Method of distributions for quantification of geologic uncertainty in flow simulations

Hyung Jun Yang[1], Francesca Boso[1], Hamdi A. Tchelepi[1], and Daniel M. Tartakovsky[1]

[1]Department of Energy Resources Engineering, Stanford University, Stanford, CA 94305, USA

Key Points:

- Our method combines the random domain decomposition and the method of distributions.
- The method accounts for geologic and parametric uncertainties.
- The method handles high heterogeneity and is orders of magnitude faster than Monte Carlo.

Corresponding author: Daniel M. Tartakovsky, tartakovsky@stanford.edu

Abstract

Probabilistic models of subsurface flow and transport are required for risk assessment and reliable decision making under uncertainty. These applications require accurate estimates of confidence intervals, which generally cannot be ascertained with such statistical moments as mean (unbiased estimate) and variance (a measure of uncertainty) of a quantity of interest (QoI). The method of distributions provides this information by computing either the probability density function or the cumulative distribution functions (CDF) of the QoI. The method can be orders of magnitude faster than Monte Carlo simulations (MCS), but is applicable to stationary, mildly-to-moderately heterogeneous porous media in which the coefficient of variation of input parameters (e.g., log-conductivity) is below three. Our CDF-RDD framework alleviates these limitations by combining the method of distributions and the random domain decomposition (RDD); it also accounts for uncertainty in the geologic makeup of a subsurface environment. For a given realization of the geological map, we derive a deterministic equation for the conditional CDF of hydraulic head of steady single-phase flow. The solutions of this equation are then averaged over realizations of the geological maps to compute the hydraulic head CDF. Our numerical experiments reveal that the CDF-RDD remains accurate for two-dimensional flow in a porous material composed of two heterogeneous hydrofacies, a setting in which the original CDF method fails. For the same accuracy, the CDF-RDD is an order of magnitude faster than MCS.

1 Introduction

Reliable and accurate predictions of subsurface flow and transport are notoriously elusive due to insufficient site characterization, which manifests itself in uncertainty about a site's geological makeup, spatial variability of its hydraulic properties, external forcings (e.g., initial and boundary conditions, recharge), etc. This uncertainty is characterized by treating input parameters as random fields (Dagan & Neuman, 1997) and solving a stochastic version of the governing equations, whose solutions are probability density functions (PDFs) or cumulative distribution functions (CDFs) of model outputs such as hydraulic head or solute concentration. While Monte Carlo simulations (MCS) can be, and often are, used to compute those, they require a large number of forward model runs (MC realizations) to converge. This number increases with the degree of uncertainty in input parameters, as quantified, e.g., by their variances.

Numerical sampling-based strategies for computing PDFs/CDFs of system states, designed to beat MCS in terms of the computational cost, include multi-level MC, various forms of quasi-MC, and stochastic collocation methods. The method of distributions obviates the need for sample generation by deriving deterministic equations for PDFs or CDFs. It relies on stochastic averaging techniques similar to those routinely used to derive (deterministic) moment differential equations for the first two statistical moments of system states (Neuman et al., 1996; Likanapaisal et al., 2012, and the references therein). The performance of these methods deteriorates with the degree of subsurface heterogeneity, as quantified by the correlation lengths and variances (or, more precisely, coefficients of variation) of the input parameters: Short correlation lengths give rise to the so-called curse of dimensionality, which makes polynomial chaos-based techniques slower than MCS. Large variances undermine the veracity of perturbation-based moment differential equations and PDF/CDF equations.

Probabilistic computations become even more challenging when parameter PDFs exhibit multimodality and/or lack of statistical homogeneity (stationarity). These are manifestations of the presence of multiple geologic materials with distinct (heterogeneous and uncertain) hydraulic and transport properties. Random domain decomposition (RDD) (Winter & Tartakovsky, 2000, 2002) ameliorates these complications by representing a heterogeneous subsurface environment as a union of distinct geological units or hydrofacies. By construction, hydraulic and transport properties of each unit are treated as unimodal, statistically homogeneous random fields with relatively small variances; boundaries between the units, reconstructed from hard and/or soft data, can be uncertain as well. RDD has been used to dramatically enhance the performance of moment differential equations (Winter et al., 2003), generalized polynomial chaos expansions (Xiu & Tartakovsky, 2004),

and stochastic collocation methods (Lin et al., 2010). Here, we use RDD to derive a deterministic equation for the CDF of hydraulic head for flow in highly heterogeneous aquifers with uncertain geology, hydraulic properties, and external forcings.

