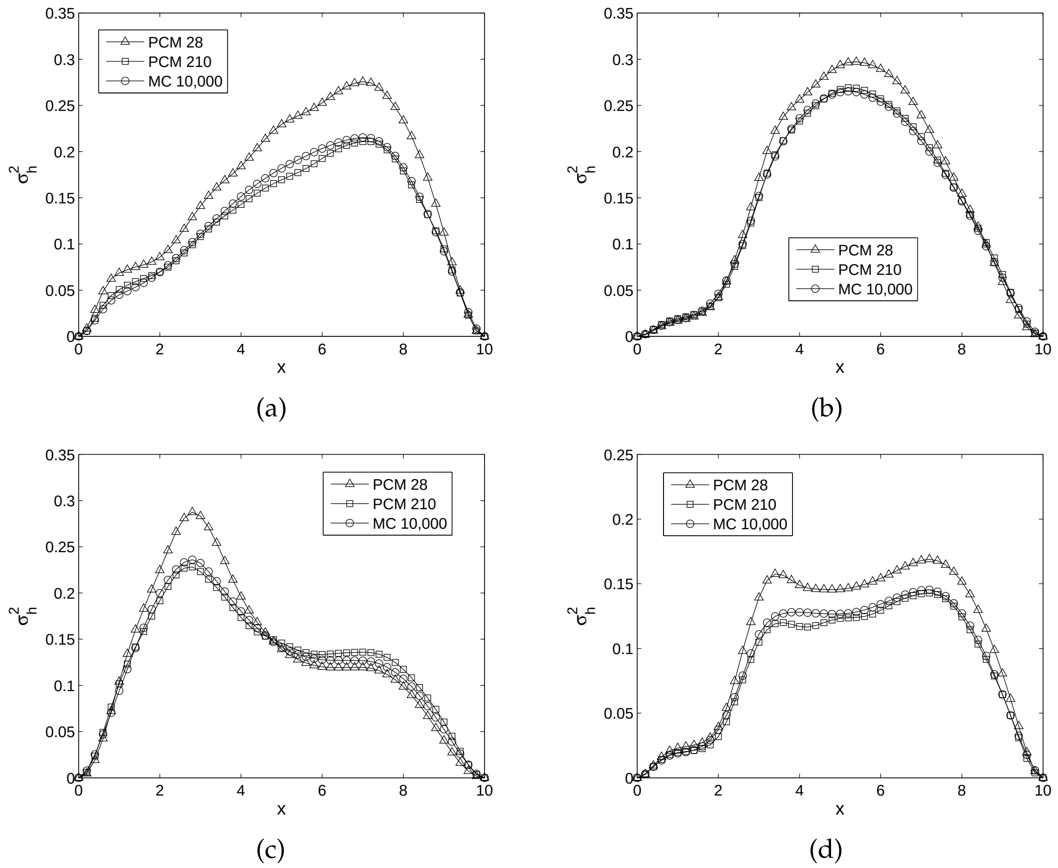


Figure 9: Variance of hydraulic head for the case with $\sigma_Y^2 = 4.0$. The conditioning points are at (a) $x=1$ and $x=7$; (b) $x=1$ and $x=3$; (c) $x=3$ and $x=7$; (d) $x=1$, $x=3$, and $x=7$.

6 Summary and conclusions

A stochastic approach has been proposed for conditional simulation of flow in random porous media by combining the Karhunen-Loeve (KL) expansion and the probabilistic collocation method (PCM). After incorporating the measurements of log hydraulic conductivity with the kriging technique, the conditional mean and covariance of the log conductivity field are obtained, by which the KL expansion is used to represent the log hydraulic conductivity field. When the unconditional covariance function is of the separable exponential form, the conditional eigenvalues and eigenfunctions can be computed from their unconditional counterparts directly. Once the conditional eigenvalues and eigenfunctions are obtained, the KL expansion provides an effective way to generate conditional realizations and parameterize the conditional random field through a set of independent Gaussian random variables. The flow response such as the hydraulic head

