COMMUNICATIONS IN COMPUTATIONAL PHYSICS
Vol. **6**, No. 3, pp. 509-535

Commun. Comput. Phys.
September 2009

A Comparative Study of Stochastic Collocation Methods for Flow in Spatially Correlated Random Fields

Haibin Chang[*,1] and Dongxiao Zhang[1,2]

[1] *Department of Energy and Resources Engineering, College of Engineering, Peking University, Beijing 100871, China.*
[2] *Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, USA.*

Received 27 February 2008; Accepted (in revised version) 24 December 2008

Available online 6 February 2009

Abstract. Stochastic collocation methods as a promising approach for solving stochastic partial differential equations have been developed rapidly in recent years. Similar to Monte Carlo methods, the stochastic collocation methods are non-intrusive in that they can be implemented via repetitive execution of an existing deterministic solver without modifying it. The choice of collocation points leads to a variety of stochastic collocation methods including tensor product method, Smolyak method, Stroud 2 or 3 cubature method, and adaptive Stroud method. Another type of collocation method, the probabilistic collocation method (PCM), has also been proposed and applied to flow in porous media. In this paper, we discuss these methods in terms of their accuracy, efficiency, and applicable range for flow in spatially correlated random fields. These methods are compared in details under different conditions of spatial variability and correlation length. This study reveals that the Smolyak method and the PCM outperform other stochastic collocation methods in terms of accuracy and efficiency. The random dimensionality in approximating input random fields plays a crucial role in the performance of the stochastic collocation methods. Our numerical experiments indicate that the required random dimensionality increases slightly with the decrease of correlation scale and moderately from one to multiple physical dimensions.

AMS subject classifications: 60H15, 65M70, 76S05, 76M22

Key words: Stochastic collocation method, probabilistic collocation method, stochastic partial differential equations, Smolyak sparse grid method.

1 Introduction

Geological formation properties are ordinarily observed at a few locations despite they exhibit a high degree of heterogeneity. This leads to uncertainty in the description of the

*Corresponding author. *Email addresses:* haibinch@usc.edu (H. Chang), donzhang@usc.edu (D. Zhang)

 ©2009 Global-Science Press

formation properties and thus in the prediction of subsurface flow and transport. Such uncertainty necessitates a stochastic description of the formation parameters, which leads to stochastic partial differential equations governing flow and transport [25].

Monte Carlo (MC) simulation is one of the most natural approaches for solving the stochastic differential equations numerically. It is a statistical method that samples a large number of realizations for the random process and approximates the moments of interest with ensemble average. Thus the number of realizations, which one chooses, controls the accuracy of MC simulation [4]. To ensure the convergence of the moments, typically a few thousand samples or more are required, which is the main disadvantage of the direct sampling MC simulation.

An alternative approach is the stochastic finite element method, which has been rapidly developed in recent years [2, 6, 9, 10]. This method employs the polynomial chaos expansion (PCE) for random processes. After truncation in probability space, its formulation fits into the traditional spectral method framework [5, 9]. However, as the deterministic spectral methods, one must solve a set of coupled equations for the deterministic coefficients of the PCE. This increases the computational effort when the number of coefficients is large.

To overcome the difficulty for solving the coupled system, Mathelin et al. [13] proposed the so-called stochastic collocation method (SCM), which has had several successful applications [1, 15, 22]. In this approach, the output random field is approximated by Lagrange polynomial interpolation in probability space. One can derive an uncoupled system to solve the function values at selected positions. The solution process is highly parallelizable and it is found to be quite promising approach on the basis of examples with low random dimensions in the input random fields. The choice of collocation points leads to a variety of collocation methods including tensor product method [1], Smolyak method [15], Stroud 2 or 3 cubature method [17], and adaptive Stroud method [7].

Another kind of collocation method is the probabilistic collocation method (PCM) introduced by Tatang et al. [18] and successfully applied to the uncertainty analysis in some fields [11, 12, 20]. In this approach, the polynomial chaos expansion is used to approximate the output random field in probability space. The PCM is used to determine the coefficients of the polynomial chaos expansion by solving for the output random field for different sets of collocation points. The solution process is also highly parallelizable. For the cases examined with low random dimensions, this approach is found to be accurate and computationally efficient.

For numerical methods, accuracy and efficiency are two important aspects. For both the SCM and PCM, the computational efficiency depends on the total number of collocation points, which depends on both the representative random dimensions in the input random fields and the order of polynomial or other expansions in the dependent random fields. For a given dimensionality of input random field, the effect of the order or level of approximations in representing the dependent random fields has been studied [1,12,22]]. However, in practical applications the exact or proper dimension of the random space is often not known *a priori*. When the underlying (input) fields are spatially correlated, they

are described with an infinite number of random variables. These random fields may be approximated with a finite number of terms, hence a finite random dimensionality (N). For both the SCM and PCM, the solution accuracy may improve with the increase of N. However, the total number of collocation points and thus the computational efforts required increase rapidly with N. That is, the SCM and PCM may suffer from the so called "curse of dimensionality". In situations where the required random dimensionality N is too large, the SCM and PCM may become either less efficient than Monte Carlo simulations or even impractical computationally. The required dimensionality depends on a number of factors such as the correlation scale relative to the domain size, the variance, and the physical dimensionality. The effect of the retained random dimensionality on solution accuracy and efficiency has not been investigated for the SCM and PCM. However, that is crucial for the application of the SCM and PCM to flow and transport as well as other phenomena in correlated random fields. In addition, most of the studies of the stochastic collocation methods are in one physical dimension. Extending to multiple physical dimensions is conceptually straightforward. However, the random dimensionality could be much higher in multiple dimensions because the rate of decay of the eigenvalues of a correlated random field is much slower. Then the stochastic collocation methods may suffer from the curse of dimensionality and become impractical computationally if the critical random dimensionality is not selected wisely. In this work, we focus on discussing the computational properties and applicable range of these methods. In particular, we explore the possibility of defining a critical input random dimensionality for achieving the balance of accuracy and efficiency for the SCM and PCM. We do so through numerical experiments under different conditions and with numerical error analyses.

This paper is organized as follows: in Section 2 we introduce the governing equations for flow in spatially correlated random fields and the general formulations of various stochastic (probabilistic) collocation methods. In Section 3, we discuss the implementation of these methods. Numerical examples and analyses are presented in Sections 4 and 5. Some conclusions are given in Section 6.

2 Mathematical formulations

2.1 Governing equations

We consider the steady-state flow in saturated porous media satisfies the following equation:

$$\nabla\cdot[K_s(\boldsymbol{x})\nabla h(\boldsymbol{x})]+g(\boldsymbol{x})=0, \tag{2.1}$$

subject to boundary conditions

$$h(\boldsymbol{x})=H(\boldsymbol{x}), \qquad \boldsymbol{x}\in\Gamma_D, \tag{2.2}$$

$$K_s(\boldsymbol{x})\nabla h(\boldsymbol{x})\cdot\boldsymbol{n}(\boldsymbol{x})=-Q(\boldsymbol{x}), \qquad \boldsymbol{x}\in\Gamma_N, \tag{2.3}$$

where $h(\boldsymbol{x})$ is the hydraulic head, $K_s(\boldsymbol{x})$ is the hydraulic conductivity, and $g(\boldsymbol{x})$ is the source (or sink) team. $H(\boldsymbol{x})$ is the prescribed head on Dirichlet boundary segment Γ_D, $Q(\boldsymbol{x})$ is the prescribed flux across Neumann boundary segment Γ_N and $\boldsymbol{n}(\boldsymbol{x})=(n_1,\cdots,n_d)^T$ is the outward unit normal to the boundary $\Gamma_D\cup\Gamma_N$.

In this study, $K_s(\boldsymbol{x})$ is treated as a spatially correlated random field, thus equation (2.1) becomes a stochastic differential equation, whose solutions are no longer deterministic values but probability distributions or related moments. We aim to estimate the statistical properties of the hydraulic head in teams of the statistic moments of the log transformed $K_s(\boldsymbol{x})$.

2.2 The standard stochastic collocation method

2.2.1 Stochastic partial differential equations (SPDE)

Let $(\Omega,\mathcal{F},\mathcal{P})$ be a complete probability space, where Ω is the event space, $\mathcal{F}\subset 2^{\Omega}$ the σ-algebra, and $\mathcal{P}$ the probability measure. We consider a d-dimensional bounded domain $D\subset R^d$ $(d=1,2,3)$ and study the following problem: find a random function, $u\equiv u(\omega,\boldsymbol{x})$: $\Omega\times D\to R$, such that for $\mathcal{P}$-almost everywhere $\omega\in\Omega$, the following equation holds:

$$\mathcal{L}(u;\omega,\boldsymbol{x})=f(\omega,\boldsymbol{x}),\quad \boldsymbol{x}\in D, \tag{2.4}$$

where $\mathcal{L}$ is an operator involving differentiation in space and can be nonlinear and $f(\omega,\boldsymbol{x})$ is a known function. Here we omit the boundary conditions for simplicity. To solve equation (2.4) numerically, we need to reduce the infinite-dimensional probability space to a finite-dimensional space. This can be accomplished by characterizing the probability space by a finite number of random variables. So we suppose that $u(\omega,\boldsymbol{x})$ depends on random variables $\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_N)^T$ which take values in space P. Then equation (2.4) can be rewritten as

$$\mathcal{L}(u;\boldsymbol{\xi}(\omega),\boldsymbol{x})=f(\boldsymbol{\xi}(\omega),\boldsymbol{x}),\qquad \boldsymbol{\xi}\in \mathrm{P},\boldsymbol{x}\in \mathrm{D}. \tag{2.5}$$

In numerical solution, Eq. (2.5) is usually represented in a weak form: to seek $\hat{u}(\boldsymbol{\xi},\boldsymbol{x})\in V$ such that

$$\int_{\mathrm{P}}\rho(\boldsymbol{\theta})\mathcal{L}(\hat{u};\boldsymbol{\theta},\boldsymbol{x})v(\boldsymbol{\theta})d\boldsymbol{\theta}=\int_{\mathrm{P}}\rho(\boldsymbol{\theta})f(\boldsymbol{\theta},\boldsymbol{x})v(\boldsymbol{\theta})d\boldsymbol{\theta},\qquad \forall v(\boldsymbol{\theta})\in W, \tag{2.6}$$

where $\hat{u}(\boldsymbol{\xi},\boldsymbol{x})$ is a approximation of $u(\boldsymbol{\xi},\boldsymbol{x})$, V is called the trial function space, W is called the test function space, v is a test function, and $\rho(\boldsymbol{\theta})$ is the probability density function of $\boldsymbol{\xi}$. Different choices of V and W lead to different kinds of methods such as polynomial chaos expansion approach (PCE), probabilistic collocation method (PCM), and stochastic collocation method (SCM). In the following subsections, we give the general formulations for the PCM and the SCM, respectively.