This approach, which we refer to as CDF-RDD, is presented in section 2. It combines the CDF method for flow in statistically homogeneous porous media (Yang et al., 2019) with RDD that accounts for geologic uncertainty (A. Guadagnini et al., 2003). In section 3, we demonstrate the accuracy and computational efficiency of CDF-RDD via a series of numerical experiments dealing with two-dimensional steady-state flow in a porous medium composed of two materials whose spatial arrangement and hydraulic conductivity are uncertain. Main findings and conclusions drawn from our study are summarized in section 4.

2 Problem Formulation and its Probabilistic Solution

Consider a subsurface environment Ω composed of N_{gu} non-overlapping geological units Ω_i. A spatial arrangement of these units can be provided by expert opinion in the form of a geological map. When sufficient data (measurements of hydraulic conductivity or other discriminating attributes of hydrofacies) are available, one might be able to reconstruct such a map by using geostatistics—e.g., indicator Kriging (L. Guadagnini et al., 2004), object-based geostatistics (Deutsch & Tran, 2002) and multi-point geostatistcs (Strebelle, 2002)—or machine learning tools such as support vector machines (Wohlberg et al., 2005) or nearest-neighbor estimators (Tartakovsky et al., 2007). Regarding of the method used, the resulting geological maps are invariably uncertain.

Steady-state d-dimensional groundwater flow in such an environment is described by

$$\nabla \cdot [K(\mathbf{x})\nabla h(\mathbf{x})] = g(\mathbf{x}), \qquad \mathbf{x} = (x_1, \dots, x_d)^\top \in \Omega, \tag{1}$$

where $K(\mathbf{x})$ is hydraulic conductivity of the porous medium Ω, $h(\mathbf{x})$ is hydraulic head, and $g(\mathbf{x})$ represents point and/or distributed sources and sinks. The groundwater flow equation (1) is subject to boundary conditions

$$h(\mathbf{x}) = \phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad -K(\mathbf{x})\nabla h(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N. \tag{2}$$

Here $\phi(\mathbf{x})$ and $\psi(\mathbf{x})$ are the hydraulic head and the normal component of the Darcy flux $\mathbf{q}(\mathbf{x}) = -K(\mathbf{x})\nabla h(\mathbf{x})$ prescribed, respectively, on the Dirichlet (Γ_D) and Neumann (Γ_N) segments of the boundary $\partial\Omega = \Gamma_D \cup \Gamma_N$ of the flow domain Ω; and $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to Γ_N.

An unknown/unknowable spatial distribution of the hydraulic conductivity $K(\mathbf{x})$ has to be estimated from measurements $K(\mathbf{x}_n)$ collected at N_{meas} (well) locations $\mathbf{x}_n \in \Omega$ ($n = 1, \dots, N_{\text{meas}}$). The presence of multiple hydrofacies Ω_i manifests itself in a histogram of the measurement set $\{K(\mathbf{x}_n)\}_{n=1}^{N_{\text{meas}}}$ (an estimate of the PDF of K) that exhibits multi-modal behavior and its overall standard deviation σ_K is large. This typical setting would increase the computational cost of MCS and invalidate the perturbation-based moment differential equations (Likanapaisal et al., 2012) and PDF/CDF equations (Yang et al., 2019), both of which require the perturbation parameter σ_Y^2 (the variance of log-conductivity $Y = \ln K$) to be relatively small.

We tackle this challenge by using the RDD described in section 2.1. It is deployed in section 2.2 to account for geologic and parametric uncertainties in the context of the method of distributions. An efficient numerical implementation of the resulting CDF-RDD approach for computing the CDF $F_h(H; \mathbf{x})$ of hydraulic head $h(\mathbf{x})$ is described in section 2.3.