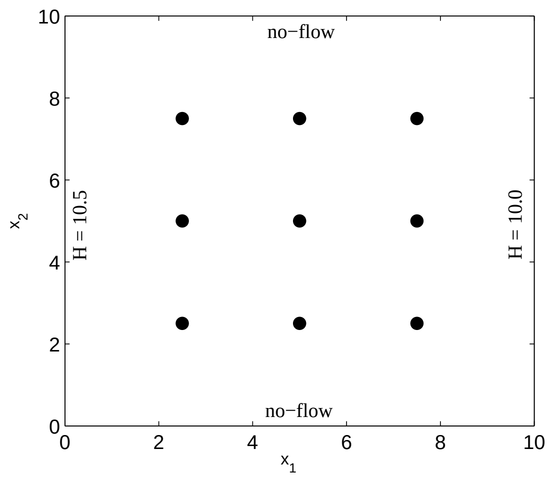


Figure 10: Problem configuration of Case 6. The solid circles are conditioning points.

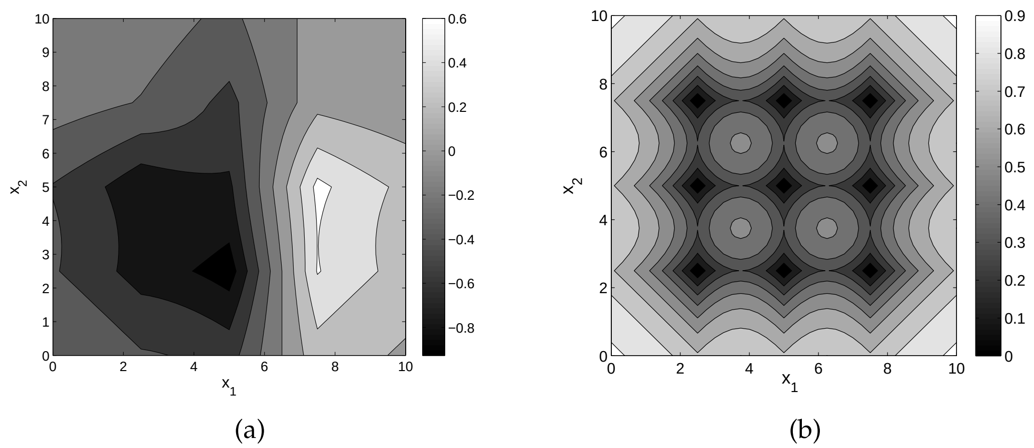


Figure 11: (a) Conditional mean of log hydraulic conductivity; (b) conditional variance of log hydraulic conductivity for Case 6.

is expressed with the polynomial chaos expansion, whereby the PCM can be employed for efficient uncertainty quantification.

The PCM is non-intrusive to the flow models in that it results in independent differential equations similar to the governing flow equations. Therefore, existing flow simulators or codes could be employed in a straightforward manner for conditional simulation. Our numerical examples reveal that with a smaller number of flow simulations, the PCM can achieve a good agreement with the Monte Carlo solutions that are involved with a large number of simulations. It is found that for moderate variability of hydraulic conductivity, the 2nd order PCM is sufficient to obtain satisfactory results. When the variance