2.2.2 The probabilistic collocation method (PCM)

The construction of the PCM is based on polynomial chaos expansion approximation in space P [12]. Now we define

$$V=span\{\Psi_i(\boldsymbol{\xi})\}_{i=1}^{M},\quad W=span\{\delta(\boldsymbol{\xi}-\boldsymbol{\theta}_i)\}_{i=1}^{M}, \tag{2.7}$$

where $\{\Psi_i(\boldsymbol{\xi})\}_{i=1}^{M}$ is a set of orthogonal polynomials in terms of the multi-dimensional random variable ξ with a specific probability distribution and $\boldsymbol{\theta}_i\in \mathrm{P}$ is a particular set selected with certain algorithm out of the random vector $\boldsymbol{\xi}$. Then the elements in $\boldsymbol{\theta}_i$ are called the set of collocation points. With the choice of V, we actually have the following approximation for $u(\boldsymbol{\xi},\boldsymbol{x})$:

$$\hat{u}(\boldsymbol{\xi},\boldsymbol{x})=\sum_{i=1}^{M}c_i(\boldsymbol{x})\Psi_i(\boldsymbol{\xi}). \tag{2.8}$$

Define the residual R as

$$R(\{c_i\},\boldsymbol{\xi})=\mathcal{L}\hat{u}-f. \tag{2.9}$$

Then Eq. (2.6) takes the following form:

$$\int_{\mathrm{P}}R(\{c_i\},\boldsymbol{\theta})\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_j)\rho(\boldsymbol{\theta})d\boldsymbol{\theta}=0,\quad j=1,\cdots,M, \tag{2.10}$$

which is equivalent to

$$R(\{c_i\},\boldsymbol{\theta}_j)=0,\quad j=1,\cdots,M, \tag{2.11}$$

resulting in a set of independent equations, evaluated at the given sets of collocation points, $\boldsymbol{\theta}_j$, where $j=1,\cdots,M$. In order to obtain the M coefficients $\{c_i\}$, where $i=1,\cdots,M$, we need to choose M sets of collocation points to solve Eq. (2.11) for M times. Once we obtain the coefficients of Eq. (2.8) from a system of M linear algebraic equations, the statistical moments of $u(\boldsymbol{\xi},\boldsymbol{x})$, such as mean and variance, can be derived as follows:

$$\langle u(\boldsymbol{\xi},\boldsymbol{x})\rangle=c_1(\boldsymbol{x}), \tag{2.12}$$

$$\sigma_u^2=\sum_{j=2}^{M}c_j(\boldsymbol{x})^2\left\langle\Psi_j^2\right\rangle. \tag{2.13}$$

The algorithm for choosing the sets of collocation points is significant to the performance of the PCM. The particular scheme described in [12] is to select the collocation points at a given order of polynomials from the roots of the next higher order of orthogonal polynomial for each uncertain parameter. Owing to the fact that the number of collocation points available is always larger than the number of collocation points needed, the algorithm is designed to select the points by keeping as many as possible of the variables at high probability values.

2.2.3 The stochastic collocation methods (SCM)

The construction of standard SCM is based on polynomial interpolation approximation in space P. Now we define

$$V = span\{L_i(\boldsymbol{\theta})\}_{i=1}^{M}, \quad W = span\{\delta(\boldsymbol{\theta}-\boldsymbol{\theta}_i)\}_{i=1}^{M}, \tag{2.14}$$

where $\{\boldsymbol{\theta}_i\}_{i=1}^{M} \subset \mathrm{P}$ be a set of prescribed interpolation nodes and $\{L_i(\boldsymbol{\theta})\}_{i=1}^{M}$ be the corresponding Lagrange interpolation basis function. Actually, we have the following approximation for $u(\boldsymbol{\xi},\boldsymbol{x})$,

$$\hat{u}(\boldsymbol{\xi},\boldsymbol{x}) = \mathcal{I}(u)(\boldsymbol{\theta},\boldsymbol{x}) = \sum_{i=1}^{M} u(\boldsymbol{\theta}_i,\boldsymbol{x}) L_i(\boldsymbol{\theta}), \tag{2.15}$$

where $\mathcal{I}$ is the Lagrange interpolation operator. Denote $u_i(\boldsymbol{x}) = u(\boldsymbol{\theta}_i,\boldsymbol{x})$, the function value at the interpolation node $\boldsymbol{\theta}_i$, which can be computed form the following equation,

$$\mathcal{L}(u_i(\boldsymbol{x});\boldsymbol{\theta}_i,\boldsymbol{x}) = f(\boldsymbol{\theta}_i,\boldsymbol{x}), \quad \boldsymbol{x} \in D, \tag{2.16}$$

which also results in a set of independent equations, evaluated at the given sets of interpolation nodes, $\boldsymbol{\theta}_i$, where $i = 1,\cdots,M$. Furthermore, we can approximate the mean and variance of function u as follows

$$\langle u(\boldsymbol{\xi})\rangle = \int_{\mathrm{P}} u(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \sum_{i=1}^{M} u(\boldsymbol{\theta}_i)\int_{\mathrm{P}} L_i(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta} = \sum_{i=1}^{M} u_i w_i, \tag{2.17}$$

$$\begin{aligned} Var(u(\boldsymbol{\xi})) &= \int_{\mathrm{P}} u^2(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta} - \left(\int_{\mathrm{P}} u(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}\right)^2 \\ &\approx \sum_{i=1}^{M} u^2(\boldsymbol{\theta}_i)\int_{\mathrm{P}} L_i(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta} - \left(\sum_{i=1}^{M} u(\boldsymbol{\theta}_i)\int_{\mathrm{P}} L_i(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}\right)^2 \\ &= \sum_{i=1}^{M} u_i^2 w_i - \left(\sum_{i=1}^{M} u_i w_i\right)^2, \end{aligned} \tag{2.18}$$

where $\{w_i\}_{i=1}^{M}$ are the corresponding quadrature weights. The construction of the SCM is composed of two sets of parameters: the interpolation nodes and the quadrature weights. The computational complexity of stochastic collocation methods is M times that of a deterministic problem, where M is the total number of collocation points. Thus we need to choose a nodal set with the fewest possible number of points under a prescribed accuracy. In the following subsections, we will present some existing choices.

Tensor product of one-dimensional nodal sets

A natural choice of the nodal set is the tensor product of one-dimensional sets. When $N = 1$, there are many good interpolation formulas for smooth functions $u : (-\infty,\infty) \to \mathrm{R}$

i.e. for each direction $i=1,\cdots,N$, we can construct a one-dimensional interpolation,

$$\mathcal{U}^i(u)=\sum_{k=1}^{m_i} u(\theta_k^i)\cdot a_k^i, \tag{2.19}$$

based on nodal sets

$$\Theta^i=(\theta_1^i,\cdots,\theta_{m_i}^i)\subset(-\infty,\infty), \tag{2.20}$$

where $a_k^i=a_k(\boldsymbol{\theta}_i)$ is the one-dimensional (1D) interpolation basis polynomial at θ_k^i. For the multivariate case, the tensor product formulas are

$$\mathcal{I}(u)=\left(\mathcal{U}^{i_1}\otimes\cdots\otimes\mathcal{U}^{i_N}\right)(u)=\sum_{k_1=1}^{m_{i_1}}\cdots\sum_{k_N=1}^{m_{i_N}} u\left(\theta_{k_1}^{i_1},\cdots,\theta_{k_N}^{i_N}\right)\cdot\left(a_{k_1}^{i_1}\otimes\cdots\otimes a_{k_N}^{i_N}\right). \tag{2.21}$$

Clearly, the above product formula needs $M=m_{i_1}\cdots m_{i_N}$ nodal points. However, if we use the same interpolating function (2.19) for each dimension with the same number of points, i.e., $m_{i_1}=\cdots=m_{i_N}\equiv m$, the total number of points is $M=m^N$. This number grows quickly in high dimensions $N\gg 1$. This property makes the tensor product algorithm impractical for high dimensions.

The Smolyak method

The Smolyak formulas are the linear combinations of the product formulas (2.21) with the following key properties: only products of a relatively small number of nodes are used and the linear combination is chosen in such a way that an interpolation property for $N=1$ is preserved for $N>1$ [16].

The Smolyak algorithm is given by [19]

$$\mathcal{I}(u)=A(q,N)(u)=\sum_{q-N+1\leqslant|i|\leqslant q}(-1)^{q-|i|}\begin{pmatrix} N-1 \\ q-|i| \end{pmatrix}\left(\mathcal{U}^{i_1}\otimes\cdots\otimes\mathcal{U}^{i_N}\right)(u), \tag{2.22}$$

where $i=(i_1,\cdots,i_N)\in\mathbb{N}_+^N$ and $|i|=i_1+\cdots+i_N$. To compute $A(q,N)(u)$, one only needs to know the function values on the "sparse grid",

$$\Theta_N\equiv\mathcal{H}(q,N)=\bigcup_{q-N+1\leqslant|i|\leqslant q}(\Theta^{i_1}\times\cdots\times\Theta^{i_N}). \tag{2.23}$$

In this paper we choose to use the Smolyak formulas based on polynomial interpolation at the zeros of the orthogonal polynomials with respect to a weight ρ. This naturally leads to the Gauss formulas that have a maximum degree of exactness of $2m_i-1$. (Here we do not use the Clenshaw-Curtis formulas because the probability space with which we deal is infinite.) Furthermore, we choose $m_1=1$ and $m_i=2^{i-1}+1$ for $i>1$.

From Fig. 1, we can see that the sparse grid method uses fewer nodes than the tensor product algorithm from the same one-dimensional nodes. This will be much obvious when the dimension is high. For example, when $N=6$, the total number of nodes of

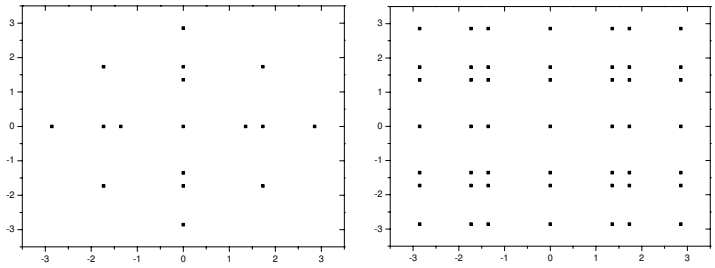

Figure 1: Two-dimensional ($N=2$) interpolation nodes. Left: sparse grid $\mathcal{H}(4,2)$ from Smolyak algorithm, the total number of points being 17. Right: tensor product algorithm from the same one-dimensional nodes, the total number of nodes being 49.

sparse grid $\mathcal{H}(8,6)$ is 97 while the corresponding number of tensor product algorithm from the same one-dimensional nodes is $7^6=117649$.