2.1 Random Domain Decomposition

RDD treats the porous medium Ω and its hydraulic conductivity $K(\mathbf{x})$ as a two-scale stochastic process. The large scale represents geologic uncertainty, such that a random label α with the PDF $f_\alpha(a)$ encapsulates alternative representations of a site's geology, i.e., uncertain spatial extent of the facies Ω_i ($i = 1, \dots, N_{\text{gu}}$). The small scale accounts for random variability of the hydraulic

conductivity $K(\mathbf{x})$ within each facies Ω_i, which is quantified by the PDF $f_K(k;\mathbf{x}\in\Omega_i)$. Combining the two scales of uncertainty, hydraulic conductivity $K(\mathbf{x})$ is characterized by the joint PDF $f_{K,\alpha}(k,a;\mathbf{x}) = f_{K|\alpha}(k;\mathbf{x}|\alpha=a)f_\alpha(a)$, where $f_{K|\alpha}$ is the PDF of K conditioned on a given geologic map with label a.

By construction, the random hydraulic conductivity $K(\mathbf{x})$ of each sub-domain Ω_i is statistically homogeneous, with unimodal conditional PDFs $f_{K|\alpha=a}(k;\mathbf{x}\in\Omega_i|\alpha=a)$ and relatively small variances $\sigma^2_{Y_i}$ of log conductivity $Y_i = Y(\mathbf{x})$ for all $\mathbf{x}\in\Omega_i$. We model the hydraulic conductivity of each facies, $K_i \equiv K(\mathbf{x})$ for all $\mathbf{x}\in\Omega_i$, as a second-order stationary multivariate log-normal field with constant mean $\bar{K}_i$ and variance $\sigma^2_{K_i}$. It has a correlation function $\rho_{K_i}(r/\ell_{K_i})$, where ℓ_{K_i} is the correlation length, and $r = |\mathbf{x}-\mathbf{y}|$ is the distance between any two points $\mathbf{x},\mathbf{y}\in\Omega_i$. To simplify the presentation, we assume $K_i(\mathbf{x})$ and $K_j(\mathbf{x})$ with $i\neq j$ to be mutually uncorrelated; RDD can readily account for cross-correlations between hydraulic properties of different facies at the cost of slightly increased mathematical complexity (Winter et al., 2006).

With these preliminaries, we replace (1) with

$$\nabla\cdot[K_i(\mathbf{x})\nabla h_i(\mathbf{x})] = g(\mathbf{x}), \qquad \mathbf{x}\in\Omega_i, \qquad i=1,\dots,N_{\text{gu}}, \tag{3}$$

which is subject to boundary conditions (2) and the continuity conditions

$$h_i(\mathbf{x}) = h_j(\mathbf{x}), \qquad K_i(\mathbf{x})\nabla h_i(\mathbf{x})\cdot\mathbf{n}_i(\mathbf{x}) = K_j(\mathbf{x})\nabla h_j(\mathbf{x})\cdot\mathbf{n}_j(\mathbf{x}), \qquad \mathbf{x}\in\Gamma_{ij}. \tag{4}$$

defined on the contact interfaces $\Gamma_{ij} = \Omega_i\cap\Omega_j$ between the adjacent facies Ω_i and Ω_j ($i\neq j$). In (3) and (4), the subscript of h indicates the hydraulic head inside the corresponding facies. This problem formulation is beneficial because it enables one to use small variances $\sigma^2_{Y_i}$ within each facies Ω_i as perturbation parameters. This has been done before to derive moment differential equations (Winter et al., 2003); here, we use it to derive a deterministic equation for the full CDF of hydraulic head.

