Journal of Computational Physics 229 (2010) 6995–7012

Contents lists available at ScienceDirect

Journal of Computational Physics

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

Uncertainty quantification via random domain decomposition and probabilistic collocation on sparse grids

G. Lin [a,*], A.M. Tartakovsky [a], D.M. Tartakovsky [b]

[a] Pacific Northwest National Laboratory, Richland, WA 99352, USA
[b] Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 27 October 2009
Received in revised form 12 April 2010
Accepted 25 May 2010
Available online 2 June 2010

Keywords:
Uncertainty quantification
Random composite
Polynomial chaos
Stochastic finite element
Stochastic collocation method

ABSTRACT

Quantitative predictions of the behavior of many deterministic systems are uncertain due to ubiquitous heterogeneity and insufficient characterization by data. We present a computational approach to quantify predictive uncertainty in complex phenomena, which is modeled by (partial) differential equations with uncertain parameters exhibiting multi-scale variability. The approach is motivated by flow in random composites whose internal architecture (spatial arrangement of constitutive materials) and spatial variability of properties of each material are both uncertain. The proposed two-scale framework combines a random domain decomposition (RDD) and a probabilistic collocation method (PCM) on sparse grids to quantify these two sources of uncertainty, respectively. The use of sparse grid points significantly reduces the overall computational cost, especially for random processes with small correlation lengths. A series of one-, two-, and three-dimensional computational examples demonstrate that the combined RDD–PCM approach yields efficient, robust and non-intrusive approximations for the statistics of diffusion in random composites.

Published by Elsevier Inc.

1. Introduction

Predictions of the behavior of deterministic physical systems are often uncertain because such systems tend to be heterogeneous and under-parameterized by data. In addition to being sparse, parameter data are often corrupted by measurement and interpretive errors. Modeling of subsurface flow and transport is a case in point. While it relies on equations (e.g. a diffusion equation used to describe single-phase flow in porous media or an advection–dispersion equation used to model solute transport) that are inherently deterministic, subsurface heterogeneity, data sparsity, and low-resolution indirect measurements of the parameters entering these equations (e.g. hydraulic conductivity and dispersion coefficients) render model predictions highly uncertain. A prevailing approach to quantifying predictive uncertainty is to treat such parameters as random fields, whose statistics are inferred from available data. The corresponding governing equations become stochastic [1,2].

Monte Carlo simulations (MCS) are routinely used to solve stochastic partial differential equations (SPDEs), but they are computationally prohibitive. Perturbation-based moment equations have been widely used to quantify uncertainty in subsurface modeling, e.g. [1–8]. Since these approaches employ the variance of log-hydraulic conductivity as a small perturbation parameter, they are formally applicable to mildly-to-moderately heterogeneous media. Moreover, perturbation-based moment solutions of some stochastic equations (e.g. advection–dispersion equations, which are the focus of this analysis)

* Corresponding author. Tel.: +1 509 372 6596.
E-mail addresses: guang.lin@pnl.gov (G. Lin), dmt@ucsd.edu (D.M. Tartakovsky).

0021-9991/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jcp.2010.05.036

may diverge for any non-zero variances of the hydraulic conductivity [9]. Stochastic finite elements represent a "nonperturbative" alternative to MCS and moment equations that, under certain conditions discussed below, is computationally more efficient than either of the two. The classical polynomial chaos [10], which relies on the Hermite orthogonal polynomials to represent a second-order stochastic process by a spectral expansion in terms of Gaussian random variables, has been used to solve SPDEs via a Galerkin projection (see [11] and the references therein). The generalized polynomial chaos [12] extends its range of applicability by employing orthogonal polynomials from the Askey scheme. Similar to moment equations, polynomial-chaos expansions result in deterministic equations that can be different from the underlying governing equations and, hence, require one to modify existing deterministic (legacy) codes. Such approaches are sometimes referred to as "intrusive".

Probabilistic collocation methods (PCMs) [13–27], which combine the strengths of MCS and stochastic Galerkin methods, provide a non-intrusive alternative. They utilize the theory of multivariate polynomial interpolations [28,29] to achieve fast convergence when solutions of SPDEs possess sufficient smoothness in the random space. Implementation of PCMs is straightforward, since it only requires one to solve of the underlying deterministic problems at pre-selected sampling points. The choice of these sampling or collocation points is based on sparse grids obtained with the Smolyak algorithm [30]. This offers high-order accuracy and convergence rates that are less dependent on dimensionality than are those of polynomial-chaos expansions.

The computational cost of stochastic finite element approaches, both intrusive and non-intrusive, might become prohibitive when they are used to analyze highly heterogeneous systems, i.e. when system parameters have a large number of random dimensions. In the present study, we address this open issue by introducing a two-scale framework that combines a random domain decomposition (RDD) [31–33] and the PCM approach on sparse grids. The RDD takes advantage of the fact that multi-modality of, and the lack of stationarity (statistical homogeneity) in, a system parameter Y typical indicates that it is sampled from a union of ensembles, each of which is unimodal and stationary (see also [34]). The RDD uses parameter data to reconstruct probabilistically the spatial extent of distinct populations [35–38], which allows one to replace a non-Gaussian, nonstationary distribution of Y with a joint distribution of random sub-domains (boundaries between populations) and sub-domains' properties. RDD has been combined with the generalized polynomial chaos to model diffusion in a one-dimensional random composite [39].

The paper is organized as follows. A general formulation of steady-state diffusion in random composites (or single-phase flow in highly heterogeneous porous media) is presented in Section 2. Section 3 contains probabilistic representations of the system parameters describing these phenomena. An outline of the PCM approach on sparse grids is provided in Appendix A for interested readers. In Section 5 we use the RDD to decrease the computational cost of the PCM approach and to extend its range of applicability. The accuracy and convergence of the resulting RDD–PCM approach are analyzed in Section 6 by considering two one-dimensional problems. Finally, Section 7 presents the results of two- and three-dimensional simulations.

2. Governing equations

Consider a Poisson equation:

$$\nabla \cdot [K(\mathbf{x};\omega)\nabla h(\mathbf{x};\omega)] + f(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathcal{D}, \quad \omega \in \Omega, \tag{1}$$

which describes steady-state single-phase flow in a (randomly heterogeneous) porous medium $\mathcal{D}$, among many other phenomena. Uncertainty in the hydraulic conductivity of a heterogeneous porous medium $K(\mathbf{x})$ is quantified by treating it as a random field $K(\mathbf{x};\omega)$, so that K varies in both the physical space, $\mathbf{x} \in \mathcal{D}$, and the sample space, $\omega \in \Omega$, with ω representing a random event in the sample space Ω. In applications, ensemble statistics of the random field $K(\mathbf{x};\omega)$, including its correlation structure, are inferred from a set of N_m measurements $K_i \equiv K(\mathbf{x}_i)$ $(i = 1,\ldots,N_m)$ by invoking ergodicity [1,2]. As a solution of (1), the hydraulic head $h(\mathbf{x})$ becomes random as well, $h(\mathbf{x};\omega)$. The source function $f(\mathbf{x})$ might be either random or deterministic.

The Poisson Eq. (1) is subject to the boundary conditions:

$$h(\mathbf{x};\omega) = H(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D, \quad \omega \in \Omega, \tag{2a}$$

$$K(\mathbf{x};\omega)\mathbf{n}(\mathbf{x}) \cdot \nabla h(\mathbf{x};\omega) = G(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N, \quad \omega \in \Omega, \tag{2b}$$

where $H(\mathbf{x})$ and $G(\mathbf{x})$ are the forcing functions prescribed on the Dirichlet, Γ_D, and Neumann, Γ_N, segments of the boundary $\partial\mathcal{D} = \Gamma_D \cup \Gamma_N$, and $\mathbf{n}$ is the outward unit normal vector. To simplify the presentation, we treat the forcing functions f, H and G as deterministic, i.e. known with certainty. Randomness (uncertainty) in these functions can be readily accounted for since their effects are additive [40].

Physical considerations dictate that (random) hydraulic conductivity be non-negative, $K(\mathbf{x};\omega) > 0$, almost everywhere (a.e.) in $\omega \in \Omega$. Some reservations (e.g. [41] and references therein) notwithstanding, it is common to treat $Y = \ln K$ as a stationary multivariate-Gaussian random field (e.g. [1,2] and references therein). Such parameterizations of (1) are expected to fail if a porous medium is highly heterogeneous, consisting of multiple inhomogeneous materials. Even after removing a deterministic trend (the ensemble mean) $\overline{Y}(\mathbf{x})$, the random increments $Y'(\mathbf{x}) \equiv Y(\mathbf{x}) - \overline{Y}(\mathbf{x})$ remain nonstationary, exhibiting variances that vary from one constitutive material to another and different degrees of correlation within each material and between different materials [42]. The stochastic algorithm described in Sections 3–5 provides an efficient computational tool for quantifying uncertainty in such systems, and allows one to deal with highly non-Gaussian system parameters.

3. Stochastic representations of system parameters

Efficient representations and generations of nonstationary (statistically inhomogeneous) random fields with arbitrary distributions and correlation structures are elusive. It is at this first stage of uncertainty quantification—probabilistic representation of uncertain system parameters—that the advantages of RDD become apparent. Suppose that an analysis of system-parameter data $\{Y_i\}_{i=1}^{N_m}$ enables one to identify probabilistically the spatial extent of N_p distinct populations Ω_n ($n = 1,\ldots,N_p$), i.e. to decompose the d-dimensional flow domain $\mathcal{D} \in \mathbb{R}^d$ into N_p sub-domains $\mathcal{D}_n$ (such that $\mathcal{D} = \cup_{n=1}^{N_p}\mathcal{D}_n$ and $\mathcal{D}_n \cap \mathcal{D}_k$ for $n \neq k$), each of which is characterized by a second-order continuous stationary multivariate-Gaussian field $Y_n(\mathbf{x};\omega)$. (Such an analysis lies outside the scope of the present study; the tools developed in [35–38], among others, can be used to achieve this goal.) Then one can utilize the well-developed theory of stationary random processes (e.g. [34,43]) to represent random fields $Y_n(\mathbf{x};\omega)$ ($n = 1,\ldots,N_p$) with a required degree of accuracy.

Specifically, if $Y_n(\mathbf{x};\omega)$ is a second-order stationary random process that is mean-square continuous on the closure of $\mathcal{D}_n$, i.e. if for any $\mathbf{x},\mathbf{y} \in \overline{\mathcal{D}}_n \equiv \mathcal{D}_n \cup \partial\mathcal{D}_n$:

$$\mathbb{E}\left[Y_n(\mathbf{x},\omega)^2\right] < \infty \quad \text{and} \quad \lim_{\mathbf{y}\to\mathbf{x}} \mathbb{E}\left[|Y(\mathbf{y},\omega) - Y(\mathbf{x},\omega)|^2\right] = 0, \tag{3}$$

then its best linear approximation (in the mean-square sense) is provided by the Karhunen–Loève (K–L) expansion [43]:

$$Y_n(\mathbf{x};\omega) = \overline{Y}_n + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\psi_i(\mathbf{x})\xi_i(\omega), \quad \mathbf{x} \in \mathcal{D}_n, \quad \omega \in \Omega_n. \tag{4}$$

Here $\xi_i(\omega)$ are mutually uncorrelated random variables with zero-mean and unit variance; and λ_i and $\psi_i(\mathbf{x})$ are the eigenvalues and corresponding eigenfunctions of the covariance function $C_{Y_n}(\mathbf{x},\mathbf{y})$, which satisfy a Fredholm equation:

$$\int_{\mathcal{D}_n} C_{Y_n}(\mathbf{x},\mathbf{y})\psi(\mathbf{y})\mathrm{d}V_{\mathbf{y}} = \lambda\psi(\mathbf{x}), \quad \mathbf{x} \in \mathcal{D}_n, \tag{5}$$

where $\mathrm{d}V_{\mathbf{y}} = \mathrm{d}y_1,\ldots,\mathrm{d}y_d$. In the analysis below, we take the random variables ξ_i to be Gaussian and the covariance function $C_{Y_n}(\mathbf{x},\mathbf{y}) \equiv C_{Y_n}(|\mathbf{x}-\mathbf{y}|)$ to be separable exponential, i.e. $C_{Y_n}(|\mathbf{x}-\mathbf{y}|) = \sigma_{Y_n}^2 \exp\left(-\sum_{i=1}^d |x_i - y_i|/l_{n_i}\right)$ where $\sigma_{Y_n}^2$ is the variance of Y in the n-th sub-domain $\mathcal{D}_n$ and $l_{n_i} (i = 1,\ldots,d)$ are the correlation lengths along the $x_1,\ldots,x_d$ coordinate axes, respectively. Note that here and below we use $\mathbb{E}[A]$ and $\overline{A}$ interchangeably to denote the ensemble mean of a random field A.