In the computation, we set $q=N+k$ and refer to k in the $A(N+k,N)$ as the "level " of Smolyak construction [22].

Stroud-2 cubature method and adaptive Stroud collocation method

For $\mathrm{P}=[-1,1]^N$, Stroud [17] constructed a set of cubature points with $(N+1)$-point that is accurate for multiple integrals of polynomials of degree 2. The degree 2 formula, termed as the Stroud-2 method hereafter, consists of points $\{\boldsymbol{\theta}_i\}_{i=0}^N$ such that

$$\theta_i^{2r-1}=\sqrt{\frac{2}{3}}\cos\frac{2ri\pi}{N+1},\quad \theta_i^{2r}=\sqrt{\frac{2}{3}}\sin\frac{2ri\pi}{N+1},\quad r=1,2,\cdots,[N/2], \tag{2.24}$$

where $[N/2]$ is the greatest integer not exceeding $N/2$, and if N is odd $\theta_i^N=(-1)^i/\sqrt{3}$. The quadrature weight is $w_i=1/(N+1)$, $i=0,\cdots,N$.

The Stroud-2 method employs the minimal number of points for its corresponding algebraic accuracy [14]. But when N is prescribed, the accuracy cannot be improved any further. An adaptive Stroud collocation method was recently put fourth in order to improve accuracy [7]. That is to subdivide the space $[-1,1]^N$ into small parts and take integral in each.

Mathematically, let $\{\mathrm{P}_j\}_{j=1}^K$ express the partitions of P. By affine transformation and using the collocation points of the Stroud-2 method, one can get the points in P_j, $\{\theta_i^j\}_{i=0}^N$. So the approximations about the mean and variance of $u(\boldsymbol{\xi})$ are

$$\langle u(\boldsymbol{\xi})\rangle=\int_{[-1,1]^N}u(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}\approx\frac{2^N}{K(N+1)}\sum_{i=0}^N\sum_{j=1}^K u(\theta_i^j)\rho(\theta_i^j), \tag{2.25}$$

$$\begin{aligned}Var(u(\boldsymbol{\xi}))&=\int_{[-1,1]^N}u^2(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}-\left(\int_{[-1,1]^N}u(\boldsymbol{\theta})\rho(\boldsymbol{\theta})d\boldsymbol{\theta}\right)^2\\&\approx\frac{2^N}{K(N+1)}\sum_{i=0}^N\sum_{j=1}^K u^2(\theta_i^j)\rho(\theta_i^j)-\left(\frac{2^N}{K(N+1)}\sum_{i=0}^N\sum_{j=1}^K u(\theta_i^j)\rho(\theta_i^j)\right)^2.\end{aligned} \tag{2.26}$$

3 KL based stochastic collocation method

3.1 Karhulen-Loeve expansion

Let $Y(\boldsymbol{x},\omega)$ be a random process, where $\boldsymbol{x}\in D$ and $\omega\in\Omega$. One may write $Y(\boldsymbol{x},\omega)=\overline{Y}(\boldsymbol{x})+Y'(\boldsymbol{x},\omega)$, where $\overline{Y}(\boldsymbol{x})$ is the mean and $Y'(\boldsymbol{x},\omega)$ is the fluctuation. The spatial structure of the random field may be described by the covariance $C_Y(\boldsymbol{x},\boldsymbol{y})=\langle Y'(\boldsymbol{x},\omega)Y'(\boldsymbol{y},\omega)\rangle$. Since the covariance is bounded, symmetric and positive-defined, it may be decomposed as [9]

$$C_Y(\boldsymbol{x},\boldsymbol{y})=\sum_{n=1}^{\infty}\lambda_n f_n(\boldsymbol{x})f_n(\boldsymbol{y}), \tag{3.1}$$

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

$$\int_D C_Y(\boldsymbol{x},\boldsymbol{y})f(\boldsymbol{x})d\boldsymbol{x}=\lambda f(\boldsymbol{y}). \tag{3.2}$$

Then the random process $Y(\boldsymbol{x},\omega)$ can be expressed as

$$Y(\boldsymbol{x},\omega)=\overline{Y}(\boldsymbol{x})+\sum_{n=1}^{\infty}\sqrt{\lambda_n}f_n(\boldsymbol{x})\xi_n(\omega), \tag{3.3}$$

where $\xi_n(\omega)$ are orthogonal Gaussian random variables with zero mean and unit variance when Y is assumed to be Gaussian. The expansion in Eq. (3.3) is called the Karhulen-Loeve (KL) expansion. The KL expansion, which is a spectral expansion, is optimal with mean square convergence when the underlying process is Gaussian [9]. As such, one may truncate the infinite series of (3.3) with a finite number (N) of terms. The rate of decay of λ_n determines the number of terms that are retained in the Karhulen-Loeve expansion, which determines the retained random dimensionality (N) of the problem. Therefore, when the input random fields are spatially correlated, the random dimensionality is not known *a priori*. In this study, we will explore the effect of correlation length, variance, and physical dimensionality on the retained random dimensionality and the impact of the latter on the solution accuracy and efficiency.

Although, in general, the eigenvalue problem (3.2) has to be solved numerically, there exist analytical or semi-analytical solutions under certain conditions. For a one-dimensional stochastic process 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 process, respectively, the eigenvalues and their eigenfunctions can be expressed as [26]

$$\lambda_n=\frac{2\eta\sigma_Y^2}{\eta^2w_n^2+1}, \tag{3.4}$$

and

$$f_n(x)=\frac{1}{\sqrt{(\eta^2w_n^2+1)L/2+\eta}}[\eta w_n\cos(w_nx)+\sin(w_nx)], \tag{3.5}$$

where w_n are positive roots of the characteristic equation

$$(\eta^2 w^2 - 1)\sin(wL) = 2\eta w \cos(wL). \tag{3.6}$$

Eq. (3.6) has an infinite number of positive roots. If the roots w_n are sorted in an increasing order, the related eigenvalues λ_n are monotonically decreasing.

3.2 The implementation of PCM

Here we follow closely the work of [12] for describing the implementation process of the PCM. Suppose the input random field, $K_s(\boldsymbol{x})$, is a lognormal random field. Let $Y(x) = \ln[K_s(\boldsymbol{x})]$ and substitute the KL expansion of $Y(\boldsymbol{x})$ into the governing equation (2.1) yields

$$\nabla\left\{\exp\left[\bar{Y}(\boldsymbol{x}) + \sum_{n=1}^{N}\sqrt{\lambda_n} f_n(\boldsymbol{x})\xi_n(\omega)\right]\nabla h(\boldsymbol{x})\right\} + g(\boldsymbol{x}) = 0. \tag{3.7}$$

Here we truncate Eq. (3.3) by keeping a finite number (N) of terms and N is the dimensionality of the input random field. We consider the approximation of the output random field, $h(\boldsymbol{x},\boldsymbol{\xi})$, with a second order polynomial chaos expansion, and express it as $\hat{h}(\boldsymbol{x},\boldsymbol{\xi})$ [8,9,21]

$$\hat{h}(\boldsymbol{x},\boldsymbol{\xi}) = a_0(\boldsymbol{x}) + \sum_{i=1}^{N} a_i(\boldsymbol{x})\xi_i + \sum_{i=1}^{N} a_{ii}(\boldsymbol{x})(\xi_i^2 - 1) + \sum_{i=1}^{N-1}\sum_{j>i}^{N} a_{ij}(\boldsymbol{x})(\xi_i\xi_j). \tag{3.8}$$

Here we use the multi-dimensional Hermite Polynomials owing to the fact that $\boldsymbol{\xi}$ is a Gaussian random vector [9]. In case of other random distributions, generalized polynomial chaos expansions [23,24] can be used to represent the random field. Eq. (3.8) can be rewritten as

$$\hat{h}(\boldsymbol{x},\boldsymbol{\xi}) = \sum_{i=1}^{M} c_i(\boldsymbol{x})\Psi_i(\boldsymbol{\xi}). \tag{3.9}$$

There is a one-to-one correspondence between the terms in Eqs. (3.8) and (3.9). The total number of terms (M) is determined by the random dimensionality (N) and the degree (d) of the polynomial chaos expansion, $M = \frac{(N+d)!}{N!d!}$.

For the probabilistic collocation method, the residual is zero at the selected set of collocation points, specified as $\boldsymbol{\theta}_j = (\theta_{1,j},\theta_{2,j},\cdots,\theta_{N,j})^T$, where $j = 1,\cdots,M$. We have the following equation

$$\nabla\cdot\left\{\exp\left[\bar{Y}(\boldsymbol{x}) + \sum_{n=1}^{N}\sqrt{\lambda_n} f_n(\boldsymbol{x})\theta_{n,j}\right]\nabla\hat{h}_j(\boldsymbol{x})\right\} + g(\boldsymbol{x}) = 0, \tag{3.10}$$

where $\hat{h}_j(\boldsymbol{x})$, $j = 1,\cdots,M$, is the function value for the j^{th} set of collocation points.

We need to choose M sets of collocation points to solve Eq. (3.10) for M times. Then we can obtain M sets of the hydraulic head field, $\hat{h}_j(\boldsymbol{x})$. Here we denote the coefficients

in Eq. (3.9) by $c_1(\boldsymbol{x}), c_2(\boldsymbol{x}),\cdots, c_M(\boldsymbol{x})$ in sequence, and $C(\boldsymbol{x})=[c_1, c_2,\cdots, c_M]^T$. The corresponding hydraulic head field at each collocation point are $\hat{h}_1(\boldsymbol{x}), \hat{h}_2(\boldsymbol{x}),\cdots, \hat{h}_M(\boldsymbol{x})$, and $h(\boldsymbol{x})=[\hat{h}_1, \hat{h}_2,\cdots, \hat{h}_M]^T$. Then we have

$$ZC(\boldsymbol{x})=h(\boldsymbol{x}), \tag{3.11}$$

where Z is a space-independent matrix of dimension $M\times M$, consisting of Hermite polynomials at the collocation points. By solving the linear system of equation, (3.11), the deterministic coefficients $C(\boldsymbol{x})$ could be obtained readily. Note that the selected collocation points should make the matrix Z satisfy the limitation of $rank(Z)=M$.

Finally we can use Eqs. (2.12) and (2.13) to evaluate the statistical properties of the hydraulic head field.

3.3 The implementation of SCM

The major steps in the implementation of SCM are: (1) Representing the input random field with the KL expansion in terms of a set of independent Gaussian random variables; (2) Approximating the output random field with the polynomial interpolation in the probability space of the same set of Gaussian random variables; (3) Determining the values of the output random field at the interpolation nodes; (4) Evaluating the statistical properties of the output random field.