2.2 Combined CDF-RDD Approach

For a given geological map, defined by the label (realization) $\alpha = a$, we show in Appendix A that the conditional CDF $F_{h|\alpha}(H;\mathbf{x}|\alpha=a)$ of hydraulic head $h(\mathbf{x})$ in (3) satisfies a deterministic $(d+1)$-dimensional differential equation

$$\nabla\cdot(\bar{K}(\mathbf{x})\nabla F_{h|\alpha}) + \frac{\partial(UF_{h|\alpha})}{\partial H} = \upsilon F_{h|\alpha}, \qquad \tilde{\mathbf{x}} = (x_1,\dots,x_d,H)^\top\in\tilde{\Omega}. \tag{5}$$

The (ensemble) averaged hydraulic conductivity $\bar{K}(\mathbf{x})$ takes the constant value of $\bar{K}_i$ for $\mathbf{x}\in\Omega_i$, where $i=1,\dots,N_{\text{gu}}$. The PDF equation (5) is defined on the domain $\tilde{\Omega}\equiv\Omega\times(H_{\min},H_{\max})$, where $H_{\min}$ and $H_{\max}$ are, respectively, the minimum and maximum values hydraulic head $h(\mathbf{x})$ can take in the simulation domain Ω; and

$$U = \upsilon(H-\bar{h}) + \nabla\cdot(\bar{K}\nabla\bar{h}) + 2g, \quad \upsilon = \frac{\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - V}{\sigma_h^2}, \quad V = -\frac{1}{2}\bar{K}\nabla^2\sigma_h^2. \tag{6}$$

The coefficients (6) contain the conditional mean, $\bar{h}$, and variance, σ_h^2, of hydraulic head $h(\mathbf{x})$ at $\mathbf{x}\in\Omega$. These statistical moments can be computed with various techniques, including MCS. In our implementation, we use deterministic moment equations Appendix B, which prove to be more computationally efficient than MCS.

The CDF equation (5) is derived by deploying the self-consistent closure approximation (Yang et al., 2019) that ensures that the differential equations for the moments $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$, obtained by integrating (5), are identical to the moment equations derived in Appendix B. This is in contrast to the interaction-with-the-mean closures (Pope, 2001; Raman et al., 2005; Haworth, 2010) used in turbulence and combustion. The latter fail to preserve a system state's variance and, under certain conditions, its mean (Boso & Tartakovsky, 2016; Yang et al., 2019).

Boundary and interfacial conditions for the CDF equation (5) are derived in Appendix A. If the boundary functions $\phi(x)$ and $\psi(x)$ are uncertain and treated as random fields with one-point

CDFs $F_\phi(\Phi;\mathbf{x})$ and $F_\psi(\Psi;\mathbf{x})$, then (5) is subject to boundary conditions

$$F_{h|\alpha} = F_\phi, \quad \mathbf{x}\in\Gamma_D; \qquad -\nabla F_{h|\alpha}\cdot\mathbf{n} = [\gamma(\mathbf{x})(H-\bar{h})+\eta(\mathbf{x})]\frac{\partial F_{h|\alpha}}{\partial H}, \quad \mathbf{x}\in\Gamma_N, \tag{7}$$

where

$$\gamma(\mathbf{x}) = \frac{\bar{K}\sigma_h^2\cdot\mathbf{n}}{2\sigma_h^2-4\bar{h}^2}, \qquad \eta(\mathbf{x}) = \bar{K}\nabla\bar{h}\cdot\mathbf{n}-\bar{\psi}, \tag{8}$$

$\bar{\psi}(\mathbf{x})$ is the mean of the boundary flux $\psi(\mathbf{x})$, and Γ_D and Γ_N are portions of the Dirichlet and Neumann boundaries, respectively, that intersect Ω. Finally, the general property of a CDF provides the remaining boundary conditions in H space,

$$F_h(H=H_{\min};\mathbf{x}) = 0, \qquad F_h(H=H_{\max};\mathbf{x}) = 1. \tag{9}$$

The straightforward formulation for the boundary conditions in the phase space is a key advantage of CDF equations over PDF equations, for which the corresponding boundary conditions may not be uniquely defined and have to be supplemented with the conservation of probability condition.