In numerical simulations, the infinite series in (4) must be truncated at a finite index number M_n:

$$Y_{M_n}(\mathbf{x};\omega) = \overline{Y}_n + \sum_{i=1}^{M_n} \sqrt{\lambda_i}\psi_i(\mathbf{x})\xi_i(\omega). \tag{6}$$

The corresponding optimal mean-square truncation error:

$$\int_{\mathcal{D}_n} \mathbb{E}\left[|Y_n(\mathbf{x};\omega) - Y_{M_n}(\mathbf{x};\omega)|^2\right]\mathrm{d}V_{\mathbf{x}} = \sum_{i=M_n+1}^{\infty} \lambda_i, \tag{7}$$

depends on the decay rate of the eigenvalues λ_i ($i \geqslant 1$). For general second-order stationary random fields, the decay rate of λ_i depends on two factors [44]: regularity of the correlation function [45] and its correlation length. For the separable exponential correlation function C_{Y_n} considered in this study, the eigenvalue decay rate is completely determined by the correlation lengths l_{n_i}.

Let us sort the eigenvalues λ_i ($i \geqslant 1$) in the monotonically decreasing order before truncating the K–L expansion. Then a truncation error criterion:

$$\sum_{i=1}^{M_n} \lambda_i \leqslant 0.9 \sum_{i=1}^{\infty} \lambda_i, \tag{8}$$

which aims to retain 90% of the full eigen-spectrum, is used to determine the appropriate number of random dimensions M_n for each sub-domain $\mathcal{D}_n$ ($n = 1,\ldots,N_p$). The criterion (8) enables one to find M_n from the eigenvalue decay rate prior to solving the stochastic problem (1)–(2). For large correlation lengths, the eigenvalues decay very fast [46], so that only a small number of terms must be kept in the K–L expansion. For small correlation lengths, the eigenvalues decay slower and more terms are needed in the K–L expansion. Moreover, the spatial resolution must be increased to capture small-scale features associated with short correlation lengths.

For a general d-dimensional domain $\mathcal{D} \in \mathbb{R}^d$ and for an arbitrary correlation function C_Y, the eigenvalue problem (5) has to be solved numerically. If the domain is a d-dimensional cube, $\mathcal{D} = \{\mathbf{x} : 0 \leqslant x_i \leqslant L_i, i = 1,\ldots,d\}$, and if the correlation function $C_Y(|\mathbf{x}-\mathbf{y}|) = \sigma_Y^2 \exp\left(-\sum_{i=1}^d |x_i - y_i|/l_i\right)$, then the eigenvalues λ_i and the corresponding eigenfunctions ψ_i are computed analytically by combining one-dimensional eigenvalues and eigenfunctions for the one-dimensional Fredholm equation (5) (e.g. [46]).

To facilitate computations further, it is common to treat the mutually uncorrelated random variables $\xi_i(\omega)$ in (6) as statistically independent (see the discussion and references in Section 3.3 of [47]). If these random variables are *multivariate* Gaussian, then this treatment is exact since the lack of correlation implies mutual independence.

4. Stochastic numerical methods

Monte Carlo simulations (MCS) and probabilistic collocation methods (PCMs) follow a similar procedure for solving stochastic differential equations. It consists of the following steps: (i) select sampling points in the probability space of random system parameters, (ii) solve corresponding deterministic governing equations at these points, and (iii) estimate statistical moments of the resulting solutions. The main difference between MCS and PCM is the choice of sampling points and appropriate weights associated with these points. To focus on this difference, we employ the same high-order numerical method—a weighted, essentially non-oscillatory scheme (WENO)—to solve deterministic equations in both MCS and PCM. A brief description of PCMs is provided below, and algorithms for choosing the sampling points are discussed in the Appendix A.

4.1. Probabilistic collocation methods

Suppose that in each sub-domain $\mathcal{D}_n$ $(n = 1,\ldots,N_p)$ the random system parameter $Y(\mathbf{x};\omega)$ can be described, with a given degree of accuracy, by M_n independent random variables $\xi_i(\omega)$ $(i = 1,\ldots,M_n)$ in accordance with (6). Then the total number of independent random variables $\xi_i(\omega)$ needed to describe $Y(\mathbf{x};\omega)$ for all $\mathbf{x} \in \mathcal{D} = \cup_{n=1}^{N_p}\mathcal{D}_n$ is $M = \sum_{n=1}^{N_p} M_n$. According to the Doob–Dynkin lemma [48], the same number of random variables is sufficient to describe the system state $h(\mathbf{x};\omega)$ given by a solution of (1)–(2). In other words, $h = h(\mathbf{x};\boldsymbol{\xi})$ where $\boldsymbol{\xi}(\omega) = [\xi_1(\omega),\ldots,\xi_M(\omega)]^T$.

Let $p_i(\xi_i) : \Gamma_i \to \mathbb{R}^+$ denote the probability density function (PDF) of the i-th random variable ξ_i $(i = 1,\ldots,M)$, where the images $\Gamma_i \equiv \xi_i(\Omega)$ are assumed to be intervals in $\mathbb{R}$. Since $\xi_i(\omega)$ $(i = 1,\ldots,M)$ are independent random variables, the joint PDF of $\boldsymbol{\xi}(\omega)$ is given by

$$p(\boldsymbol{\xi}) = \prod_{i=1}^{M} p_i(\xi_i), \quad \boldsymbol{\xi} \in \Gamma, \tag{9}$$

and has the support:

$$\Gamma \equiv \prod_{i=1}^{M} \Gamma_i \subset \mathbb{R}^M. \tag{10}$$

Applying the weighted residual method in the probability space to (1), we obtain:

$$\int_\Gamma \vartheta(\boldsymbol{\xi})\{\nabla\cdot[K(\mathbf{x};\boldsymbol{\xi})\nabla h(\mathbf{x};\boldsymbol{\xi}] + f(\mathbf{x})\}\mathrm{d}V_\xi = 0, \tag{11}$$

where $\mathrm{d}V_\xi \equiv \mathrm{d}\xi_1,\ldots,\mathrm{d}\xi_M$ and $\vartheta(\boldsymbol{\xi})$ represents a family of weight functions. In the Galerkin projection method, weight functions $\vartheta(\boldsymbol{\xi})$ are taken to be basis functions, e.g. the generalized polynomial chaos (gPC) basis functions [12]. The orthogonality of (gPC) basis functions allows one to reduce the stochastic integral Eq. (11) to a set of deterministic equations. These equations can have a functional form that is different from the underlying deterministic equation. Hence one needs to modify an existing numerical code used to solve the deterministic version of (1), a procedure that is sometimes called "intrusive".

In contrast to the Galerkin projection, the collocation formulation employs the Dirac delta functions $\delta(\xi - \xi_k)$ with $k = 1,\ldots,M$ as the weight functions. Applying the collocation projection, $\vartheta(\boldsymbol{\xi}) = \delta(\xi - \xi_k)$, to (11) yields a set of M deterministic equations:

$$\nabla\cdot[K(\mathbf{x};\xi_k)\nabla h(\mathbf{x};\xi_k)] + f(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathcal{D}, \quad k = 1,\ldots,M. \tag{12}$$

The main advantages of the collocation projection method over the Galerkin projection method (e.g. [11,12]) are that this set of deterministic equations is decoupled in random space, and that each of these equations has the same functional form as the underlying governing equation. These advantages hold regardless of whether a differential operator in the governing equation is linear (the case considered here) or nonlinear.

A general PCM procedure is similar to that for MCS, except for the choice of the sampling points ξ_k and the corresponding weights. The procedure consists of the following three steps:

(1) Generate N_c collocation points ξ_k $(k = 1,\ldots,N_c)$ in the probability space of random parameters as independent random inputs based on either the Clenshaw–Curtis formulae for nested sparse grids or the Gauss quadrature formula for non-nested sparse grids.
(2) Solve a deterministic problem (12) at each collocation point ξ_k.
(3) Compute the statistics of $h(\mathbf{x};\omega)$ using the corresponding quadrature rule, e.g.

$$\bar{h}(\mathbf{x}) = \int_\Gamma h(\mathbf{x};\boldsymbol{\xi})p(\boldsymbol{\xi})\mathrm{d}V_\xi \approx \sum_{k=1}^{N_c} \vartheta_k h(\mathbf{x},\xi_k), \tag{13}$$

$$\sigma_h^2(\mathbf{x}) = \int_\Gamma \left[h(\mathbf{x};\boldsymbol{\xi}) - \bar{h}(\mathbf{x})\right]^2 p(\boldsymbol{\xi})\mathrm{d}V_\xi \approx \sum_{k=1}^{N_c} \vartheta_k h(\mathbf{x},\xi_k)^2 - \bar{h}(\mathbf{x})^2, \tag{14}$$

where $\{\vartheta_k\}_{k=1}^{N_c}$ is a set of weights corresponding to the set of quadrature points $\{\xi_k\}_{k=1}^{N_c}$. Extensive reviews on construction of quadrature formulae may be found in [49,50].

4.2. Selection of collocation points

The computational cost of the PCM approach is the number of collocation points times the cost of computing the corresponding deterministic problem. For a given required accuracy, our goal is to select a collocation point set with the minimal number of collocation points. The Appendix A contains a brief overview of two different methods for the selection of collocation point sets: tensor products of one-dimensional collocation point sets (Section A.1) and a sparse grid strategy for high dimensionality (Section A.2).

In the simulations presented below, we employ non-nested sparse grid points, which are selected as follows (see Section A.4 for more details). First, a one-dimensional Gaussian-quadrature rule is constructed for a given probabilistic distribution function. Second, the Smolyak algorithm is employed to select sparse combinations in multi-dimensions.

5. Random domain decomposition (RDD)

We use a RDD [31,32] to improve the efficiency and to speed up the convergence of the PCM for random parameters with multimodal nonstationary distributions. Within the combined RDD–PCM framework, the uncertainty in a system parameter $Y(\mathbf{x})$ is treated as a two-scale random process. Uncertainty in the spatial extent of N_p sub-domains $\{\mathcal{D}_i\}_{i=1}^{N_p}$ is quantified by treating boundaries $\alpha_{ij}(\mathbf{x})$ between sub-domains $\mathcal{D}_i$ and $\mathcal{D}_j (i \neq j)$ as random fields $\alpha_{ij}(\mathbf{x};\omega)$ with a joint PDF $p_\alpha(\alpha)$. Uncertainty in $Y(\mathbf{x})$ within each sub-domain $\{\mathcal{D}_i\}_{i=1}^{N_p}$ is quantified by treating it as a random field with a joint conditional PDF $p_{Yc}(y|\alpha)$. Since each sub-domain $\mathcal{D}_j$ represents a distinct population Ω_i, the conditional PDF $p_{Yc}(y|\alpha)$ is expected to be unimodal and stationary. This process replaces $p_Y(y)$, a multimodal nonstationary distribution for $Y(\mathbf{x})$, with a joint distribution $p_Y(y,\alpha) = p_{Yc}(y|\alpha)p_\alpha(\alpha)$.

The stochastic flow Eq. (1) can now be rewritten as

$$\nabla \cdot [K_i(\mathbf{x};\omega)\nabla h(\mathbf{x};\omega)] + f(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathcal{D}_i, \quad \omega \in \Omega_i, \quad i = 1, \ldots, N_p, \tag{15}$$

where each $Y_i(\mathbf{x}) = \ln K_i(\mathbf{x})$ is a multivariate-Gaussian stationary field. Depending on application, the fields $Y_i(\mathbf{x})$ and $Y_j(\mathbf{x})$ $(i \neq j)$ might or might not be cross-correlated. The relative importance of such a cross-correlation has been studied in [51]. The boundary conditions (2) are now supplemented with the conditions of continuity of the state variable h and the normal component of the flux $K\nabla h$ along each realization of the random boundaries α_{ij}.

Most previous applications of the RDD dealt with MCS and/or perturbation solutions of the moment equations in one and two spatial dimensions [32,33,35,52]. To circumvent the limitations of moment equations and to alleviate the computational burden of MCS, an intrusive RDD-polynomial-chaos approach has been investigated in [39]. Here we introduce a new *non-intrusive* two-scale RDD–PCM framework that combines the RDD and the PCM approach on sparse grids to improve the ability of the latter to handle problems with large numbers of random dimensions.

Let $N_{c,\alpha_{ij}}$ and $N_{c,i}$ $(i, j \in N_p)$ denote the number of collocation points used to represent boundaries $\alpha_{ij}(\mathbf{x};\omega)$ and random parameters $Y_i(\mathbf{x};\omega)$, respectively. Then the total number of collocation points is the tensor product of each set of collocation points, i.e. $\Pi_{i,j\in N_p} N_{c,\alpha_{ij}} \times \Pi_{i=1}^{N_p} N_{c,i}$. To simplify the presentation, we consider flow domains $\mathcal{D} \subset \mathbb{R}^d$ consisting of two sub-domains, $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, separated by a random boundary $\alpha \equiv \alpha_{12}$. In this case, the total number of collocation points is $N_{c,\alpha} \times N_{c,1} \times N_{c,2}$.