The implementation of the SCM is similar to that of the PCM, so we do not need to discuss them in detail. But we must pay attention to the Stroud method whose interpolation nodes are chosen from the probability space $[-1,1]^N$ of uniform random vector with N elements, not the probability space $(-\infty,+\infty)^N$ of Gaussian random vector. Owing to the assumption that both the random fields $\ln[K_s(\boldsymbol{x},\omega)]$ and $h(\boldsymbol{x},\omega)$ depend on Gaussian random vector $\boldsymbol{\xi}$, we need to establish the relationship between uniform and Gaussian random variables before using this method [7].

Denote the cumulative distribution function (CDF) of a Gaussian random variable ξ_i by $\mathcal{F}_(\xi_i)$ and suppose

$$F(\xi_i)=2\mathcal{F}(\xi_i)-1. \tag{3.12}$$

It is straightforward to know that $\eta_i=F(\xi_i)$ is uniformly distributed in $[-1,1]$. Denote the j^{th} interpolation node of Stroud method as $\boldsymbol{\eta}_j=(\eta_{1,j},\eta_{2,j},\cdots,\eta_{N,j})$. Then the function value at the j^{th} interpolation node, $\hat{h}_j(\boldsymbol{x})$, can be solved from the following equation

$$\nabla\cdot\left\{\exp\left[\overline{Y}(\boldsymbol{x})+\sum_{n=1}^{N}\sqrt{\lambda_n}f_n(\boldsymbol{x})F^{-1}(\eta_{n,j})\right]\nabla\hat{h}_j(\boldsymbol{x})\right\}+g(\boldsymbol{x})=0. \tag{3.13}$$

Eqs. (2.17) and (2.18) can be used to evaluate the statistical properties of the hydraulic head field for all types of SCM.

3.4 The computational effort for PCM and SCM

In the following, we denote the interpolation nodes of the SCM as collocation points also for simplicity and express the total number of the collocation points as M. The implementation of the PCM and SCM is nearly the same. But there is one more step in the PCM, which is to solve the coefficients of Eq. (3.9) after obtaining the values of the hydraulic head for each collocation point set. This step can be done according to Eq. (3.11). Note that Z is a space-independent matrix of dimension $M\times M$, that is to say, Z is a fixed matrix when we have chosen the collocation points already. We only need to deal with matrix Z once so this step will not increase much computational effort for the PCM.

For both the PCM and SCM, computing the values of the hydraulic head at all physical nodes for each collocation point is the most time consuming step, especially in the presence of large number of physical nodes. So the computational effort for both the PCM and SCM depends on the total number of collocation points, M, which strongly depends on the order or level of approximations in the dependent variables and the random dimensionality (N) retained for the input random fields.

4 Numerical results

In the following subsections, we use both the SCM and the PCM to solve the problem of flow in porous media. We aim to compare these methods about their accuracy and efficiency. We consider the steady state flow in one-dimensional domain of length $L=10$[L] (where [L] denotes any consistent length unit) and assume the source (sink) term to be zero. The boundary conditions are prescribed heads on the two ends, $H_0=7$[L] and $H_L=5$[L]. The correlation length is given as $\eta=4$[L]. The mean of the log hydraulic conductivity is given as $<Y>=0$. We choose three different values for the spatial variability of Y (i.e., the variance $\sigma_Y^2=0.3$, 1.0 or 2.0) in the computations.

We need to decide the random dimensionality according to the correlation length. The eigenvalue and eigenfunction, λ_n and $f_n(x)$, n=1,2,$\cdots$, can be determined analytically by solving Eqs. (3.4) and (3.5). The eigenvalues are monotonically decreasing as illustrated in Fig. 2(a) with different correlation lengths ($\eta=1.0$ and 4.0). Fig. 2(b) shows the sum of the eigenvalues as a function of number of terms included. From Eq. (3.1), we have $\sigma_Y^2=\sum_{n=1}^{\infty}\lambda_n f_n^2(x)$. Integrating this equation yields $D\sigma_Y^2=\sum_{n=1}^{\infty}\lambda_n$, where D is the measure of the domain size (length, area, or volume for 1D, 2D, or 3D domain, respectively). On the other hand, λ_n expresses the energy and input information for each term. It is clear that when we deal with the small correlation length condition, more terms must be included in the KL expansion to retain enough energy. The salient question is how many terms are needed to retain in order to keep the balance between solution accuracy and efficiency.

In the following computations, the finite difference method is used to solve Eq. (3.10) or (3.13) and 151 physical nodes are chosen for all SCMs and the PCM. We solve the same problem with the direct sampling Monte Carlo method as a reference solution. We

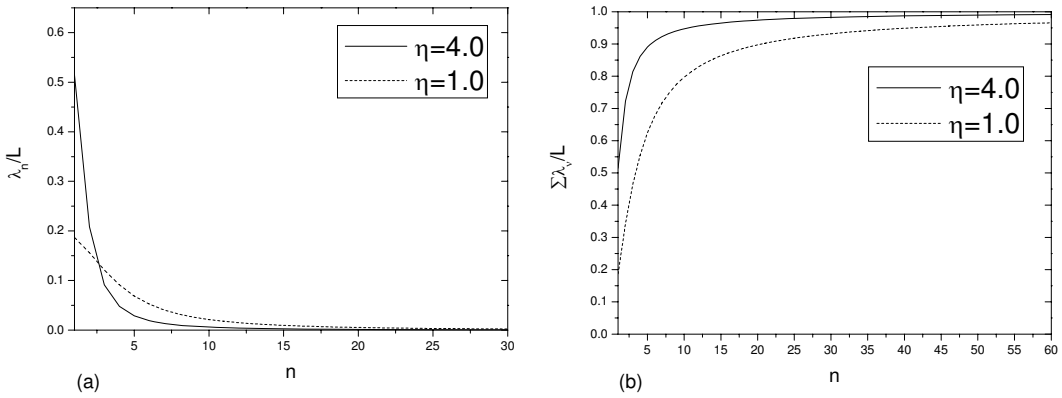

Figure 2: Series of eigenvalues (a) and finite sum (b), for $\eta=4.0$ and $\eta=1.0$.

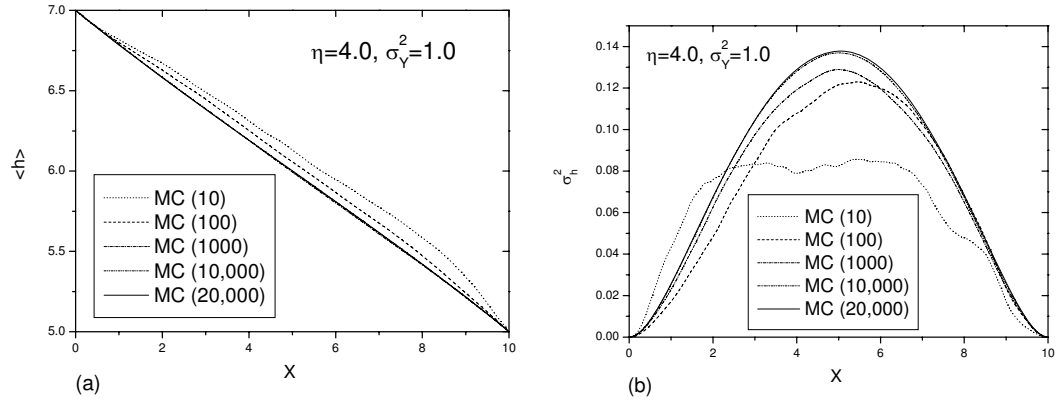

Figure 3: The mean and variance of hydraulic head derived from 5 sets of Monte Carlo simulations, corresponding to 10, 100, 1000, 10,000 and 20,000 realizations, for $\eta=4.0$ and $\sigma_Y^2=1.0$.

truncate the KL expansion to 100 terms to generate the random field of the log hydraulic conductivity based on Eq. (3.3) and solve multiple realizations.

Fig. 3 illustrates 5 sets of Monte Carlo simulations, corresponding to 10, 100, 1000, 10,000 and 20,000 realizations. It can be found that Monte Carlo simulations with 10 or 100 realizations do not lead to statistically accurate results and that even the variance from 1000 realizations deviates from that obtained with 10,000 realizations. The results from 10,000 and from 20,000 realizations coincide with each other and are believed to have converged statistically. Therefore, in this study the Monte Carlo simulations with 10,000 realizations are regarded as the reference solutions in all cases.

Here we choose a moderate correlation length, $\eta=4.0$, for the random input field. Owing to the rapid decay of the eigenvalues, only the first 6 terms are retained in the KL expansion. That is, the random dimensionality is $N=6$. Owing to the particular boundary conditions in our examples, the mean head obtained from different approaches are very close to each other. We thus focus our discussion only on the head variance in the