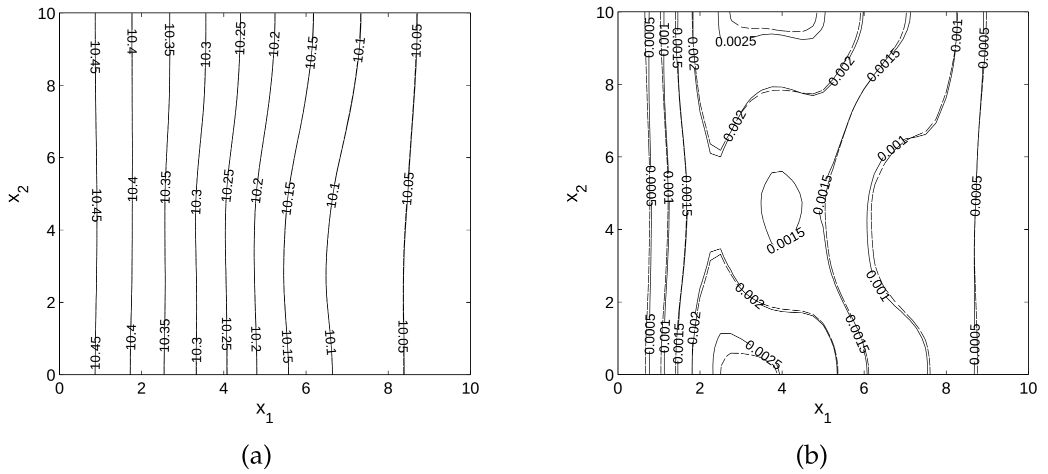


Figure 12: Comparisons of (a) conditional mean head and (b) conditional head variance for Case 6 obtained from PCM (231 simulations) and MC (10,000 simulations): solid curves for PCM and dashed curves for MC.

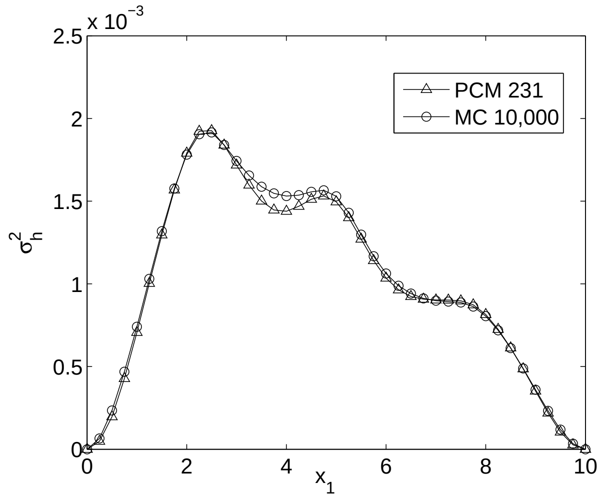


Figure 13: Variance of hydraulic conductivity along the profile $x_2 = 5.0$ for Case 6 computed from the PCM (231 simulations) and MC (10,000 simulation)

of log conductivity becomes large (e.g, with the coefficient of variation of the hydraulic conductivity being larger than 253%), a high order of PCM may be needed to achieve higher accuracy. The conditioning on the measured hydraulic conductivity can reduce the overall uncertainty of the hydraulic conductivity field, and that of flow response, i.e., the hydraulic head. As more conditioning points are taken into consideration, the uncertainty would be further reduced.

Acknowledgments

This work is partially funded by the National Science and Technology Major Project of China through Grants 2011ZX05009-006 and 2011ZX05052, by the National Key Technology R&D Program of China (Grant No. 2012BAC24B00), and by the National Natural Science Foundation of China (Grant No. 51204008).

References

[1] F. Ballio and A. Guadagnini. Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. Water Resour. Res., 40, 2004. W04603, doi:10.1029/2003WR002876.

[2] M. Chen, A. A. Keller, D. Zhang, Z. Lu, and G. A. Zyvoloski. Stochastic analysis of transient two-phase flow in heterogeneous porous media. Water Resour. Res., 42, 2006. W03425, doi:10.1029/2005WR004257.

[3] M. Chen and D. Zhang. Stochastic analysis of two phase flow in heterogeneous media by combining Karhunen-Loeve expansion and perturbation method. Water Resour. Res., 41, 2005. W01006, doi:10.1029/2004WR003412.

[4] J. H. Cushman. The Physics of Fluids in Hierarchical Porous Media: Angstroms to Miles. Kluwer Academic, Norwell, MA., 1997.

[5] G. Dagan. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 1. Conditional simulation and the direct problem. Water Resour Res, 18(4), 1982, 813–33.

[6] G. Dagan. Flow and Transport in Porous Formations. Springer, New York, 1989.

[7] L. W. Gelhar. Stochastic Subsurface Hydrology. Prentice-Hall, Englewood Cliffs, NJ., 1993.

[8] R. Ghanem and P. Spanos. Stochastic Finite Element: A Spectral Approach. Springer, New York, 1991.

[9] W. D. Graham and D. McLaughlin. Stochastic analysis of nonstationary subsurface solute transport: 1. Unconditional moments. Water Resour. Res., 25(2), 1989, 215–232.