5.1. Combined RDD–PCM framework

We assume that the random boundary $\alpha(\mathbf{x};\omega)$ can be parameterized by either a Gaussian or a uniformly-distributed random variable ξ_α. For every realization of α, the K–L expansion (6) is employed to represent the random fields $Y_i(\mathbf{x};\omega)$ within each sub-domain $\mathcal{D}_i$ $(i = 1, 2)$. Let M_{Y_i} denote the number of random dimensions that are required to represent the multivariate-Gaussian field $Y_i(\mathbf{x};\omega)$ $(i = 1, 2)$ with a given degree of accuracy. Then the total number of random dimensions that are require to represent the field $Y(\mathbf{x};\omega)$ in the whole domain $\mathcal{D}$ is $M_T = M_{Y_1} + M_{Y_2} + 1$.

5.1.1. Conditional statistics

Within each sub-domain $\mathcal{D}_i$ $(i = 1, 2)$, the random state variable $h(x;\omega)$ is represented by a polynomial-chaos expansion:

$$h_i(\mathbf{x};\boldsymbol{\xi}) = \sum_{j=1}^{M_{Y_i}} \tilde{h}_{i,j}\phi_j(\boldsymbol{\xi}), \quad \mathbf{x} \in \mathcal{D}_i, \quad \boldsymbol{\xi}(\omega) \in \Omega_i, \tag{16}$$

where $\boldsymbol{\xi}(\omega)$ is a vector of independent random variables. Substituting (16) into (1) and applying a collocation projection onto each of the basis $\{\phi_j\}_{j=0}^{M_{Y_i}}$, we obtain:

$$\nabla \cdot [K_i(\mathbf{x};\boldsymbol{\xi}_k)\nabla h_i(\mathbf{x};\boldsymbol{\xi}_k)] + f = 0, \quad \mathbf{x} \in \mathcal{D}_i, \quad \boldsymbol{\xi}_k(\omega) \in \Omega_i, \quad i = 1, 2, \tag{17}$$

where $\boldsymbol{\xi}_k(\omega)$ represents the k-th set of collocation sampling points. These equations are subject to the boundary conditions (2). At the (random) interface α between the sub-domains $\mathcal{D}_1$ and $\mathcal{D}_2$, the state variable h and the normal component of the (random) flux $\mathbf{J} = -K\nabla h$ are continuous, i.e.

$$h_{1,k}^- = h_{2,k}^+, \quad \mathbf{n} \cdot \mathbf{J}_{1,k}^- = \mathbf{n} \cdot \mathbf{J}_{2,k}^+, \tag{18}$$

where the subscript k represents the k-th set of sparse grid points in random space.

For each set of sparse grid points a system of algebraic equations constructed from (17), (18) and the external boundary conditions (2) is solved. Conditional statistics of h (the conditioning on α is denoted below by the hat $\hat{\cdot}$) are obtained as

$$\widehat{\bar{h}}(\mathbf{x}|\alpha) \approx \sum_{k=1}^{N_{c,i}} \vartheta_{i,k} h_k\left[\mathbf{x}, \boldsymbol{\xi}(\omega)_{i,k}\right], \quad \hat{\sigma}_h^2(\mathbf{x}|\alpha) \approx \sum_{k=1}^{N_{c,i}} \vartheta_{i,k} h_k^2\left[\mathbf{x}, \boldsymbol{\xi}(\omega)_{i,k}\right] - \widehat{\bar{h}}^2, \tag{19}$$

for $i = 1, 2$. Here $x \in D_i$ and $N_{c,i}$ represents the number of sparse collocation points in the i-th sub-domain.

Since the contact boundary α is random, the size of each sub-domain is also subject to uncertainty. For a fixed correlation length of $Y_i(\mathbf{x}; \omega)$ in the sub-domain $\mathcal{D}_i$, the number of the K–L modes required to capture a fixed percentage of the integrated input variance varies depending on the domain size. Hence the number of prematurely truncated K–L modes for each sub-domain varies from one realization of α to another. The combined RDD–PCM approach enables one to obtain the K–L modes in the pre-process step. Such K–L modes are truncated when the smallest eigenvalue is less than 10% of the largest eigenvalue, which varies between different realizations of α. Therefore, large uncertainty in α can lead to different number of prematurely truncated K–L modes for different realizations of the same sub-domain.

5.1.2. Averaging over random α

At the last step, statistics of h are obtained by averaging its conditional statistics over realizations of the random boundary $\alpha(\mathbf{x},\omega)$:

$$\begin{aligned} \bar{h}(\mathbf{x}) &= \int \widehat{\bar{h}}(\mathbf{x}; \xi_\alpha) p_\alpha(\xi_\alpha) \mathrm{d}\xi_\alpha \approx \sum_{j=1}^{N_{c,\alpha}} \vartheta_j \widehat{\bar{h}}\left(\mathbf{x}|\alpha_j\right), \\ \sigma_h^2(\mathbf{x}) &= \int \hat{\sigma}_h^2(\mathbf{x}; \xi_\alpha) p_\alpha(\xi_\alpha) \mathrm{d}\xi_\alpha \approx \sum_{j=1}^{N_{c,\alpha}} \vartheta_j \hat{\sigma}_h^2\left(\mathbf{x}|\alpha_j\right). \end{aligned} \tag{20}$$

6. Accuracy of RDD–PCM approach in one dimension

To analyze the accuracy and efficiency of the combined RDD–PCM approach, we consider the one-dimensional version of (1) with $f = 0$ that is defined on $\mathcal{D} = \{x : 0 \leqslant x \leqslant L\}$. The first computational example (Section 6.1) deals with a random contact point α separating two sub-domains $\mathcal{D}_1 = \{x : 0 \leqslant x < \alpha\}$ and $\mathcal{D}_2 = \{x : \alpha \leqslant x \leqslant L\}$ with random log conductivities $Y_1(x)$ and $Y_2(x)$, respectively. In the second example (Section 6.2), a randomly heterogeneous inclusion $\mathcal{D}_1 = \{x : \alpha - L/6 \leqslant x \leqslant \alpha + L/6\}$ centered at a random point α is embedded into a randomly heterogeneous material $\mathcal{D}_2 = \mathcal{D}/\mathcal{D}_1$. Fig. 1 provides a schematic representation of both examples.

In both examples, the log conductivity $Y_i(x)$ of sub-domain $\mathcal{D}_i$ $(i = 1, 2)$ is treated as a multivariate-Gaussian random field with an exponential correlation function $\rho_{Y_i}(|x - y|) = \exp(-|x - y|/l_i)$ and correlation length l_i. We assume that the random fields Y_1 and Y_2 are mutually uncorrelated. Finally, in both examples the external boundary conditions:

$$K\frac{\mathrm{d}h}{\mathrm{d}x}(x = 0) = -q_0, \quad h(x = L) = 0, \tag{21}$$

are supplemented with the continuity conditions (18) at the random interfaces.

The composite nature of the porous media in Fig. 1 indicates that if log conductivity $Y(x)$ were to be measured at a few locations throughout $\mathcal{D}$ and these data were to be used to compute sample statistics of the random field $Y(x; \omega)$, it would result in a bimodal nonstationary distribution. Fig. 2 in [39] demonstrates that a Gaussian cumulative distribution function (CDF) is accurately approximated with a fifth-degree Hermite polynomial, while even a tenth-degree Hermite polynomial is not adequate to represent a bimodal CDF. This indicates that direct applications of PCMs to systems with bimodal distributions significantly undermines their efficiency. Instead, we use the RDD to decrease the computational cost of the PCM approach and to extend its range of applicability.

All simulations reported in this section (Figs. 2–7 below) correspond to $L = 6.0$, $\bar{\alpha} = 3.0$, $q_0 = 1$, and multivariate log-normal conductivities $Y_1 \sim N\{0,0.1\}$ and $Y_2 \sim N\{2,0.2\}$, while other statistical parameters vary. The uncertain contact point

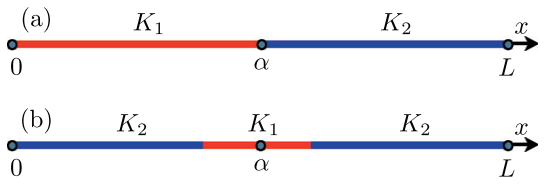

Fig. 1. One-dimensional composites in which (a) two randomly heterogeneous materials are joined at a random location α and (b) a randomly heterogeneous inclusion centered around random α is embedded into another randomly heterogeneous material.

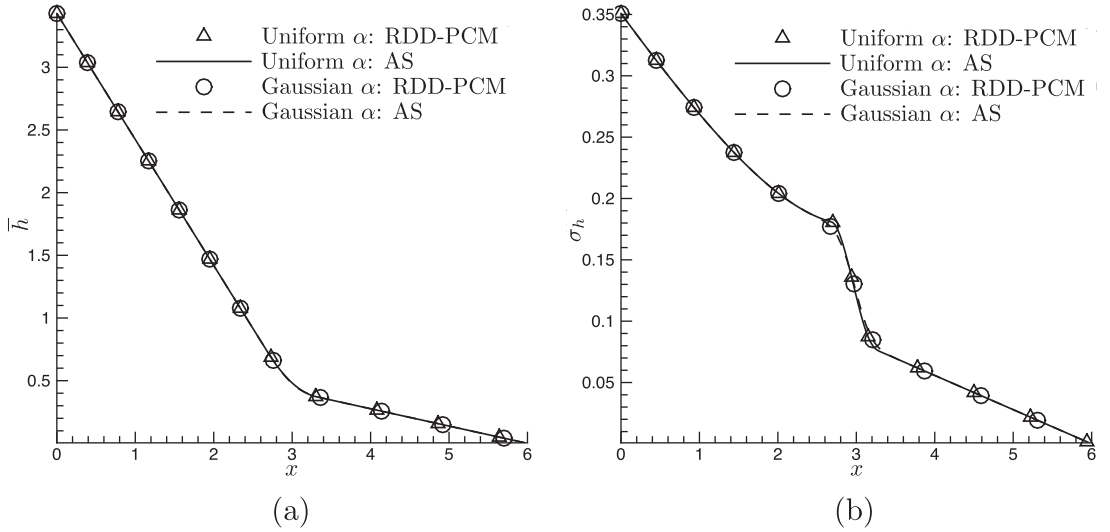

Fig. 2. Contact-point problem: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach and the analytical solutions (AS). Gaussian and uniform distributions of the contact point α are considered.

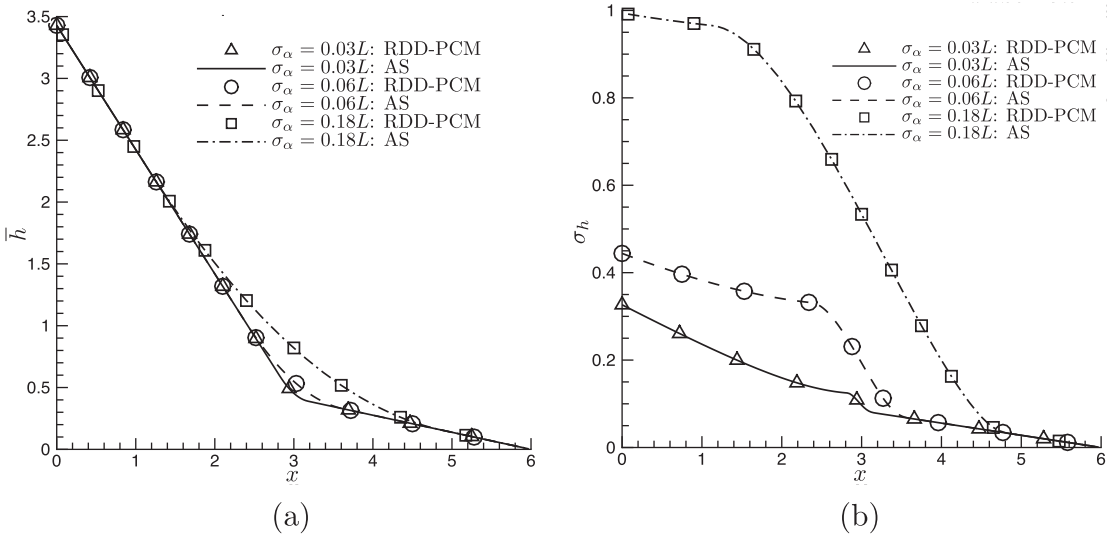

Fig. 3. Contact-point problem: numerical (RDD–PCM) and analytical (AS) solutions for (a) mean $\bar{h}$ and (b) standard deviation σ_h corresponding to several levels of uncertainty in the position of the contact point α.