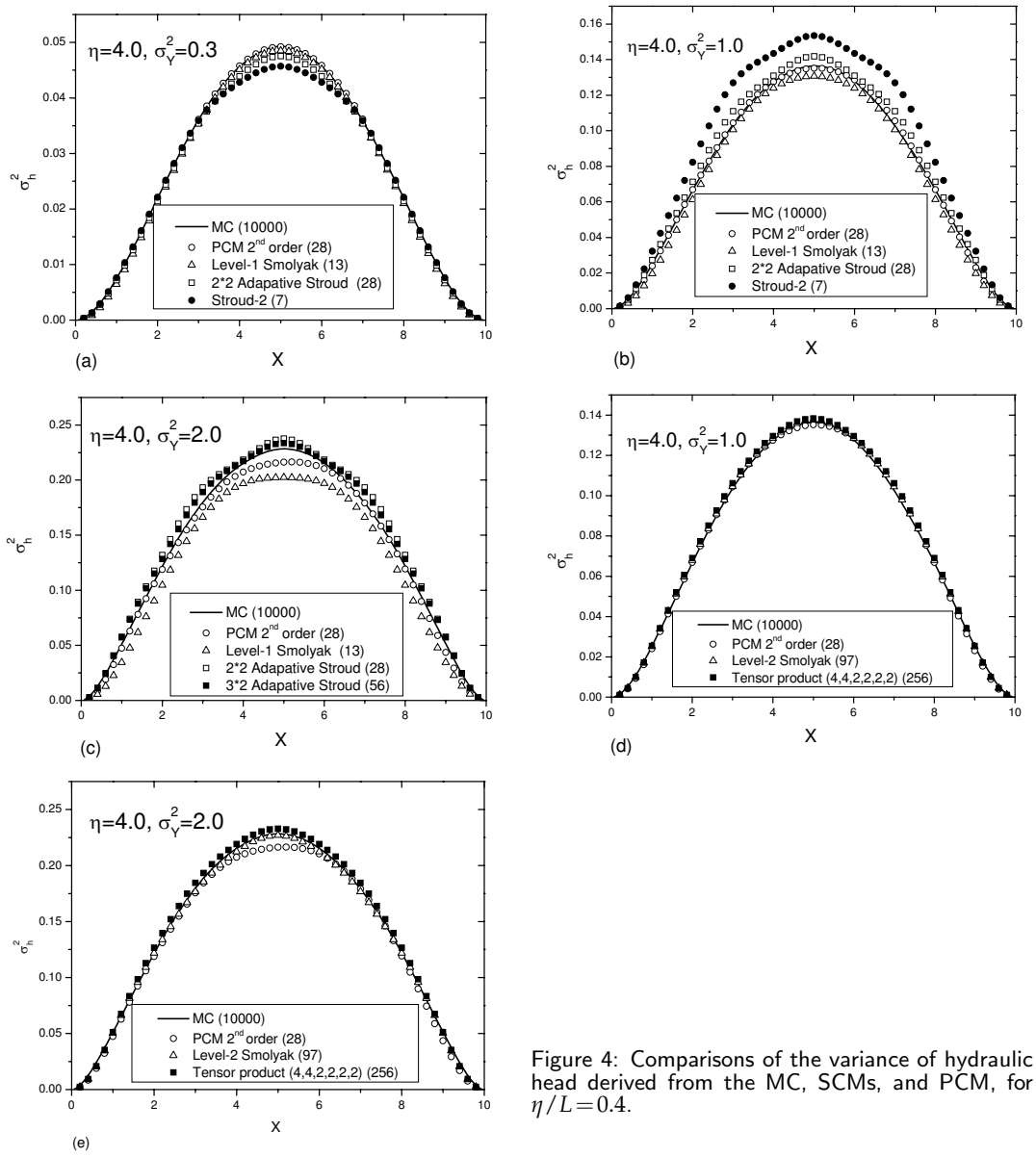

Figure 4: Comparisons of the variance of hydraulic head derived from the MC, SCMs, and PCM, for $\eta/L=0.4$.

following sections. For numerical modeling and simulation, accuracy and efficiency are two important aspects. As such, we provide two groups of comparisons for the stochastic collocation methods. In the first group, the methods all use small number of collocation points, which also means a small computational effort, so we compare their accuracy based on Figs. 4(a), (b) and (c). In the second group, all the methods compared can obtain accurate results, so we compare their efficiency based on Figs. 4(d) and (e). The observations are given in the following paragraphs.

It is clear that Stroud method uses the fewest collocation points but it is not accurate even under the condition of small variability, $\sigma_Y^2 = 0.3$. As such, we will not discuss the Stroud method in the following comparisons. An improvement is found in the 2×2 adaptive Stroud method [7] relative to the Stroud method, where $'2$ means the first two dimensions of the random variable ξ are subdivided, and $2'$ means each dimension is divided into 2 parts. But the variance from the 2×2 adaptive Stroud method still deviates from the Monte Carlo result in the middle part of the domain for the case of small variability, $\sigma_Y^2 = 0.3$. As shown in Fig. 4(b), the variances from the adaptive Stroud and level-1 Smolyak are acceptable but not as accurate as the second order PCM whose result agrees with the Monte Carlo result fairly well. The superiority of the second order PCM can also be seen from Fig. 4(c) where we increase the variability to 2, $\sigma_Y^2 = 2.0$. We perform the 3×2 adaptive Stroud method but we do not find much improvement (Fig. 4(c)).

Figs. 4(d) and (e) show the variances of hydraulic head obtained from the Monte Carlo, the second order PCM, the level-2 Smolyak, and the tensor product method. Here we use the anisotropic full tensor product algorithm [3], which is implemented with the following strategy: for a given accuracy requirement, it increases the number of interpolation nodes in one dimension as much as possible before considering the next direction. The notation [4,4,2,2,2,2] means the number of interpolation nodes we used for each dimension. As shown in Figs. 4(d) and (e), the results derived from all the methods considered agree with the Monte Carlo result fairly well. So we only need to compare their efficiency. As mentioned before, the computational complexity of the SCMs and the PCM depends on the total number of collocation points. The total number of collocation points of the second order PCM, the lever-2 Smolyak and the tensor product are 28, 97 and 256, respectively. So for this case, the second order PCM is the most efficient method.

5 Discussions

5.1 Effect of small correlation length η

As shown in Fig. 2, the correlation length L relative to the domain length L controls the rate of decay in the eigenvalues. To further test the effect of correlation length on the SCMs and the PCM, three cases for $\eta/L = 0.1$ with different spatial variability $\sigma_Y^2 = 0.3$, 1.0 and 2.0, are performed. We also use the Monte Carlo simulation result for comparison. For these cases, we choose the random dimensionality as $N = 9$. The strategy for determination of random dimensionality will be discussed in Section 5.3. In these cases, the tensor product method will not be considered because there will be $2^9 = 512$ collocation points even if only 2 interpolation nodes are used in each dimension.

As shown in Fig. 5(a), in the case of small spatial variability, the level-1 Smolyak method is attractive because it uses the fewest collocation points and its result is good. From Fig. 5(b), we can see that the results derived from the 6×2 adaptive Stroud method and level-1 Smolyak method deviate from the Monte Carlo result obviously. Notice that the 6×2 adaptive Stroud method uses more than 11 times of collocation points compared

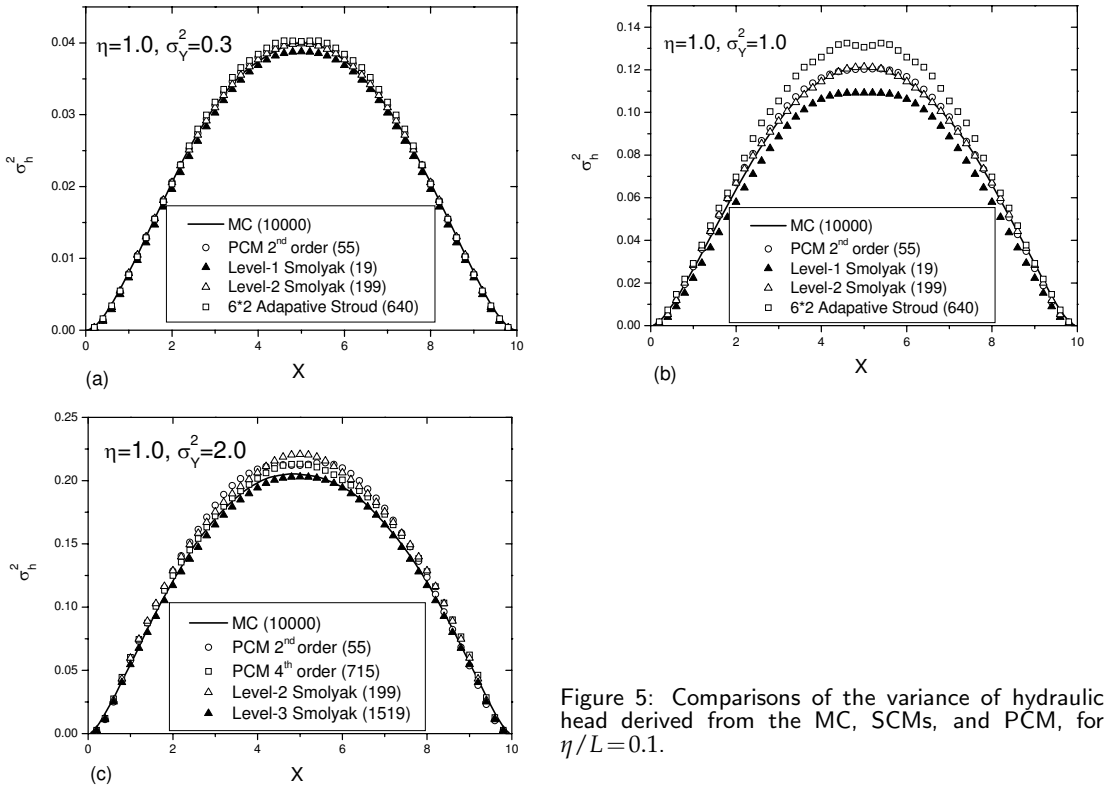

Figure 5: Comparisons of the variance of hydraulic head derived from the MC, SCMs, and PCM, for $\eta/L=0.1$.

to the second order PCM, so the adaptive Stroud method is not attractive for the high dimensional problem. The second order PCM is superior to the level-2 Smolyak method for using fewer collocation points to get nearly the same accuracy in the case of moderate spatial variability. As shown in Fig. 5(c), when σ_Y^2 is large, i.e., $\sigma_Y^2=2.0$, the results derived from the second order PCM, the fourth order PCM and the level-2 Smolyak method all somewhat deviate from the Monte Carlo result and only the level-3 Smolyak method provides a close match.

5.2 Effect of large spatial variability σ_Y^2

As illustrated in the previous sections, for moderate spatial variability, the second order PCM can obtain quite accurate results with fewer collocation points compared to other methods. However, for large spatial variability, there are some deviations in the moderate correlation length case and the deviation is even larger in the case of small correlation length. In this section, we examine some cases with an even larger spatial variability, i.e., $\sigma_Y^2=4.0$.

As shown in Fig. 6(a), for the case of large spatial variability and moderate correlation length, the results derived from the second order PCM and tensor product deviate from

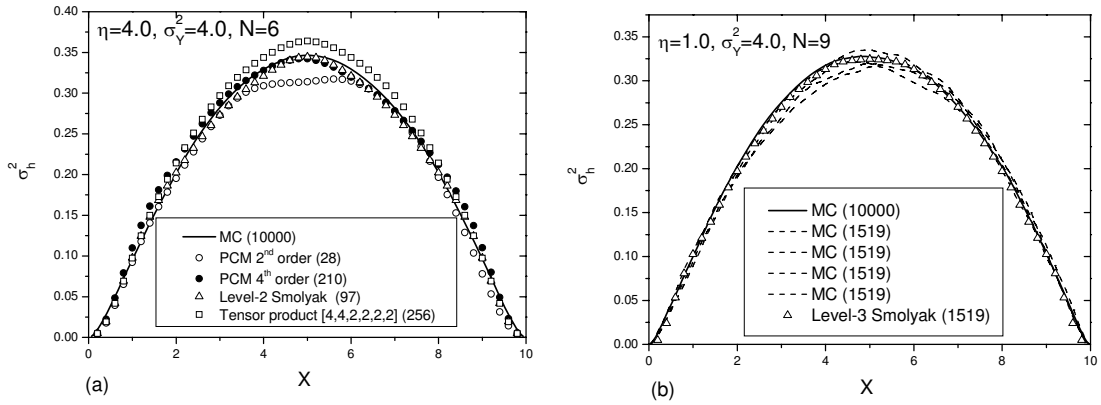

Figure 6: Comparisons of the variance of hydraulic head derived from the MC, SCMs, and PCM, for $\sigma_Y^2 = 4.0$.

the Monte Carlo result while the fourth order PCM and the level-2 Smolyak method agree with the Monte Carlo result well. So, in this case, the level-2 Smolyak method is superior to the fourth order PCM for its less computational effort. When dealing with large spatial variability and small correlation length, the level-3 Smolyak method may be appropriate. There are five sets of results from the Monte Carlo simulation: the dash lines come from the Monte Carlo simulation with 1519 realizations and the solid line comes from the Monte Carlo simulation with 10,000 realizations, which we use as the benchmark. As shown in Fig. 6(b), with 1519 realizations, the Monte Carlo simulation does not converge. The result from the level-3 Smolyak method agree with the benchmark fairly well, so it is superior to the direct Monte Carlo simulation with the same computational effort.