[10] W. D. Graham and D. McLaughlin. Stochastic analysis of nonstationary subsurface solute transport: 2. Conditional moments. Water Resour. Res., 25(11), 1989, 2331–55.

[11] A. Guadagnini and S. P. Neuman. Nonlocal and localized analysis of conditional mean steady state flow in bounded, randomly nonuniform domains: 1. Theory and computational approach. Water Resour. Res., 35(10), 1999, 2999–3018.

[12] D. G. Krige. A statistical analysis of some mine valuation and allied problems on the witwatersrand. Master's thesis, University of Witwatersrand, 1951.

[13] H. Li and D. Zhang. Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. Water Resour. Res., 43, 2007. W09409, doi:10.1029/2006WR005673.

[14] H. Li and D. Zhang. Efficient and accurate quantification of uncertainty for multiphase flow with the probabilistic collocation method. SPE Journal, 14(4), 2009, 665–679.

[15] W. Li, Z. Lu, and D. Zhang. Stochastic analysis of unsaturated flow with probabilistic collocation method. Water Resour. Res., 45, 2009. W08425, doi:10.1029/2008WR007530.

[16] G. Liu, Z. Lu, and D. Zhang. Stochastic uncertainty analysis for solute transport in randomly heterogeneous media using a Karhunen-Loeve based moment equation approach. Water Resour. Res., 43, 2007. W07427, doi:10.1029/2006WR005193.

[17] Z. Lu and D. Zhang. Conditional simulations of flow in randomly heterogeneous porous media using a KL-based moment-equation approach. Adv. Water Res., 27, 2004, 859–874.

[18] Z. Lu and D. Zhang. Stochastic simulations for flow in nonstationary randomly heterogeneous porous media using a KL-based moment-equation approach. SIAM Multiscale Model. Simul., 6(1), 2007, 228–245.

[19] S. P. Neuman. Eulerian-Lagrangian theory of transport in spacetime nonstationary velocity fields: Exact nonlocal formalism by conditional moments and weak approximations. Water Resour. Res., 29, 1993, 633–645.

[20] S. P. Neuman. Stochastic approach to subsurface flow and transport: A view to the future. In G. Dagan and S. P. Neuman, editors, Subsurface Flow and Transport: A Stochastic Approach, pages 231–241. Cambridge University Press, New York, 1997.

[21] Y. Rubin. Prediction of tracer plume migration in disordered porous media by the method of conditional probabilities. Water Resour Res, 27(6), 1991, 1291–1308.

[22] Y. Rubin. Applied Stochastic Hydrogeology. Oxford University Press, New York, 2003.

[23] L. Shi, J. Yang, D. Zhang, and H. Li. Probabilistic collocation method for unconfined flow in heterogeneous media. J. Hydrol., 365, 2009, 4–10.

[24] M. A. Tatang, W. Pan, R. G. Prinn, and G. J. McRae. An efficient method for parametric uncertainty analysis of numerical geophysical models. J. Geophy. Res., D18, 1997, 21925–21931.

[25] D. Xiu and G. E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. Comput. Methods Appl. Mech. Eng., 191, 2002, 4927–4948.

[26] J. Yang, D. Zhang, and Z. Lu. Stochastic analysis of saturated-unsaturated flow in heterogeneous media by combining Karhunen-Loeve expansion and perturbation method. J. Hydrol., 294, 2004, 18–38.

[27] D. Zhang. Numerical solutions to statistical moment equations of groundwater flow in nonstationary, bounded heterogeneous media. Water Resour. Res., 34, 1998, 529–538.

[28] D. Zhang. Nonstationary stochastic analysis of transient unsaturated flow in randomly heterogeneous media. Water Resour. Res., 35, 1999, 1127–1141.

[29] D. Zhang. Stochastic Methods for Flow in Porous Media: Coping With Uncertainties. Academic Press, San Diego, CA, 2002.

[30] D. Zhang and Z. Lu. An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loeve and polynomial expansions. J. Comput. Phys., 194, 2004, 773–794.

[31] D. Zhang and S. P. Neuman. Eulerian-Lagrangian analysis of transport conditioned on hydraulic data, 1. Analytical-numerical approach. Water Resour. Res., 31, 1995, 39–51.