α is treated alternatively as either a Gaussian or a uniformly-distributed random variable, both distributions having assigned the same mean ($\bar{\alpha} = 3.0$) and variance σ_α^2.

6.1. Random contact point

Consider a one-dimensional version of (1) with $f = 0$ that is defined on $\mathcal{D} = \{x : 0 \leqslant x \leqslant L\}$, subject to the boundary conditions (21) and the continuity conditions at a random contact point $x = \alpha$ separating two sub-domains $\mathcal{D}_1 = \{x : 0 \leqslant x < \alpha\}$ and $\mathcal{D}_2 = \{x : \alpha \leqslant x \leqslant L\}$ with random log conductivities $Y_i(x;\omega) = \ln K_i(x;\omega)$ $(i = 1, 2)$, respectively.

Since all realizations of the random contact point α must fall within the computational domain $0 < \alpha < L$, its mean $\bar{\alpha}$ and standard deviation σ_α cannot be chosen independently. For example, let α be a uniformly distributed random variable with $\bar{\alpha} = L/2$, and let ξ_α be a uniformly-distributed random variable, $U\{-\sqrt{3}, \sqrt{3}\}$, with $\overline{\xi_\alpha} = 0$ and $\sigma_{\xi_\alpha} = 1$. The former is expressed in terms of the latter as $0 < \alpha \equiv L(0.5 + \sigma_\alpha \xi_\alpha) < L$, which yields an upper bound on σ_α:

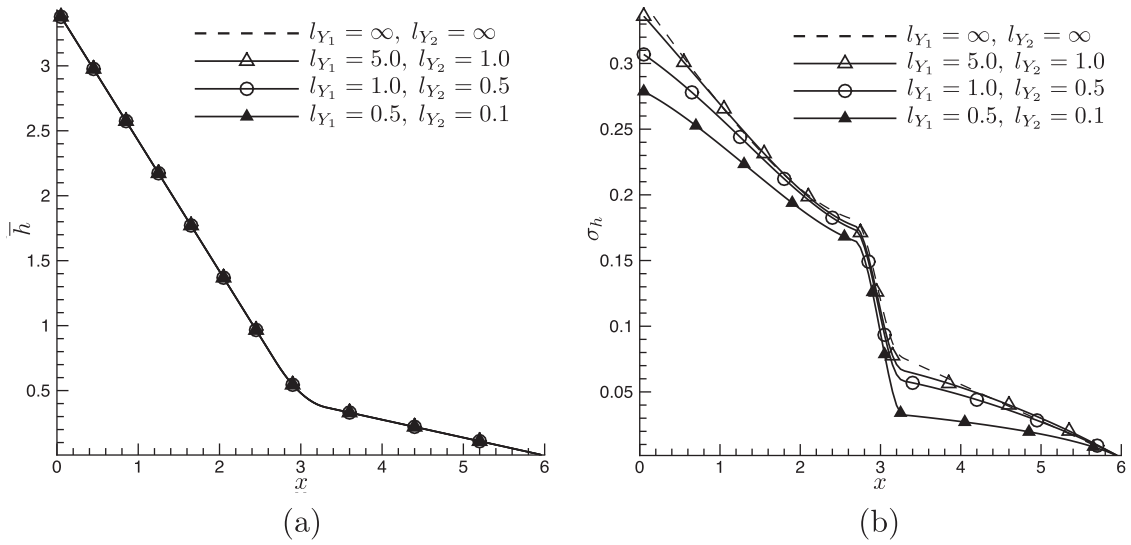

Fig. 4. Contact-point problem: (a) mean $\bar{h}$ and (b) standard deviation σ_h corresponding to different levels of spatial correlation of log conductivity.

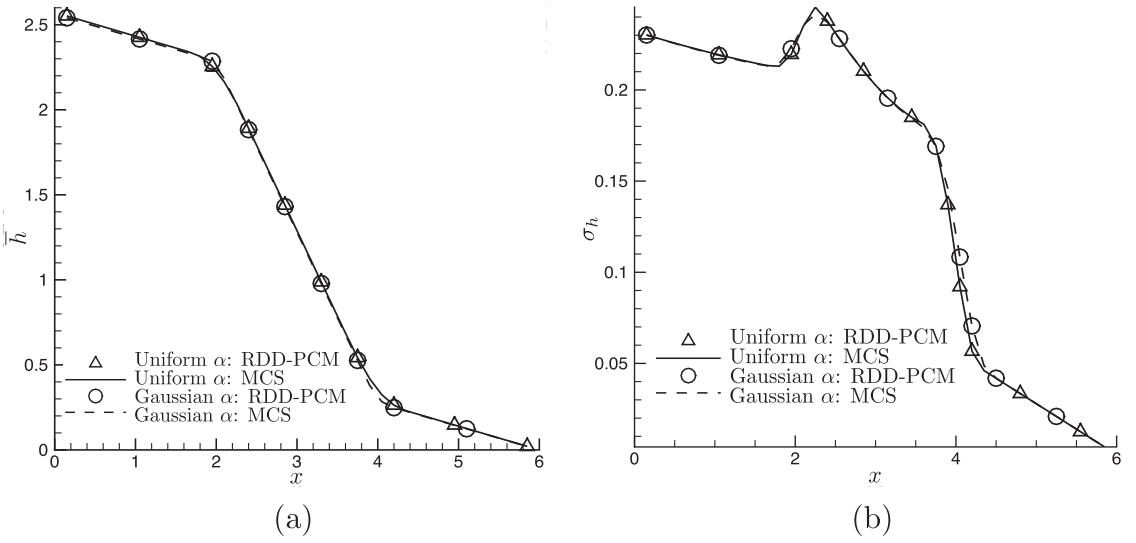

Fig. 5. Inclusion problem: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach and Monte Carlo simulations (MCS).

$$\sigma_\alpha < \frac{\sqrt{3}}{6}. \tag{22}$$

For arbitrary statistical distributions of Y_1, Y_2, and α, the boundary-value problem under consideration admits an analytical solution [32]:

$$h(x) = q_0 \mathcal{H}(\alpha - x)\left[\int_x^\alpha \frac{ds}{K_1(s)} + \int_\alpha^L \frac{ds}{K_2(s)}\right] + q_0 \mathcal{H}(x - \alpha)\int_x^L \frac{ds}{K_2(s)}, \tag{23}$$

where $\mathcal{H}(z)$ is the Heaviside function:

$$\mathcal{H}(z) = \begin{cases} 1, & z \geqslant 0, \\ 0, & z \leqslant 0. \end{cases} \tag{24}$$

The ensemble statistics of $h(x;K_1,K_2,\alpha)$ can now be computed in two steps. First, we obtain conditional statistics by averaging in the space of $K_i(x;\omega)$ $(i = 1, 2)$, while keeping α fixed. Then we average the conditional statistics over realizations of α. This procedure applied to (23) leads to analytical expressions for the conditional ensemble mean of hydraulic head h:

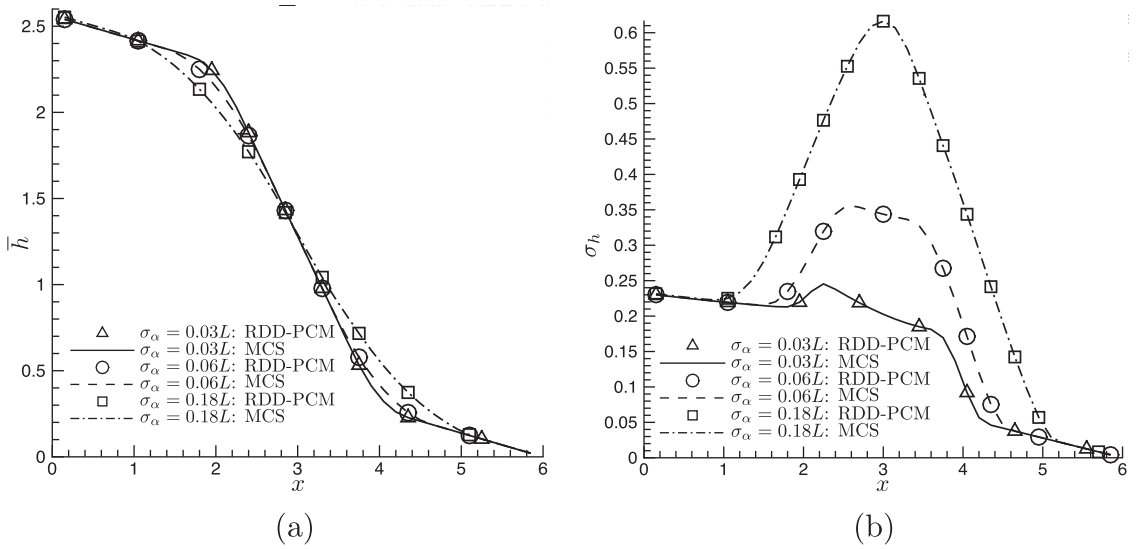

Fig. 6. Inclusion problem: (a) mean $\bar{h}$ and (b) standard deviation σ_h corresponding to several levels of uncertainty in the position of the inclusion's center α.

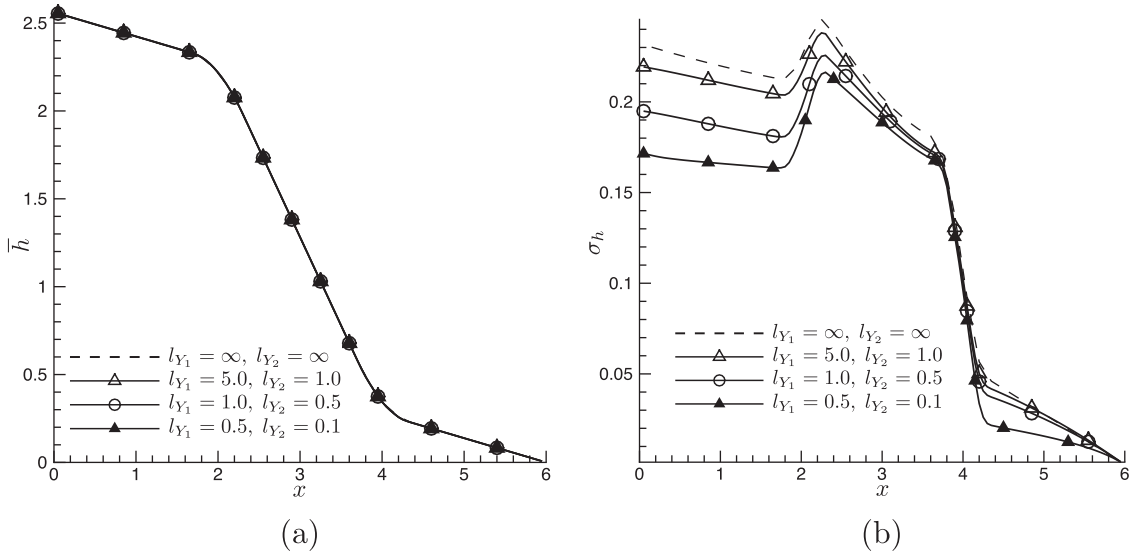

Fig. 7. Inclusion problem: (a) mean $\bar{h}$ and (b) standard deviation σ_h corresponding to several levels of spatial correlation of log conductivity.

$$\hat{\bar{h}} = q_0\mathcal{H}(\alpha - x)\left(\frac{\alpha - x}{K_{h_1}} + \frac{L - \alpha}{K_{h_2}}\right) + q_0\mathcal{H}(x - \alpha)\frac{L - x}{K_{h_2}}, \tag{25}$$

and the conditional hydraulic head variance $\hat{\sigma}^2_{\bar{h}}(x) = \widehat{\bar{h}^2} - \hat{\bar{h}}^2$, with

$$\widehat{\bar{h}^2} = 2q_0^2\mathcal{H}(\alpha - x)\left[\frac{1}{K_{h_1}^2}\int_0^{\alpha - x}(\alpha - x - s)e^{\sigma^2_{Y_1}\rho_{Y_1}(s)}ds + \frac{(\alpha - x)(L - \alpha)}{K_{h_1}K_{h_2}} + \frac{1}{K_{h_2}^2}\int_0^{L-\alpha}(L - \alpha - s)e^{\sigma^2_{Y_1}\rho_{Y_1}(s)}ds\right]$$
$$+ 2q_0^2\frac{\mathcal{H}(x - \alpha)}{K_{h_2}^2}\int_0^{L-x}(L - x - s)e^{\sigma^2_{Y_1}\rho_{Y_1}(s)}ds. \tag{26}$$

Here $K_{h_i} \equiv K_{g_i}\exp(-\sigma^2_{Y_i}/2)$ and $K_{g_i} \equiv \exp(\overline{Y}_i)$ are the harmonic and geometric means of conductivities $K_i(x)$ ($i = 1, 2$), respectively. It is worthwhile pointing out that the analytical solutions (25) and (26) are exact, while their analytical counterparts in [32] are first-order (in $\sigma^2_{Y_i}$) approximations.