5.3 The error comparison between PCM and Smolyak method

As discussed above, the PCM and Smolyak method are two accurate and efficient methods compared to other methods. When the random input fields can be represented isotropically with a fixed (and known) random dimensionality, there is a rigorous error analysis for the Smolyak method [15]. Such an error analysis is not applicable to spatially correlated random fields because after representing the latter with the KL expansion, the random space becomes anisotropic owing to different eigenvalues and eigenfunctions and a proper random dimensionality is neither fixed nor known *a priori*. For flow in correlated random fields, both the choice of the random dimensionality and that of the order of polynomials contribute to the errors. In this study, we introduce the following error definition:

$$\varepsilon = \frac{1}{\hat{M}} \sum_{j=1}^{\hat{M}} \left| y(x_j) - y_{N,d}(x_j) \right|, \tag{5.1}$$

where $\hat{M}$ is the total number of physical nodes, $y(x_j)$ are the reference values, $y_{N,d}(x_j)$ are the approximate function values, and subscripts N and d denote the random dimension-

ality and the order (level) of a given method, respectively. Owing to the fact that there is no analytical solution for our problem, we use the high-resolution Monte Carlo results as the reference solution. Note that, here we only analyze the error in the random space by using the same numerical solver on the same grid.

It is seen from Eq. (5.1) that, in order to perform the error analysis, for a given order of method, we must first determine the proper random dimensionality, N, for each case. As we know, for both the SCM and PCM, the solution accuracy and efficiency depends on the total number of collocation points, which is a strong function of the representative random dimensions in the input random fields. As such, we introduce the following error definition,

$$\varepsilon_i = \frac{1}{\hat{M}} \sum_{j=1}^{\hat{M}} \left| y_{i,d}(x_j) - y_{i-n,d}(x_j) \right| \cdot \frac{1}{\sigma_Y}, \tag{5.2}$$

where $y_{i,d}(x_j)$ is the value of statistical moment (being the variance in our cases) of the random variable at x_j when the random dimensionality is i and the order of the method is d, and $n \in \mathbb{N}_+$ is the step size of the increment of the random dimensionality.

In practice, we need to choose an error criterion, ε_c, for the determination of N. In the following, we determine a critical random dimensionality so that the results from each method are converged to a given criterion. We first investigate the influences of spatial variability and correlation length. As show in Figs. 7(a) and (b), for a fixed correlation length, with the increase of the spatial variability the critical random dimensionality should also be increased for a given ε_c. If we choose a very small criterion error, the difference between the random dimensionality for different spatial variability is neglectable. As show in Figs. 7(c) and (d), for a fixed spatial variability, with the decrease of correlation length the critical random dimensionality should be increased for a given ε_c. We can see that the correlation length is a more important parameter than the spatial variability for determining the critical N. It should be noted that the PCM and Smolyak method possess nearly the same rate of convergence with respect to N.

In order to better understand why we need the convergence criterion mentioned above to find the critical random dimensionality at various correlation lengths, we introduce another criterion for choosing the random dimensionality that is to retain the same energy level in the KL expansion. That is, a random dimensionality, N_c, is selected to make sure that a certain level of energy is preserved in the input field. We denote this energy as E_c, which can be computed as

$$E_c = \frac{\sum_{n=1}^{N_c} \lambda_n}{\sum_{n=1}^{\infty} \lambda_n} = \frac{\sum_{n=1}^{N_c} \lambda_n}{D\sigma_Y^2}.$$

It is seen from Fig. 8(a) that with the 90% energy criterion (curve with squares), the retained random dimensionality increases exponentially with the decrease of the correlation length. This energy criterion for choosing random dimensionality would lead to the problem of the curse of dimensionality for the collocation methods (and other stochastic expansion methods). On the other hand, we may adopt the convergence criterion

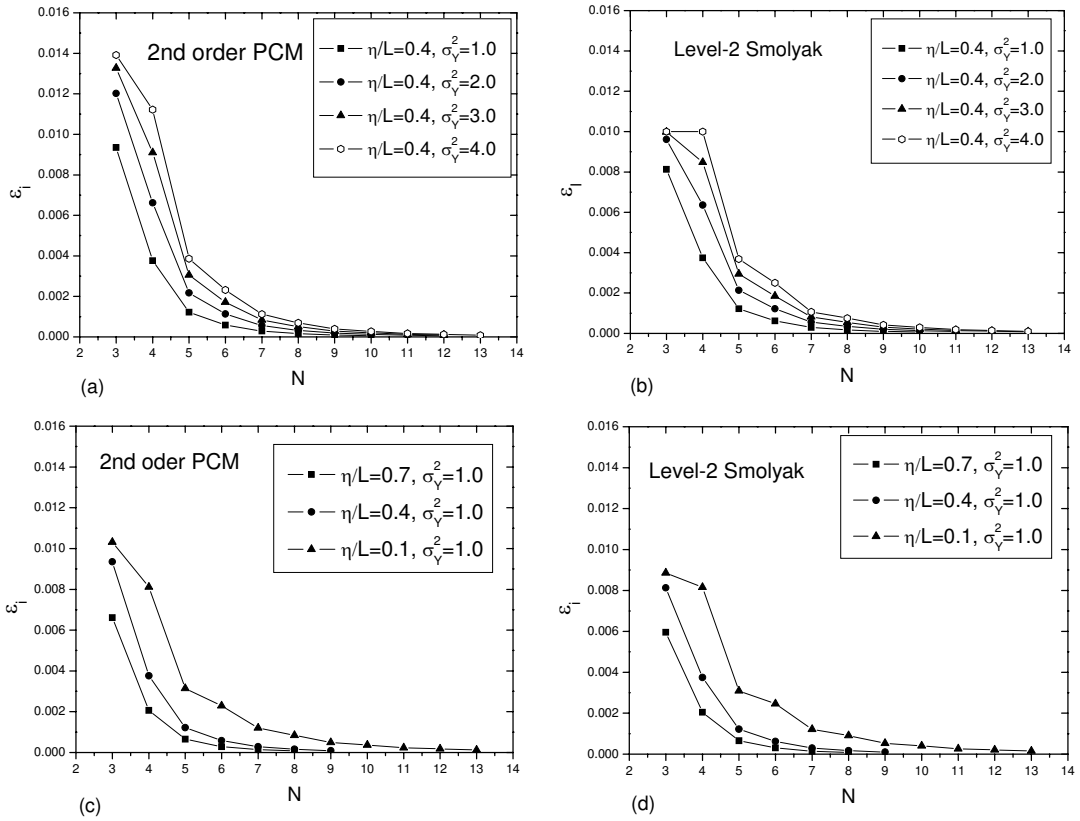

Figure 7: Convergence analyses for both PCM and Smolyak method with respect to N. The convergence rate for the PCM (a) and Smolyak method (b) with fixed correlation length and different spatial variability levels. The convergence rate for the PCM (c) and Smolyak method (d) with fixed spatial variability and different correlation lengths.

mentioned above to investigate how much energy is really needed at various correlation lengths. The minimum N, denoted as N_c, can be found from (5.2) to achieve a given error, say, $\varepsilon_c = 5\times10^{-4}$. The critical energy E_c can then be calculated corresponding to this critical random dimensionality N_c. As shown in Fig. 8(a) (curve with triangles), the critical energy decreases exponentially with the decrease of the correlation length. The combination of the slower decay of eigenvalues and the lower level energy required leads to only a slight increase in the critical random dimensionality N_c in the case of a smaller correlation scale (shown in Fig. 8(b)). By comparing the curves of critical dimensionality get from these two criterions, we can see, with a given energy, one may not obtain converged results for the case of large correlation length whereas one would retain too many terms in the KL expansion for the case of small correlation length and thus require unnecessarily high computational efforts. Therefore, the energy criterion is not appropriate for determining the critical random dimensionality. This finding has not been reported in the literature. But it has a significant implication on the application of the SCM and

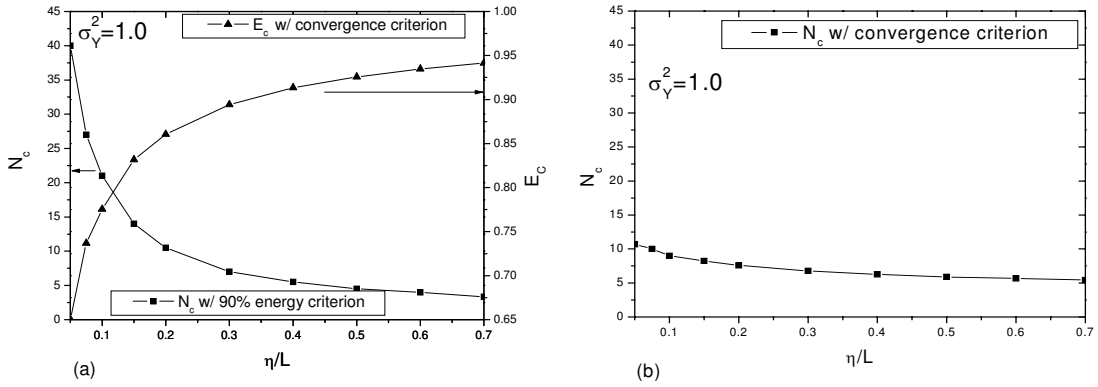

Figure 8: (a) The critical random dimensionality N_c obtained with the 90% energy criterion (curve with squares); the critical energy E_c obtained with the convergence criterion (curve with triangles). (b) The critical random dimensionality N_c obtained with the convergence criterion.