We recall that a combination of the solutions to (5)–(9) in each subdomain Ω_i is the conditional CDF $F_{h|\alpha}$, i.e., the CDF of h conditioned on a given geological map with the label $\alpha = a$. This boundary-value problem has to be solved repeatedly for different geological realization α. The average of these solutions over all possible realizations of α is the CDF $F_h(H;\mathbf{x})$,

$$F_h(H,\mathbf{x}) = \int F_{h|\alpha}(H;\mathbf{x}|\alpha=a)f_\alpha(a)\mathrm{d}a. \tag{10}$$

The latter provides a probabilistic prediction of hydraulic head $h(\mathbf{x})$, which accounts for uncertainty in both a site's geology and hydraulic conductivity.

2.3 Numerical Implementation

Numerical solution of the boundary-value problem (5)–(9) is computed in three steps. The first step involves finite-volume solutions of the moment equations (B4)–(B13), i.e., provides numerical approximations of the mean and variance of the hydraulic head, $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$. This step relies on the research code developed in (Likanapaisal et al., 2012).

The second step consists of numerical solution of (5) and (6). Among the plethora of schemes for solving a linear advection-diffusion-reaction equation, such as (5), we utilize a finite-volume scheme in which each facies $\tilde{\Omega}$ is divided into N_{fv} non-overlapping domains $\tilde{\Omega}_1,\ldots,\tilde{\Omega}_{N_{\text{fv}}}$ forming a partition $\mathcal{P}_i$ of $\tilde{\Omega}_i$. A finite-volume solution of (5) is obtained by integrating this equation over each element $\tilde{\Omega}_k$ of the partition $\mathcal{P}$ and using the Gauss-Ostrogradsky theorem to replace the volume integrals over $\tilde{\Omega}_k$ with the surface integrals over their surface $\partial\tilde{\Omega}$,

$$\int_{\partial\tilde{\Omega}_k}\boldsymbol{\mathcal{F}}\cdot\mathbf{n}_e\mathrm{d}\tilde{\mathbf{x}} = \int_{\tilde{\Omega}_k}\beta F_{h|\alpha}\mathrm{d}\tilde{\mathbf{x}}, \qquad \boldsymbol{\mathcal{F}}(\tilde{\mathbf{x}}) = (\bar{K}\nabla_{\mathbf{x}}F_{h|\alpha}, UF_{h|\alpha})^\top \tag{11}$$

where $\mathbf{n}_e$ is the outward unit normal vector of the interface $\partial\tilde{\Omega}_{k_i}$. The discrete form of (11) is

$$\sum_{j\in\text{adj}(i)} T_{ij}(F_{h|\alpha}^{(j)}-F_{h|\alpha}^{(i)}) + U_e^+F_{h|\alpha}^{(i)}A_{ij} + U_e^-F_{h|\alpha}^{(j)}A_{ij} = \beta_i F_{h|\alpha}^{(i)}V_{k_i}, \qquad i=1,\ldots,N_{\text{fv}}, \tag{12}$$

where $\text{adj}(i)$ is the set of neighbors of i, V_{k_i} is the volume of $\tilde{\Omega}_{k_i}$, and

$$T_{ij} = \frac{K_{ij}A_{ij}}{\delta x_{ij}}, \qquad U_e^+ = \frac{U_e+|U_e|}{2}, \qquad U_e^- = \frac{U_e-|U_e|}{2}, \qquad U_e = Un_{e,H}. \tag{13}$$

Here $n_{e,H}$ is the H-direction component of $\mathbf{n}_e$; and K_{ij}, A_{ij}, and δx_{ij} are harmonically averaged hydraulic conductivity, differential element cross-sectional area, and the distance between the spatially connected computational nodes i and j, respectively. A main advantage of our finite volume implementation for (5) is that it does not require any continuity conditions across the interfaces

between the adjacent facies. This results in considerable computational speed-up over other RDD-based numerical schemes that enforce the continuity iteratively (A. Guadagnini et al., 2003; Winter et al., 2003). Combining (12) into a single system of linear algebraic equations, we obtain $\mathbf{A}\mathbf{f} = \mathbf{0}$ where $\mathbf{A}$ is the $N_{\text{fv}} \times N_{\text{fv}}$ coefficient matrix and $\mathbf{f}$ is the $N_{\text{fv}} \times 1$ solution vector for the conditional CDF $F_{h|\alpha}$. We use the bi-conjugate gradient stabilized method to solve this system. Since the coefficients of the CDF equation (5) involve ensemble averages (e.g., $\bar{K}$), they are smoother than their randomly fluctuating counterparts (e.g., K). Consequently, coarser meshes (smaller values of N_{fv}) can be used to solve (5) than to solve MC realizations, providing an additional boost to the computational efficiency of the CDF method.