Averaging (25) and (26) over the random α yields the mean hydraulic head:

$$\bar{h} = \int_0^L \widehat{\bar{h}}(\alpha) p_\alpha(\alpha) \mathrm{d}\alpha \approx \sum_{j=1}^{N_\alpha} \vartheta_j \widehat{\bar{h}}(\alpha_j), \tag{27}$$

and the hydraulic head variance $\sigma_h^2(x) = \overline{h^2} - \bar{h}^2$, with

$$\overline{h^2} = \int_0^L \widehat{\overline{h^2}}(\alpha) p_\alpha(\alpha) \mathrm{d}\alpha \approx \sum_{j=1}^{N_\alpha} \vartheta_j \widehat{\overline{h^2}}(\alpha_j). \tag{28}$$

The integrals in (27) and (28) are evaluated numerically with an N_α-point quadrature, wherein $\{\alpha_j, \vartheta_j\}_{j=1}^{N_\alpha}$ is a set of quadrature points and the corresponding weights. Alternatively, one can compute these integrals analytically. For example, if α is a truncated Gaussian variable on $[0,L]$, then regardless of the functional form of the correlation functions of Y_i ($i = 1, 2$) the ensemble mean of hydraulic head is given by

$$\bar{h} = \frac{\sqrt{2} q_0 \sigma_\alpha}{\sqrt{\pi}\mathcal{W}} \left(\frac{1}{K_{h_1}} - \frac{1}{K_{h_2}}\right) \mathrm{e}^{-u^2 - \mathrm{e}^{-u_1^2}} - q_0 \frac{\bar{\alpha} - x}{\mathcal{W}} \left(\frac{1}{K_{h_1}} - \frac{1}{K_{h_2}}\right) \mathrm{erf}(u) + \frac{q_0}{\mathcal{W}} \left(\frac{\bar{\alpha} - x}{K_{h_1}} - \frac{L - \bar{\alpha}}{K_{h_2}}\right) \mathrm{erf}(u_1) - \frac{q_0}{\mathcal{W}} \frac{L - x}{K_{h_2}} \mathrm{erf}(u_0), \tag{29}$$

where

$$u = \frac{x - \bar{\alpha}}{\sqrt{2}\sigma_\alpha}, \quad u_0 = -\frac{\bar{\alpha}}{\sqrt{2}\sigma_\alpha}, \quad u_1 = \frac{L - \bar{\alpha}}{\sqrt{2}\sigma_\alpha}, \quad \mathcal{W} = \mathrm{erf}(u_1) - \mathrm{erf}(u_0). \tag{30}$$

A closed-form analytical expression for $\sigma_h^2(x)$ is obtained by integrating (28). Likewise, one can derive analytical expressions for $\bar{h}$ and σ_h^2 if α is a uniformly-distributed random variable. These analytical solutions are used to verify the RDD–PCM approach.

Figs. 2 and 3 demonstrate that the ensemble mean $\bar{h}(x)$ and the standard deviation $\sigma_h(x)$ obtained with the RDD–PCM approach coincide with their analytical counterparts, which verifies the accuracy of the RDD–PCM algorithm. In these simulations, we set the correlation lengths of random fields $Y_1(x,\omega)$ and $Y_2(x,\omega)$ to $l_{Y_1} = 5.0$ and $l_{Y_2} = 1.0$, respectively. This necessitated the use of *4* collocation points to represent realizations of each $Y_i(x)$ ($i = 1, 2$) and 8 collocation points for realizations of the interface α. Hence the total number of collocation points for the whole random composite is $4 \times 4 \times 8 = 128$.

Fig. 2 also reveals that the mean and standard deviation of h are practically unaffected by the choice of a statistical model for the contact point α (the curves corresponding to the Gaussian, $\alpha \sim N\{0.5L, (0.03L)^2\}$, and uniform, $\alpha \sim U\{(0.5 - 0.03\sqrt{3})L, (0.5 + 0.03\sqrt{3})L\}$, distributions of α nearly overlap), as long as $\bar{\alpha}$ and σ_α remain the same. This finding is important for uncertainty quantification, since available data typically allow one to provide the "best" prediction of the interface location as $\bar{\alpha}$ and to quantify the predictive uncertainty in terms of σ_α (e.g. [36]), while estimation of the full distributions of α is more elusive and subjective [35].

The effects of uncertainty in the position of the contact point α on the statistical moments of the system state h are explored in Fig. 3. As can be expected, larger parametric uncertainty (σ_α) translates into larger predictive uncertainty (σ_h), although this effect is clearly nonlinear. Geometric uncertainty (as quantified by σ_α) impacts the predictive uncertainty (i.e. σ_h) much more strongly than it does the predictions themselves (i.e. $\bar{h}$). This finding is in line with the observations reported in [33]. Fig. 3 also demonstrates that uncertainty in the position of the contact point α smoothes the predicted (mean) head $\bar{h}(x)$, whose spatial derivative would have a jump discontinuity at $x = \alpha$ if α were deterministic (i.e. known with certainty). While not shown in this figure, we found that for $\sigma_\alpha = 0.06L$ and $0.18L$ the choice between the Gaussian and uniform distributions of α has the same negligible effect on $\bar{h}$ and σ_h as that exhibited in Fig. 2 for $\sigma_\alpha = 0.03L$.

Next, we explore how the ensemble mean $\bar{h}(x)$ and the standard deviation $\sigma_h(x)$ depend on the correlation lengths of log conductivities $Y_i(x;\omega)$. The random (Gaussian or uniformly distributed) contact point $\alpha(\omega)$ was assigned the mean $\bar{\alpha} = L/2$ and the standard deviation $\sigma_\alpha = 0.03L$; it was represented with 8 collocation points. As discussed in Section 3, the correlation lengths l_{Y_1} and l_{Y_2} determine the number of points that are required to represent the random fields $Y_1(x;\omega)$ and $Y_2(x;\omega)$ with the degree of accuracy prescribed by (8). This number also varies from one realization of α, as reflected by the ratios l_{Y_1}/α and $l_{Y_2}/(L - \alpha)$ for $Y_1(x;\omega)$ and $Y_2(x;\omega)$, respectively. For $\alpha = \bar{\alpha}$, it took 4, 7, 25, and 52 collocation points to represent $Y_1(x;\omega)$ with the correlation lengths $l_{Y_1} = \infty$, 5.0, 1.0, and 0.5, respectively; and 4, 25, 52, and 87 collocation points to represent $Y_2(x;\omega)$ with the correlation lengths $l_{Y_2} = \infty$, 1.0, 0.5, and 0.1, respectively. The total number of collocation points used is 128 for the $l_{Y_1} = l_{Y_2} = \infty$ case, 1400 for the $l_{Y_1} = 5$ and $l_{Y_2} = 1$ case, 10400 for the $l_{Y_1} = 1$ and $l_{Y_2} = 0.5$ case, and 36192 for the $l_{Y_1} = 0.5$ and $l_{Y_2} = 0.1$ case. These numbers differ somewhat for other realizations of α and are mentioned here to provide some indication of the computational effort involved.

The results of these simulations are presented in Fig. 4. Predictions of the system state h (i.e. $\bar{h}$) are independent of the correlation structure of $Y(x;\omega)$, which is in full agreement with the analytical solution (29). However, the correlation of Y does affect the predictive uncertainty (i.e. σ_h): smaller correlation lengths lead to higher predictive uncertainty (larger values of σ_h). While not shown in this figure, we found the same level of agreement between the analytical and RDD–PCM solutions and between the solutions corresponding to the Gaussian and uniform distributions of α as that exhibited in Fig. 2.

6.2. Random inclusion

Consider a one-dimensional domain $\mathcal{D} = \{x : 0 \leqslant x \leqslant L\}$ consisting of a random inclusion $\mathcal{D}_1 = \{x : \alpha - L/6 \leqslant x \leqslant \alpha + L/6\}$ embedded into a random material $\mathcal{D}_2 = \mathcal{D}/\mathcal{D}_1$ (Fig. 1(b)). The one-dimensional flow Eq. (1) with $f = 0$ is subject to the boundary conditions (21) and the continuity conditions (18) at the random interfaces $x = \alpha \pm L/6$.

For the uniformly distributed α, the upper limit of σ_α is now defined from the condition $0 < \alpha = L(0.5 + \sigma_\alpha \xi_\alpha) \pm L/6) < L$, where ξ_α is the zero-mean uniformly-distributed random variable introduced in Section 6.1. This condition yields:

$$\sigma_\alpha < \frac{\sqrt{3}}{9}. \tag{31}$$

Both distributions have the same mean, $0.5L$, and variance, $(0.03L)^2$.

To further validate the RDD–PCM approach, we compare its estimates of the ensemble mean $\bar{h}(x)$ and the standard deviation $\sigma_h(x)$ with those obtained with Monte Carlo simulations consisting of 10000 realizations. We set $\sigma_\alpha = 0.03L$, and the correlation lengths of random fields $Y_1(x,\omega)$ and $Y_2(x,\omega)$ to $l_{Y_1} = 5.0$ and $l_{Y_2} = 1.0$, respectively. As discussed in Section 6.1, this requires the RDD–PCM approach to employ 4 collocation points to represent realizations of $Y_i(x;\omega)$ for each subdomain, and 8 collocation points to represent realizations of $\alpha(\omega)$. The total number of collocation points is 128. Fig. 5 demonstrates that the two numerical approaches yield identical solutions for $\bar{h}$ and σ_h. Both statistics are relatively insensitive to the choice of the PDF of $\alpha(\omega)$—Gaussian or uniform—as long as they have the same mean and variance.

Fig. 6 exhibits the impact of uncertainty in the inclusion's location on the predictive uncertainty of h. In analogy with the contact-point problem, higher geometric uncertainty (σ_α) leads to higher predictive uncertainty (σ_h), and increasing values of σ_α have a much more pronounced effect on predictive uncertainty (σ_h) than on predictions ($\bar{h}$). While the latter are limited

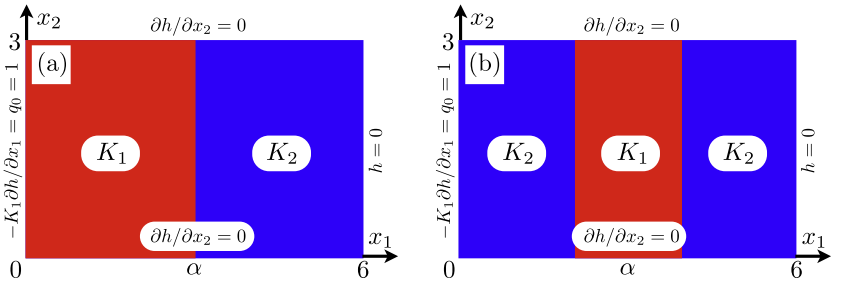

Fig. 8. Two-dimensional composites in which (a) two randomly heterogeneous materials are separated by a random contact line $x_1 = \alpha$ and (b) a randomly heterogeneous inclusion centered around random $x_1 = \alpha$ is embedded into another randomly heterogeneous material. Flow is perpendicular to stratification.

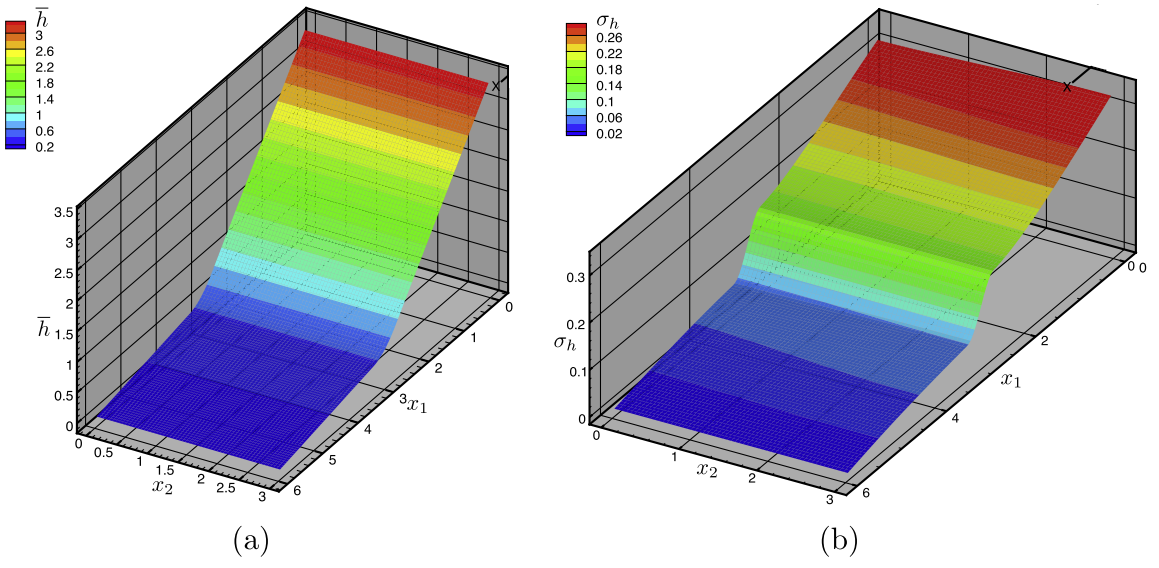

Fig. 9. Two-dimensional flow perpendicular to the contact line: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach.

to the immediate vicinity of the interfaces $\alpha \pm L/6$, the former manifest themselves throughout much of the flow domain, peaking at its center. In both problems, uncertainty in the interface position and the conductivities of the sub-domains smoothes the head predictions $\bar{h}(x)$, which otherwise would have kinks at the interfaces.