PCM: The curse of dimensionality may be less severe than previously thought in the case of small correlation scale. Since it is only based on numerical experiments, it cannot be regarded as a concrete conclusion. But it may provide a practical guidance for the applications of the SCM and PCM and for further theoretical studies.

With the critical random dimensionality needed to retain, we can now perform comparisons between PCM and Smolyak method about their error properties. Denote $d1$ as the order of polynomial chaos expansion and $d2$ as the level of the Smolyak method. We know that the total number of collocation points, M, depends on $d1$ (or $d2$) and the random dimensionality, N. We compute the errors of the 2^{nd} order, 4^{th} order, and 6^{th} order PCM as well as the level-1, level-2, and level-3 Smolyak methods according to equation (5.1). Here we only use the even order PCM on the basis of the findings in [12]. As shown in Fig. 9, when the number of the collocation points is small, the error of the PCM is less than that of the Smolyak method for the cases of different correlation lengths and different spatial variability levels. When the number of the collocation points is large, the Smolyak method is more accurate than the PCM. As shown in Fig. 9(d), for the case of large spatial variability, $\sigma_Y^2 = 4.0$, the error of PCM does not seem to converge. To further test this property, we investigate another case whose correlation length equals to 2 and find the same phenomena. As shown in Fig. 9(f), the error increases as the order of PCM increases. It is also found that the error of Smolyak method always decreases as the number of collocation points increases. Therefore, for the case of large spatial variability if a high accuracy is desired and a large computational effort can be afforded, the high level Smolyak method is a better choice compared to the high order PCM.

5.4 Illustrative examples in 2D

In this section, we consider four cases in a two-dimensional (2D) domain of saturated heterogeneous medium, which is a square of size $L_1 = L_2 = 10$[L], uniformly discretized into

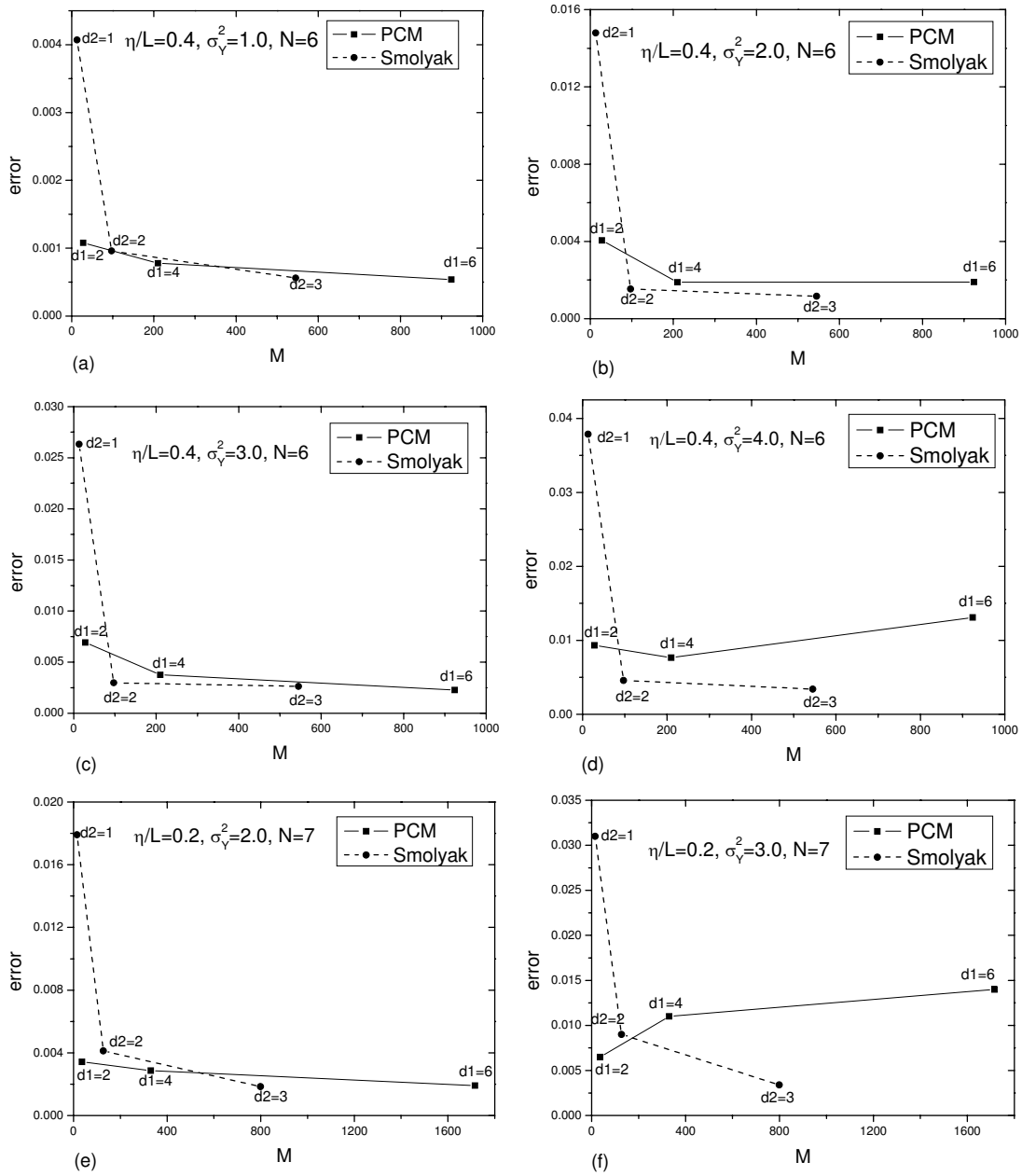

Figure 9: Comparisons of the error of the hydraulic head variance derived from the PCM and Smolyak method for the cases of different correlation lengths and spatial variability levels.

40×40 square elements. The no-flow conditions are prescribed at two lateral boundaries. The hydraulic head is prescribed at the left and right boundaries as 7[L] and 5[L], respectively. The mean of the log hydraulic head is given as $<Y>=0.0$. Assume the covariance

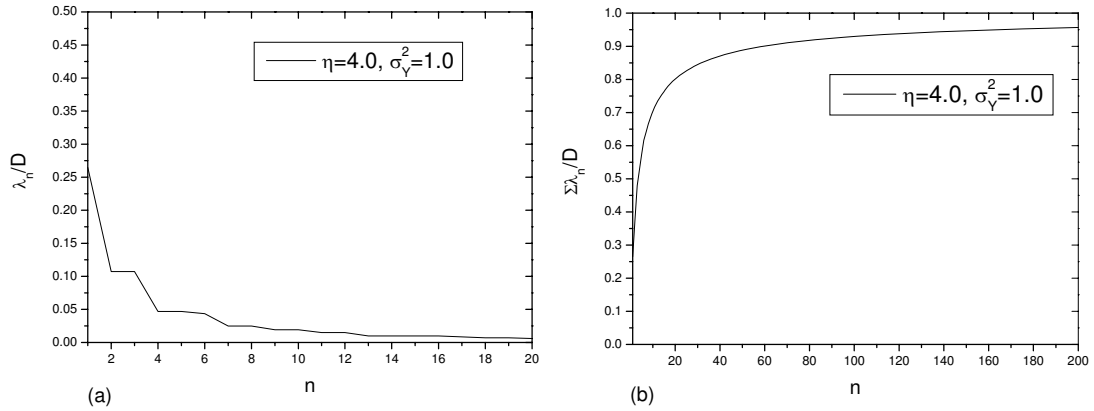

Figure 10: Series of eigenvalues and their finite sums for two-dimensional square flow domain with a separable covariance function, for $\eta/L=0.4$ and $\sigma_Y^2=1.0$.

function of the log hydraulic conductivity is $C_Y(\boldsymbol{x},\boldsymbol{y})=\sigma_Y^2\exp(-|x_1-y_1|/\eta_1-|x_2-y_2|/\eta_2)$ where $\eta_1=\eta_2=\eta=4.0$ or 1.0 and $\sigma_Y^2=1.0$ or 2.0. There is no source term, i.e., $g(x)=0$. Fig. 10 shows the series of eigenvalues and their finite sums. By comparing Fig. 2(a) and Fig. 10(a), it is found that the eigenvalue decay ratio is slower in 2D. One would have to retain a large number of terms in the KL expansion, thus a large random dimensionality (say, $N=100$), for a certain level of energy (say, 90%). This would lead to a huge computational effort for the SCM and PCM and may even make them either less efficient than Monte Carlo or impractical computationally.

As in 1D, we use the convergence criterion to determine the critical random dimensionality. We choose $\varepsilon_c=5\times10^{-4}$ and $n=5$. Here we use a value of n other than one due to the fact that for the 2D cases some of the neighboring eigenvalues may be the same or very similar, which may lead to false convergence if comparing them. In Table 1, we list the values of N_c and E_c for one and two dimensional cases with three different correlation lengths. It is seen that we need to use a larger random dimensionality in the two dimensional cases than in the one dimensional cases with the same correlation length. For E_c, when $\eta/L=0.7$ or 0.4, we have $E_{c2}\approx E_{c1}^2$. But when $\eta/L=0.1$, E_{c2} is less than E_{c1}^2, where the digitals 1 and 2 denote the one and two dimensional case, respectively. This finding is both new and surprising. It may lead to lessening the computational requirements for the SCM and PCM. As most of previous studies of the SCM and PCM are in one physical dimension, this finding may provide a practical guidance for extending the SCM and PCM to multiple dimensions while still retaining their computational efficiency compared to the MC.

Fig. 11 compares the head variance from the Monte Carlo simulations, the second order PCM and the level-2 Smolyak method along the cross section $x_2=5.0$. For comparison, we conduct Monte Carlo simulations using 10,000 two-dimensional realizations generated on the grid of 41×41 nodes with the separable covariance function, based on equation (3.3) with 200 terms. It is shown that the second order PCM and the level-2

Table 1: The random dimensionality and energy needed for one and two dimensional cases with respect to different correlation lengths. In the subscript of the variables, the digitals 1 and 2 denote the one and two dimensional case respectively.

η/L	N_{c1}	N_{c2}	E_{c1}	E_{c2}
0.7	5	15	0.94	0.87
0.4	6	20	0.91	0.80
0.1	9	30	0.77	0.44

Smolyak method agree with the Monte Carlo result fairly well for $\sigma_Y^2=1.0$. For the case of larger spatial variability, $\sigma_Y^2=2.0$, the second order PCM is not as accurate as the level-2 Smolyak method but still acceptable. It is seen that the second order PCM uses a much smaller number of collocation points than the level-2 Smolyak method. This leads to a significant saving in computational effort for the reason that the two dimensional case has much more physical nodes than does the one dimensional case.