The third, and last, step is to compute the hydraulic head CDF F_h from its conditional counterpart $F_{h|\alpha}$. This step involves numerical evaluation of the integral in (10). We approximate this integration with the Monte Carlo average of N_α realizations of the geological map with the label α. Numerical integration and differentiation of $F_{h|\alpha}$, used to compute the conditional mean and variance of h, are carried out with the Gaussian quadrature rule and central difference, respectively.

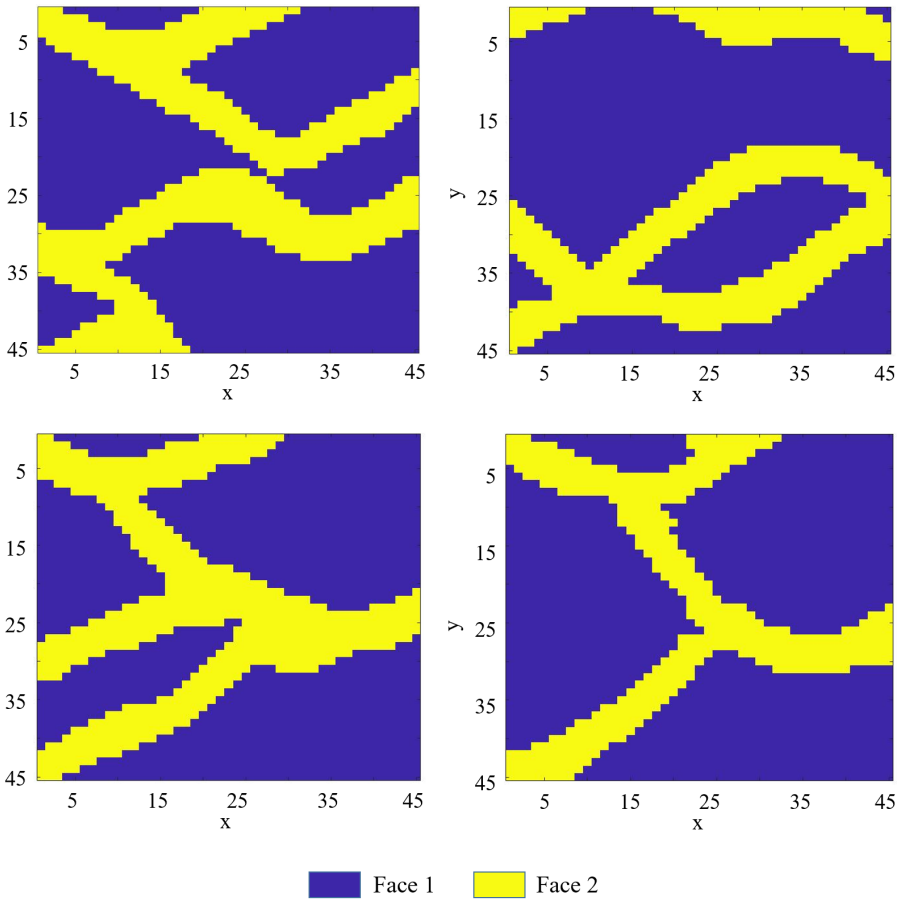

Figure 1: Equiprobable geological models used in numerical experiments.

3 Numerical Experiments

We illustrate the accuracy and efficiency of the CDF-RDD approach on two examples dealing with two-dimensional mean-uniform and convergent flows in a statistically inhomogeneous environment composed of distinct heterogeneous facies. These examples represent two typical flow

scenarios. The first is regional flow driven by internally imposed hydraulic head gradient, the second is radial flow towards a pumping well.