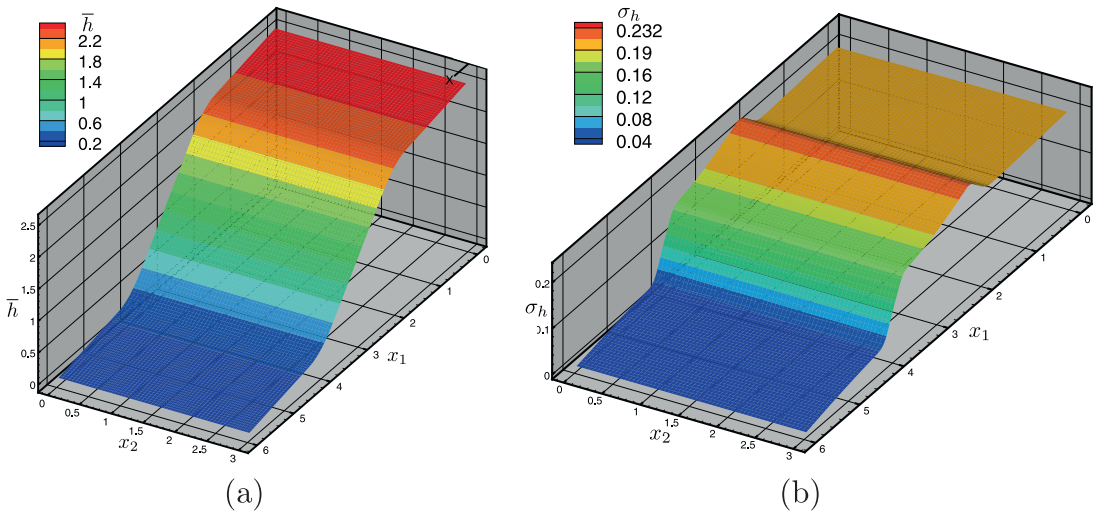

Fig. 10. Two-dimensional flow perpendicular to the inclusion: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach.

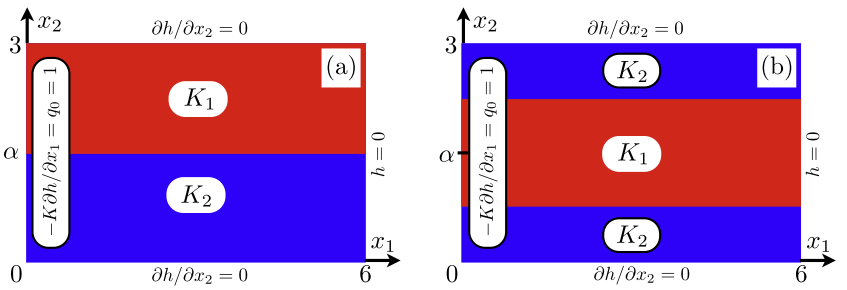

Fig. 11. Two-dimensional composites in which (a) two random materials are separated by a random contact line $x_1 = \alpha$ and (b) a random inclusion centered around random $x_1 = \alpha$ is embedded into another random material. Flow is parallel to stratification.

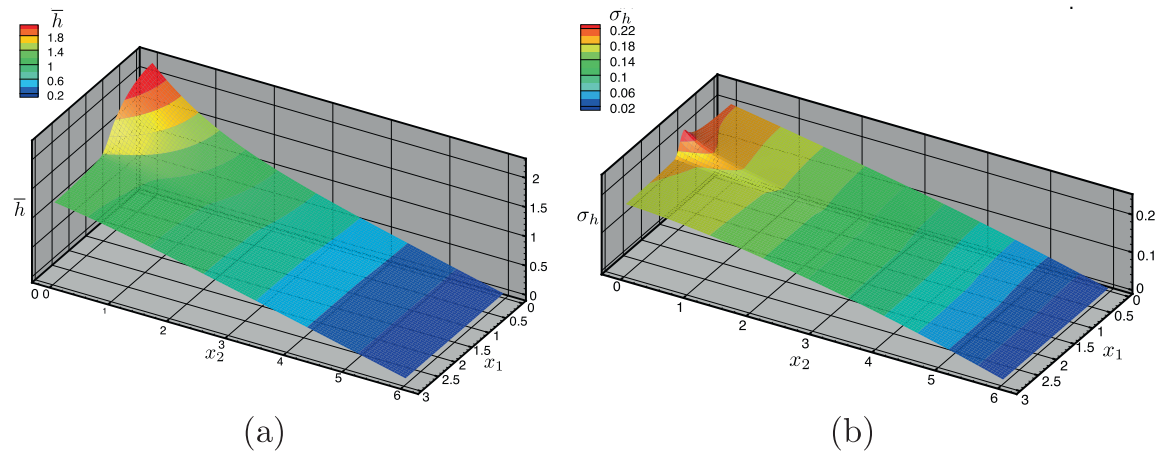

Fig. 12. Two-dimensional flow parallel to the contact line: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach.

Fig. 7 demonstrates the effects of various degrees of the spatial correlation, l_{Y_i}, of log conductivities $Y_i(x;\omega)$ on the statistical moments of h for (Gaussian and uniform) α with $\sigma_\alpha = 0.03L$. Similar to the contact-point problem, estimates of hydraulic head ($\bar{h}$) are independent from the correlation lengths l_{Y_i}, while their impact on predictive uncertainty (σ_h) is significant.

7. Multi-Dimensional computational examples

We proceed to demonstrate the ability of the RDD–PCM approach to handle two- and three-dimensional problems. Two-dimensional domains with the contact line and inclusion and the corresponding boundary conditions for the two-dimensional flow equation are shown in Fig. 8.

The ensemble mean $\bar{h}$ and the standard deviation σ_h, computed with the PCM-RDD approach, are presented in Figs. 9 and 10 for the contact line (Fig. 8(a)) and inclusion (Fig. 8(b)) problems, respectively. In these simulations, $L = 6$, $\alpha(\omega)$ is modeled as a uniform random variable $U\{(0.5 - 0.03\sqrt{3})L, (0.5 + 0.03\sqrt{3})L\}$, $Y_1(\mathbf{x};\omega)$ as a multivariate-Gaussian random field $N\{0,0.01\}$ with correlation length $l_{Y_1} = 5$, and $Y_2(\mathbf{x};\omega)$ as a multivariate-Gaussian random field $N\{2,0.04\}$ with correlation

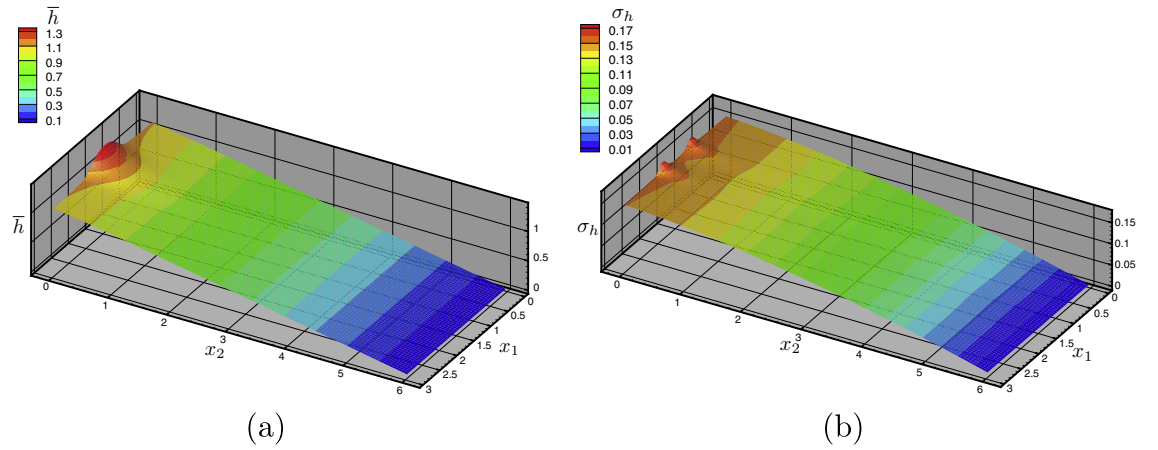

Fig. 13. Two-dimensional flow parallel to the inclusion: (a) mean $\bar{h}$ and (b) standard deviation σ_h computed with the RDD–PCM approach.

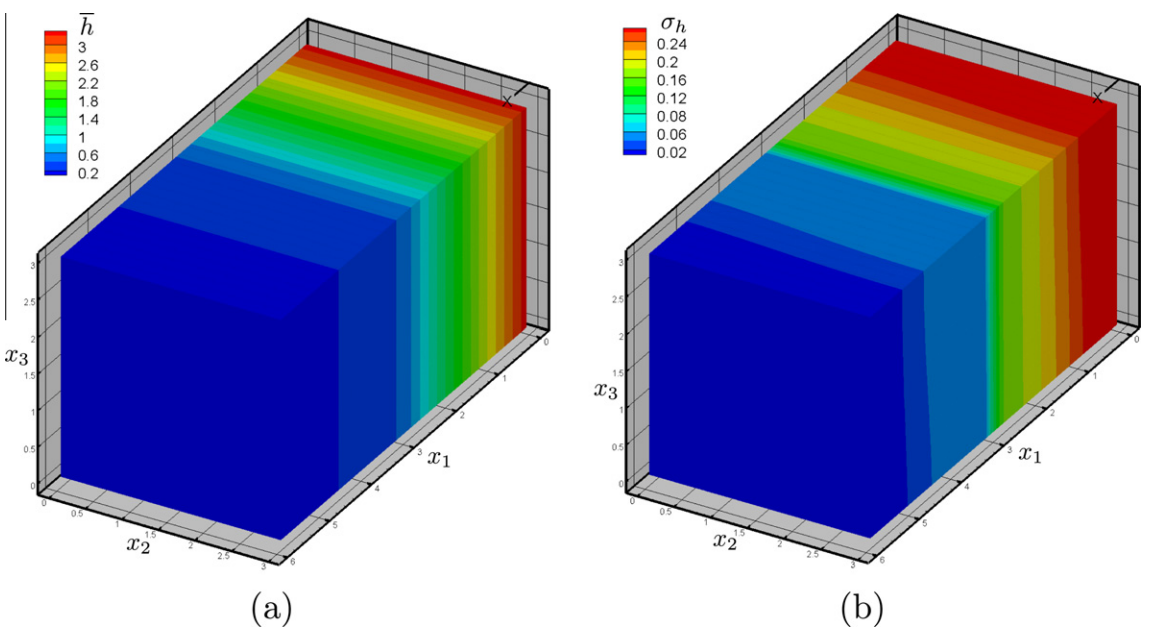

Fig. 14. Three-dimensional flow perpendicular to the random contact surface $x_1 = \alpha$: (a) mean $\bar{h}$ and (b) standard deviation $\sigma(h)$ computed with the RDD–PCM approach.

length $l_{Y_2} = 1$. We used 25 and 52 collocation points to represent realizations of Y_1 and Y_2, respectively; and 8 collocation points to represent α. This yields the total of 10400 collocation points.

Next, we consider two-dimensional flow parallel to stratification (Fig. 11). The mean and standard deviation of $h(\mathbf{x};\omega)$ for these contact line (Fig. 11(a)) and inclusion (Fig. 11(b)) problems are shown in Fig. 12 and 13, respectively. In these simulations, $\alpha(\omega)$ is modeled as a uniformly-distributed random variable, $Y_1(\mathbf{x};\omega)$ as a multivariate-Gaussian field $N\{0,0.01\}$ with correlation length $l_{Y_1} = 5$, and $Y_2(\mathbf{x};\omega)$ as a multivariate-Gaussian field $N\{2,0.04\}$ with correlation length $l_{Y_2} = 1$. We used 25 and 52 collocation points to represent realizations of Y_1 and Y_2, respectively; and 8 collocation points to represent realizations of α. The total number of collocation points is 10400.