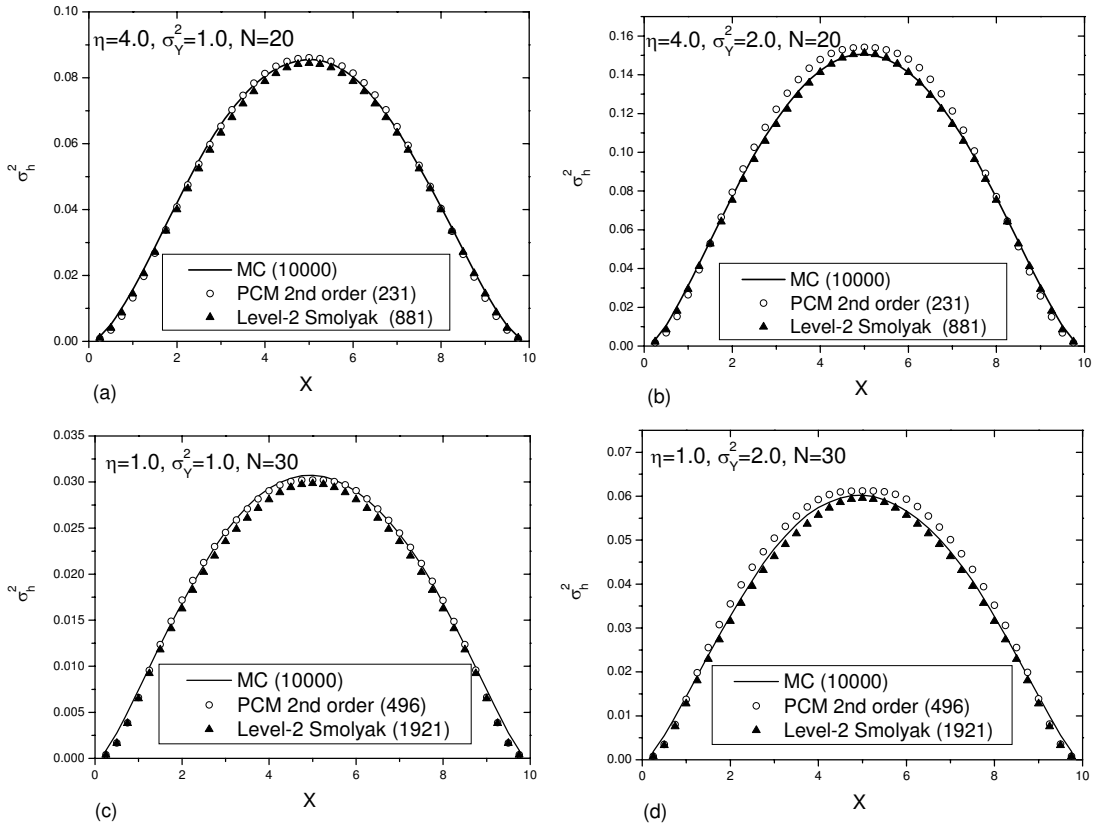

Figure 11: Head variance derived from the MC, PCM and Smolyak method at $x_2=5.0$.

6 Conclusions

In this study, we applied the stochastic collocation methods (SCM) and probability collocation method (PCM) to the problem of flow in spatially correlated random fields, with different spatial variabilities and correlation lengths. We performed comparisons between them about their accuracy and efficiency. This study leads to the following observations and conclusions:

1. Both the second order PCM and the level-1 Smolyak method require fewer collocation points and thus lead to less computational efforts compared to other methods. At small spatial variability, both the level-1 Smolyak method and the second order PCM are efficient and adequately accurate. At moderate spatial variability, the second order PCM is superior to the level-1 Smolyak in terms of its performance.

2. For the case of large spatial variability, the fourth order PCM and the level-2 Smolyak method are required to obtain accurate results. Although the two methods have nearly the same accuracy, the level-2 Smolyak method requires a smaller number of collocation points and is thus computationally more efficient than the fourth order PCM.

3. In the case of extremely large spatial variability, the Smolyak method improves with the level of the method while the PCM may not yield better results with the increase of its order. Therefore, when a high accuracy is desirable and if a large computational effort can be afforded, the high level Smolyak method is a better choice than the high order PCM.

4. The error studies indicate that the second order PCM provides much better approximations that does the level-1 Smolyak method and yields almost the same accuracy as the level-2 Smolyak method but with a much less computational effort. Compared to the second order PCM the higher order PCM does not improve the accuracy significantly (or, leads to a worse performance in the case of extremely large variability) despite a large increase in the computational requirement. Therefore, the second order PCM provides a practical approach, which is adequately accurate and sufficiently efficient, for large-scale problems.

5. It is found that Stroud method uses the fewest collocation points but gives unacceptable accuracy for our problem. The adaptive Stroud method leads to a significant improvement compared to the Stroud method. The adaptive Stroud method uses the same number of collocation points as the second order PCM for the case $N=6$, but its result is not as accurate as the latter. Furthermore, the adaptive Stroud method may not be applicable for the problems with high random dimensionality.

6. A stochastic collocation method cannot give accurate results when the spatially correlated input random fields are not approximated with the adequate number of terms (modes) in the Karhunen-Loeve expansion, which is the so called retained random dimensionality. It is found that the higher the (retained) random dimensionality, the better the results. However, for all the stochastic collocation methods discussed in this study, the computational effort increases rapidly with this random dimensionality. Therefore, it is desirable to have a priori error estimator, or a practical guidance, for estimating the

optimal random dimensionality that leads to sufficiently accurate results with minimum computational efforts.

7. It is well known that in order to retain the same level of energy in the KL expansion, the number of required modes increases exponentially with the decrease of the correlation scale relative to the domain size. However, it is found on the basis of our numerical experiments that the level of energy required in the truncated expansion decreases exponentially with the decrease of the correlation scale. This leads to the net effect of that the required random dimensionality increases only slowly with the decrease of the correlation scale. This new finding is important for the application of the SCM and PCM to random fields of small correlation length and may lead to lessening the curse of dimensionality for the SCM and PCM.

8. Most of previous studies in the SCM and PCM are in one physical dimension. The KL expansion decays much more slowly in multiple physical dimensions than in one dimension, thus requiring a large number of terms to keep the same level of energy. It is well know that in the case of large random dimensionality, the SCM and PCM approaches may become either less efficient than Monte Carlo or impractical computationally. Our numerical experiments reveal that the required level of energy is significantly reduced from one to two dimensions so that the random dimensionality only increases moderately. This has a great implication on the extension of the SCM and PCM to multiple physical dimensions.

9. In this study, the findings are based on numerical experiments and associated analyses. Although they may provide some practical guidance on the applications of the stochastic collocations methods, these findings cannot be regarded as concrete conclusions without further rigorous mathematical analysis. However, this study motivates theoretical studies on these important issues.

Acknowledgments

The authors are grateful to the supports by Natural Science Foundation of China through grant 50688901, the Chinese National Basic Research Program through grant 2006CB705800, and the U. S. National Science Foundation through grant 0801425. The first author acknowledges the support by China Scholarship Council through grant 2007100458.

References

[1] I. Babuska, F. Nobile and R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. Technical report, MOX, Dipartimento di Matematica, 2005.

[2] I. Babuska, R. Tempone and G. E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, SIAM J. Number. Anal. 42 (2004), 800-825.

[3] I. Babuska, R. Tempone and G. E. Zouraris, Solving elliptic boundary value problems with uncertain coefficients by the finite element method: The stochastic formulation. Comput. Methods Appl. Mech. Engrg., 194(12-16) (2005), 1251-1294.
[4] 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.
[5] C. Canuto, M. Y. Hussaini, A. Quateroni and T. A. Zang, Spectral Methods in Fluid Dynamics, Springer-Verlag, New York, 1988.
[6] M. K. Deb, I. Babuska and J. T. Oden, Solution of stochastic partial differential equations using Galerkin finite element techniques, Comput. Methods Appl. Mech. Engrg., 190 (2001), 6359-6372.
[7] Y. Ding, T. Li, D. Zhang and P. Zhang, Adaptive Stroud stochastic collocation method for flow in random porous media via Karhunen-Loeve expansion. Commun. Comput. Phys., 4(2008), 102-123.
[8] R. Ghanem, Scales of fluctuation and the propagation of uncertainty in random porous media, Water Resour. Res., 34(9) (1998), 2123-2136.
[9] R. Ghanem and P. Spanos, Stochastic Finite Element: A Spectral Approach, Springer, New York, 1991.
[10] A. Keese, A review of recent developments in the numerical solution of stochastic partial differential equations (stochastic finite elements), Institute fur Wissenschaftliches Rechnen, 2003.
[11] Y. S. Kumar, J. Li, C. Talarico and J. Wang, A probabilistic collocation method based statistical gate delay model considering process variations and multiple input switching, Proceedings of Design, Automation and Test in Europe, 7-11 March (2005), 770-775.
[12] H. Li and D. Zhang, Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods. Water Resources Research, 43 (2007), W09409, doi:10.1029/2006WR005673.
[13] L. Mathelin, M. Y. Hussaini and T. A. Zang, Stochastic approaches to uncertainty quantification in CFD simulations, Numerical Algorithms, 38 (2005), 209-236.
[14] I. P. Mysovskih, Proof of the minimality of the number of nodes in the cubature formula for a hypersphere, USSR Comput. Math. Math. Phys., 6 (1966), 15-27.
[15] F. Nobile, R. Tempone and C. G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Sci. Comput., 46 (2008), 2309-2345.
[16] S. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet Math. Dokl., 4 (1963), 240-243.
[17] A. H. Stroud, Remarks on the disposition of points in numerical integration formulas, Mathematical Tables and Other Aids to Computation, 11 (1957), 257-261.
[18] M. A. Tatang, W. Pan, R. G. Prinn and G. J. McRae, An efficient method for parametric uncertainty analysis of numerical geophysical models, J. Geophys. Res., 102(D18) (1997), 21,925-21,931.
[19] G. W. Wasilkowski and H. Wozniakowski, Explicit cost bounds of algorithms for multivariate tensor product problems. J. Complexity, 11 (1995), 1-56.
[20] M. Webster, M. A. Tatang and G. J. Mcrae, Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model, MIT Joint Program on the Science and Policy of Global Change Report Series No. 4, Massachusetts Institute of Technology, 1996.

[21] N. Wiener, The homogeneous chaos, Am. J. Math., 60 (1938), 897-936.
[22] D. Xiu and J. S. Hesthaven, High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput., 27(3) (2005), 1118-1139.
[23] 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.
[24] D. Xiu and G. E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput., 24 (2002), 619-644.
[25] D. Zhang, Stochastic Methods for Flow in Porous Media: Coping with Uncertainties, Academic Press, San Diego, CA, 2002.
[26] 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.