Finally, the mean and standard deviation of $h(\mathbf{x};\omega)$ for flow in a three-dimensional random composite consisting of two materials separated by the plane $x_1 = \alpha$ are shown in Fig. 14. In this example, $\alpha(\omega)$ is modeled as a uniform random variable $U\{(0.5-0.03\sqrt{3})L,(0.5+0.03\sqrt{3})L\}$, $Y_1(x;\omega)$ as a multivariate-Gaussian field $N\{0,0.01\}$ with correlation length $l_{Y_1} = 5$, and $Y_2(\mathbf{x};\omega)$ as a multivariate-Gaussian field $N\{2,0.04\}$ with correlation length $l_{Y_2} = 1$. We used 52 and 87 collocation points to represent realizations of Y_1 and Y_2, respectively; and 8 collocation points to represent realizations of α. The total number of collocation points is 36192.

8. Summary

Probabilistic collocation methods (PCMs) on sparse grids is an efficient numerical method to quantify uncertainty in complex heterogeneous phenomena. However, for highly heterogeneous systems, e.g. random composites, probabilistic representations of uncertainty may become highly non-Gaussian, multimodal, and nonstationary (statistically inhomogeneous), exhibiting small spatially-varying correlation lengths (large number of random dimensions). To expand the capability of PCMs, we have introduced a new robust and efficient framework, which combines random domain decompositions (RDDs) and PCMs. The combined RDD–PCM approach, implemented on sparse grids, is able to handle problems with multimodal distribution functions and small correlation lengths (large number of random dimensions).

A new set of analytical ensemble mean and standard deviation solutions have been derived to verify the RDD–PCM approach for a one-dimensional contact-point problem. Monte Carlo simulations (MCS) were used to verify the accuracy of the RDD–PCM for a one-dimensional random inclusion problem. These comparisons demonstrate that the RDD–PCM approach provides robust and efficient approximations for the ensemble statistics of solutions of steady-state diffusion problems in random composites.

We systematically analyzed the effects of correlation lengths of a system parameter Y (log-hydraulic conductivity) on the ensemble mean and standard deviation of a system state h (hydraulic head). We also investigated the impact of different probabilistic distributions used to quantify predictive uncertainty associated with geometric uncertainty (spatial distributions of constitutive materials in a composite).

To demonstrate the robustness of the RDD–PCM approach, it was used to simulate flow in two- and three-dimensional random composites. Different from the *intrusive* Galerkin projection based polynomial-chaos approach [39], the RDD–PCM approach provides an efficient *non-intrusive* tool for uncertainty quantification, which is used in combination with existing deterministic legacy codes.

The fast convergence of the sparse grid PCM approach also relies on smoothness of the solution in the random space with the assumption of equal importance in each direction. Ultimately, adaptivity in random space is necessary for regions with less regularity or directions with more importance. A combined RDD–PCM approach with adaptive sparse grids (similar to the approaches in [53–56]) can greatly improve the efficiency of the method.

Acknowledgments

This work was supported by Applied Mathematics program of the US DOE Office of Advanced Scientific Computing Research. Simulations were performed using the computational resources of the Environmental Molecular Sciences Laboratory. The Pacific Northwest National Laboratory is operated by Battelle for the US Department of Energy under Contract DE-AC05-76RL01830.

Appendix A. Selection of collocation points

Two different methods for the selection of collocation point sets are discussed here: tensor products of one-dimensional collocation point sets (Section A.1) and a sparse grid strategy for high dimensionality (Section A.2). In the present analysis, we employ the sparse grid strategy; tensor products of one-dimensional collocation point sets are discussed here only for the purpose of completeness.

A.1. Tensor products of one-dimensional collocation point sets

Consider the ensemble mean $\mathbb{E}(g)$ of a function $g(\boldsymbol{\xi})$ that is smooth with respect to all of its random arguments $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_M)^T$. Recalling (9), it is given by

$$\mathbb{E}(g) = \int_{\Gamma} g(\xi_1, \ldots, \xi_M) p_1(\xi_1) p_2(\xi_2) \cdots p_M(\xi_M) \mathrm{d}V_\xi. \tag{A.1}$$

The M-dimensional integral in (A.1) is represented by the tensor product:

$$\mathbb{E}(g) = (\mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_M)(g), \tag{A.2}$$

of one-dimensional integrals:

$$\mathcal{U}_j(g) = \int_{\Gamma_j} g(\xi) p_j(\xi_j) \mathrm{d}\xi_j, \quad j = 1, \ldots, M. \tag{A.3}$$

Let $i_j \in \mathbb{N}$ specify "the degree of the integration" in the j-th random dimension, i.e. n_{i_j} be the number of quadrature points used to approximate the integral in (A.3). Then $\mathcal{U}_j^{i_j}(g)$ is an approximation of $\mathcal{U}_j(g)$ given by

$$\mathcal{U}_j(g) \approx \mathcal{U}_j^{i_j}(g) = \sum_{k_j=1}^{n_{i_j}} \vartheta_{k_j}^{i_j} g\left(\xi_{k_j}^{i_j}\right), \tag{A.4}$$

where $k_j \in [1, n_{i_j}]$, and $\xi_{k_j}^{i_j}$ is the k_j-th quadrature point in the j-th random dimension, with a weight $\vartheta_{k_j}^{i_j}$. Substitution of (A.4) into (A.2) yields:

$$\mathbb{E}(g) \approx \left(\mathcal{U}_1^{i_1} \otimes \cdots \otimes \mathcal{U}_M^{i_M}\right)(g) = \sum_{k_1=1}^{n_{i_1}} \cdots \sum_{k_M=1}^{n_{i_M}} \left(\vartheta_{k_1}^{i_1} \otimes \cdots \otimes \vartheta_{k_M}^{i_M}\right) g\left(\xi_{k_1}^{i_1}, \ldots, \xi_{k_M}^{i_M}\right). \tag{A.5}$$

The total number of collocation (quadrature) points in (A.5) is $Z_M = n_{i_1} \cdots n_{i_M}$. If the same number (say, $k+1$) of collocation points is used in all random dimensions, $n_{i_1} = \cdots = n_{i_M} = k+1$, then the total number of collocation points is $Z_M^k = (k+1)^M$. For a small number of random dimensions, e.g. $M \leqslant 4$, the tensor product of one-dimensional collocation point sets is a good choice for collocation point sets. However, since Z_M^k grows exponentially with M, the efficiency of a quadrature using the tensor product of one-dimensional collocation point sets decreases exponentially with the number of random dimensions M.

A.2. High dimensionality and sparse grids

The number of random dimensions needed to avoid erroneous oscillations for both mean and variance estimates increases fast as the correlation length decreases. In such high-dimensional cases, the reliance on tensor products of one-dimensional collocation points leads to a prohibitively large number of collocation points. Following [13], we use the Smolyak formula [30] to obtain a point set that has a significantly smaller number of points than the tensor product set does. Resulting sparse grids do not depend as strongly on the dimensionality of the random space, which makes them more suitable for applications with high-dimensional random inputs.

The one-dimensional quadrature formula (A.4) serves as a building block for the Smolyak formula [30]. The latter approximates the M-dimensional integral in (A.1)–(A.3) with a linear combination of tensor products as

$$\mathbb{E}(g) \approx A_{q,M}(g) \equiv \sum_{q-M+1 \leqslant |i| \leqslant q} (-1)^{q-|i|} \binom{M-1}{q-|i|} \left(\mathcal{U}_1^{i_1} \otimes \cdots \otimes \mathcal{U}_M^{i_M}\right)(g). \tag{A.6}$$

Here the sparseness parameter $q \geqslant M$ determines the order of the formula, $k = q - M$ is called the "level" of the Smolyak formula, and $|i| = i_1 + i_2 + \cdots + i_M$. Next, we set $n_{i_j} (j = 1, \ldots, M)$, the number of quadrature (collocation) points in (A.4), to $n_1 = 1$ and $n_k = 2^{k-1} + 1$ for $k > 1$. Then, to compute $A_{q,M}(g)$, we only need to evaluate g on the Smolyak "sparse grid", i.e. on the collocation point set Θ_M of size M:

$$\Theta_M \equiv \cup_{q-M+1 \leqslant |i| \leqslant q} \left(\Theta_1^{i_1} \times \cdots \times \Theta_1^{i_M}\right), \tag{A.7}$$

where $\Theta_1^{i_j}$ represents a one-dimensional collocation point set in the j-th random dimension. The grid point sets obtained with both the Smolyak formula and the tensor products of one-dimensional collocation point sets are shown in Fig. A.1 for $M = 2$ random dimensions. The cross and circle nodes in Fig. A.1(a) correspond to sparse grids of levels $k = 5$ and $k = 6$, respectively. In Fig. A.1(b), the cross and circle nodes correspond to the full tensor product based on the Gauss quadrature rule with levels 10 and 11, respectively.

A convergence criterion in terms of the sparse grid level k is to require the absolute value of the difference of $A_{q,M}(g)$ obtained with two successive sparse grid levels be smaller than a specified threshold. See an extensive review of the adaptivity of sparse grids for details in [53].

Below we discuss two main categories of sparse grids: nested and non-nested sparse grids corresponding to different random distributions. In particular, for a uniform distribution, sparse collocation point sets are generated by the nested Chebyshev quadrature points using the Clenshaw–Curtis formulae. For beta and Gaussian distributions, sparse collocation point sets are generated by Gaussian-quadrature points. The 2^i degree of exactness is achieved by choosing $M = 2^{i-1} + 1$ quadrature points.

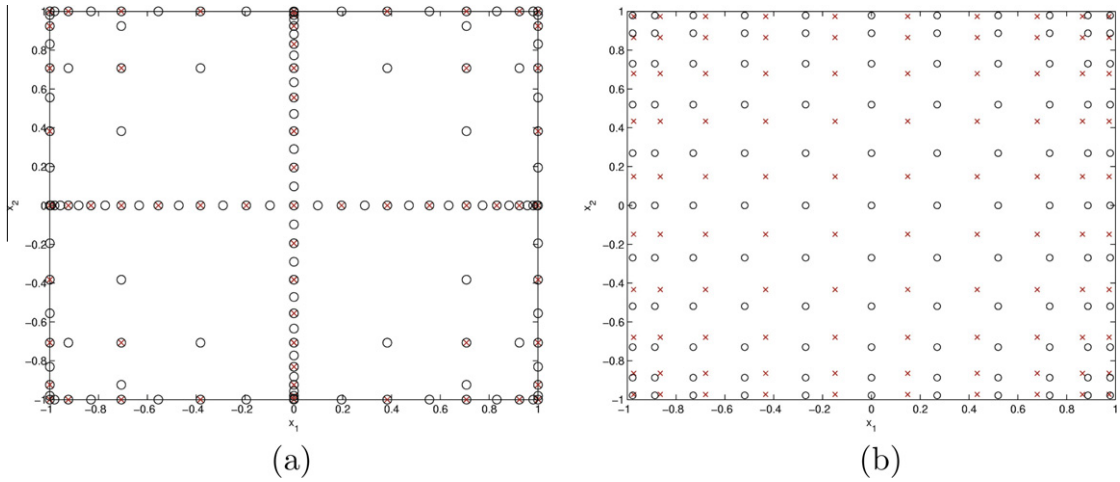

Fig. A.1. Grid points for $M = 2$ random dimensions obtained with (a) the Smolyak formula (the cross and circle nodes correspond to sparse grids of levels $k = 5$ and $k = 6$, respectively) and (b) the full tensor product (the cross and circle nodes correspond to the Gauss quadrature rule of levels 10 and 11, respectively).

A.3. Nested sparse grids

A set of sparse grid points in the j-th random dimension, $\Theta_1^{i_j}$, is called nested if $\Theta_1^{i_j} \in \Theta_1^{i_j+1}$. As in Section A.2, we set n_{i_j} $(j = 1, \ldots, M)$, the number of quadrature (collocation) points in (A.4), to $n_1 = 1$ and $n_k = 2^{k-1} + 1$ for $k > 1$, which implies that the same quadrature rule is used in all j random dimensions so that the subscript j is dropped. Among many alternatives, we take the quadrature points ξ_k^i in (A.4) to be the Clenshaw–Curtis points. These are defined as the extreme points of the Chebyshev polynomials [57]:

$$\xi_1^1 = 0, \quad \xi_k^i = -\cos\left[\frac{\pi(k-1)}{n_i - 1}\right], \quad k = 2, \ldots, n_i. \tag{A.8}$$

The corresponding weights in (A.4) are given by

$$\vartheta_1^i = \vartheta_{n_i}^i = \frac{1}{n_i(n_i - 2)}, \quad i = 2, \ldots, n_i - 1, \tag{A.9a}$$

$$\vartheta_k^i = \frac{2}{n_i - 1} + \frac{4}{n_i - 1} \sum_{k=1}^{(n_i-1)/2}{}' \frac{1}{1 - 4k^2} \cos\left[\frac{2\pi k(i-1)}{n_i - 1}\right], \tag{A.9b}$$

where the symbol $\sum'$ indicates that the last term of the sum is halved.

The one-dimensional quadrature formula (A.4) with the Clenshaw–Curtis points (A.8)–(A.9) integrates exactly all polynomials of degree less than n_i. For the Smolyak Clenshaw–Curtis grid, the accuracy of the Smolyak formula is estimated as follows. Let $\sigma = \text{floor}\,(q/M)$ and let $\tau = q \bmod (M)$. Then $A_{q,M}$ has the degree of exactness [58]:

$$m(q, M) = \begin{cases} 2(q - M) + 1, & \text{if } q \leqslant 4M, \\ 2^{\sigma-1}(M + 1 + \tau) + 2M - 1, & \text{otherwise.} \end{cases} \tag{A.10}$$

Additionally, the nested Clenshaw–Curtis points can greatly reduce the total number of collocation points.

A.4. Non-nested sparse grids

An alternative way for generating sparse nodal sets is to use one-dimensional Gaussian-quadrature points in (A.4) as a building block and then employ the Smolyak formula [30] to approximate the M-dimensional integral in (A.1)–(A.3) with a linear combination of tensor products. In this paper, the random field of Y has normal distribution. Thus, non-nested Gaussian sparse grids are used to represent the random field of Y. The 2^i degree of exactness is achieved by choosing $2^{i-1} + 1$ Gaussian-quadrature points. The nested property may be lost, which makes such nodal sets larger than their nested counterparts. However, for random inputs with arbitrary PDFs, a sparse quadrature rule is constructed in a way that the weight function coincides with the PDF. Gaussian-quadrature nodal sets for arbitrary weight functions is obtained in an efficient way [59].

For a Smolyak Gaussian grid, the accuracy of the Smolyak formula is obtained as follows. Let $\sigma = \text{floor}\,(q/M)$ and let $\tau = q \bmod M$. Then $A_{q,M}$ has the degree of exactness [58]:

$$m(q,M) = \begin{cases} 2(q-M)+1, & \text{if } q \leqslant 3M, \\ 2^{\sigma-1}(M+1+\tau)-1, & \text{otherwise.} \end{cases} \tag{A.11}$$

References

[1] G. Dagan, Flow and Transport in Porous Formations, Springer, 1989.
[2] G. Dagan, S.P. Neuman (Eds.), Subsurface Flow and Transport: A Stochastic Approach, Cambridge University Press, New York, 1997.
[3] D.M. Tartakovsky, A. Guadagnini, M. Riva, Stochastic averaging of nonlinear flows in heterogeneous porous media, J. Fluid Mech. 492 (2003) 47–62.
[4] A.M. Tartakovsky, S.P. Neuman, R. Lenhard, Immiscible front evolution in randomly heterogeneous porous media, Phys. Fluids 15 (11) (2003) 3331–3341.
[5] D.M. Tartakovsky, Z. Lu, A. Guadagnini, A.M. Tartakovsky, Unsaturated flow in heterogeneous soils with spatially distributed uncertain hydraulic parameters, J. Hydrol. 275 (2003) 182–193.
[6] D.M. Tartakovsky, A. Guadanini, F. Ballio, A.M. Tartakovsky, Localization of mean flow and apparent transmissivity tensor for bounded randomly heterogeneous aquifers, Transport Porous Med. 49 (1) (2002) 41–58.
[7] A.M. Tartakovsky, P. Meakin, H. Huang, Stochastic analysis of immiscible displacement of the fluids with arbitrarily viscosities and its dependence on support scale of hydrological data, Adv. Water Resour. 27 (2004) 1151–1166.
[8] A.M. Tartakovsky, L. Garcia-Naranjo, D.M. Tartakovsky, Transient flow in a heterogeneous vadose zone with uncertain parameters, Vadose Zone J. 3 (1) (2004) 154–163.
[9] K.D. Jarman, A. Tartakovsky, Divergence of solutions to solute transport moment equations, Geophys. Res. Lett. 35 (2008) L15401, doi:10.1029/2008GL034495.
[10] N. Wiener, The homogeneous chaos, Am. J. Math. 60 (1938) 897–936.
[11] R.G. Ghanem, P. Spanos, Stochastic Finite Eelements: A Spectral Approach, Springer-Verlag, New York, 1991.
[12] D. Xiu, G.E. Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[13] D. Xiu, J.S. Hesthaven, High order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[14] B. Ganapathysubramanian, N. Zabaras, Sparse grid collocation schemes for stochastic natural convection problems, J. Comput. Phys. 225 (1) (2007) 652–685.
[15] I. Babuska, F. Nobile, R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data, SIAM J. Numer. Anal. 45 (3) (2007) 1005–1034.
[16] M.D. Webster, A.P. Sokolov, A methodology for quantifying uncertainty in climate projections, Climatic Change 46 (4) (2000) 417–446.
[17] B.D. Phenix, J.L. Dinaro, M.A. Tatang, J.W. Tester, J.B. Howard, G.J. McRae, Incorporation of parametric uncertainty into complex kinetic mechanisms: application to hydrogen oxidation in supercritical water, Combust. Flame 112 (1-2) (1998) 132–146.
[18] O.P.L. Maitre, M.T. Reagan, H.N. Najm, R.G. Ghanem, O.M. Knio, A stochastic projection method for fluid flow: II. Random process, J. Comput. Phys. 181 (1) (2002) 9–44.
[19] B.J. Debusschere, H.N. Najm, P.P. Pebay, O.M. Knio, R.G. Ghanem, O.P.L. Maitre, Numerical challenges in the use of polynomial chaos representations for stochastic processes, SIAM J. Sci. Comput. 26 (2) (2004) 698–719.
[20] M.A. Tatang, R.G. Prinn, G.J. McRae, An efficient method for parametric uncertainty analysis of numerical geophysical models, J. Geophys. Res. 102 (D18) (1997) 21925–21932.
[21] S. Isukapalli, A. Roy, P.G. Georgopoulos, Stochastic response surface methods (SRSM) for uncertainty characterization and propagation: application to environmental and biological systems, Risk Anal. 18 (1998) 351–363.
[22] F. Nobile, R. Tempone, C.G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2309–2345.
[23] F. Nobile, R. Tempone, C.G. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (5) (2008) 2411–2442.
[24] D.M. Ghiocel, R.G. Ghanem, Stochastic finite-element analysis of seismic soil-structure interaction, J. Eng. Mech. 128 (1) (2002) 66–77.
[25] M.T. Reagan, H.N. Njam, R.G. Ghanem, O.M. Knio, Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection, Combust. Flame 132 (2003) 545–555.
[26] G. Lin, A.M. Tartakovsky, An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow in random porous media, Adv. Water Resour. 32 (5) (2009) 712–722.
[27] G. Lin, A.M. Tartakovsky, Numerical studies of three-dimensional stochastic darcy's equation and stochastic advection–diffusion–dispersion equation, J. Sci. Comput. 43 (1) (2010) 92–117.
[28] C. Bernardi, Y. Maday, Properties of some weighted Sobolev spaces and application to spectral approximations, SIAM J. Numer. Anal. 26 (1989) 769–829.
[29] C. Canuto, A. Quarteroni, Approximation results for orthogonal polynomials in Sobolev spaces, Math. Comput. 38 (1982) 67–86.
[30] S. Smolyak, Quadrature and interpolation formulas for tensor products of certain classes of functions, Soviet Math. Dokl. 4 (1963) 240–243.
[31] C.L. Winter, D.M. Tartakovsky, Mean flow in composite porous media, Geophys. Res. Lett. 27 (12) (2000) 1759–1762.
[32] C.L. Winter, D.M. Tartakovsky, Groundwater flow in heterogeneous composite aquifers, Water Resour. Res. 38 (8) (2002), doi:10.1029/2001WR000450.
[33] C.L. Winter, D.M. Tartakovsky, A. Guadagnini, Numerical solutions of moment equations for flow in heterogeneous composite aquifers, Water Resour. Res. 38 (5) (2002), doi:10.1029/2001WR000222.
[34] A.M. Yaglom, Correlation Theory of Stationary and Related Random Functions. I. Basic Results, Springer-Verlag, New York, 1987.
[35] L. Guadagnini, A. Guadagnini, D.M. Tartakovsky, Probabilistic reconstruction of geologic facies, J. Hydrol. 294 (2004) 57–67.
[36] B.E. Wohlberg, D.M. Tartakovsky, A. Guadagnini, Subsurface characterization with support vector machines, IEEE Trans. Geosci. Remote Sens. 44 (1) (2006) 47–57.
[37] D.M. Tartakovsky, B.E. Wohlberg, A. Guadagnini, Nearest neighbor classification for facies delineation, Water Resour. Res. 34 (2007) L05404, doi:10.1029/2007GL029245.
[38] B.E. Wohlberg, D.M. Tartakovsky, Delineation of geological facies from poorly differentiated data, Adv. Water Resour. 32 (2) (2009) 225–230.
[39] D. Xiu, D.M. Tartakovsky, A two-scale nonperturbative approach to uncertainty analysis of diffusion in random composites, Multiscale Model. Simul. 2 (4) (2004) 662–674.
[40] D.M. Tartakovsky, S.P. Neuman, Extension of "transient flow in bounded randomly heterogeneous domains 1. Exact conditional moment equations and recursive approximations", Water Resour. Res. 35 (6) (1999) 1921–1925.
[41] F.J. Molz, H. Rajaram, S. Lu, Stochastic fractal-based models of heterogeneity in subsurface hydrology: origins, applications, limitations, and future research questions, Rev. Geophys. 42 (2004), doi:10.1029/2003RG000126.
[42] C.L. Winter, D.M. Tartakovsky, A. Guadagnini, Moment equations for flow in highly heterogeneous porous media, Surv. Geophys. 24 (1) (2003) 81–106.
[43] M. Loève, Probability Theory, fourth ed., Springer-Verlag, New York, 1977.
[44] P. Frauenfelder, C. Schwab, R.A. Todor, Finite elements for elliptic problems with stochastic coefficients, Comput. Meth. Appl. Mech. Eng. 194 (2005) 205–228.
[45] P. Strait, Sample function regularity for Gaussian processes with the parameter in a Hilbert space, Pac. J. Math. 19 (1) (1966) 159–173.

[46] S. Huang, S. Quek, K. Phoon, Convergence study of the truncated Karhunen–Loève expansion for simulation of stochastic processes, Int. J. Numer. Meth. Eng. 52 (9) (2001) 1029–1043.
[47] D. Xiu, D.M. Tartakovsky, Numerical methods for differential equations in random domains, SIAM J. Sci. Comput. 28 (3) (2006) 1167–1185.
[48] B. Oksendal, Stochastic Differential Equations. An Introduction with Applications, fifth ed., Springer-Verlag, Berlin, 1998.
[49] R. Cools, Monomial cubature rules since 'stroud': a compilation, J. Comput. Appl. Math. 48 (1993) 309–326.
[50] R. Cools, Monomial cubature rules since 'stroud': a compilation – part 2, J. Comput. Appl. Math. 112 (1999) 21–27.
[51] C.L. Winter, A. Guadagnini, D. Nychka, D.M. Tartakovsky, Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers, J. Comput. Phys. 217 (1) (2006) 166–175.
[52] A. Guadagnini, L. Guadagnini, D.M. Tartakovsky, C.L. Winter, Random domain decomposition for flow in heterogeneous stratified aquifers, Stoch. Environ. Res. Risk Assess. 17 (6) (2003) 394–407.
[53] M. Griebel, Adaptive sparse grid multilevel methods for elliptic pdes based on finite differences, Computing 61 (2) (1998) 151–180.
[54] T. Gerstner, M. Griebel, Dimension-adaptive tensor-product quadrature, Computing 71 (1) (2003) 65–87.
[55] X. Ma, N. Zabaras, An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations, J. Comput. Phys. 228 (8) (2009) 3084–3113.
[56] J.Y. Foo, Multi-Element Probabilistic Collocation in High Dimensions: Applications to Systems Biology and Physical Systems, Ph.D. Thesis, Brown University, Division of Applied Mathematics, 2008.
[57] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, Adv. Comput. Math. 12 (2000) 273–288.
[58] E. Novak, K. Ritter, Simple cubature formulas with high polynomial exactness, Constr. Approx. 15 (1999) 499–522.
[59] W. Gautschi, On generating orthogonal polynomials, SIAM J. Sci. Stat. Comput. 3 (3) (1982) 289–317.