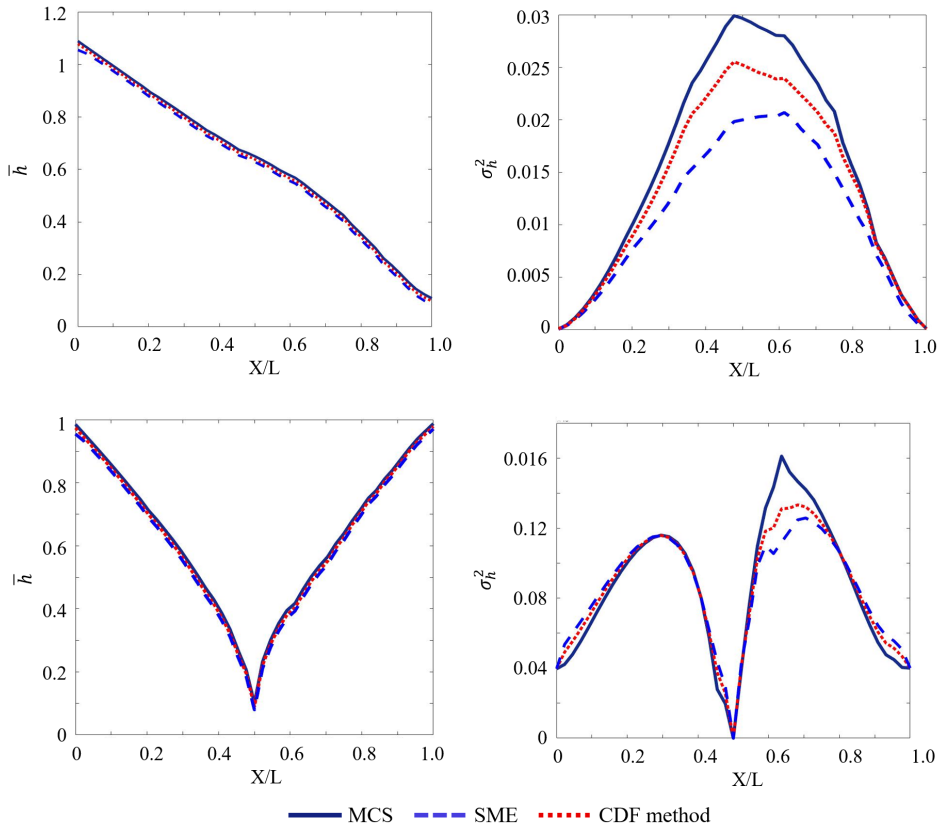

Figure 2: Mean (left column) and variance (right column) of hydraulic head in mean uniform flow (top row) and flow to a well (bottom row). These moments are alternatively computed with Monte Carlo simulations (MCS), the RDD-enhanced SME (SME-RDD), and evaluating the moments of F_h in the CDF equation with and without RDD (CDF and CDF-RDD, respectively).

In both flow regimes, the flow domain Ω, a square of unit dimensionless length (normalized with the domain size L), is composed of $N_{\text{gu}} = 2$ facies whose uncertain spatial arrangement is represented by four different equiprobable geological models labeled by α (fig. 1). These models are generated using multi-point geostatistics, specifically SNESIM algorithm (Strebelle, 2002). The log-hydraulic conductivity of each sub-domain, $Y_i(\mathbf{x}) = \ln K_i(\mathbf{x})$ $(i = 1, 2)$, is a second-order stationary multivariate Gaussian field with an isotropic exponential correlation function $\rho_{Y_i}(r) = \exp(-|\mathbf{x}-\mathbf{y}|/\ell_{Y_i})$ and the dimensionless (normalized with L) correlation length ℓ_{Y_i}. In the simulations reported below, we set $\bar{Y}_1 = 0$, $\bar{Y}_2 = 5$, $\sigma^2_{Y_1} = \sigma^2_{Y_2} = 1$, and $\ell_{Y_1} = 0.3$.

The mean uniform flow is driven by a constant hydraulic head gradient $J \equiv (h_{\text{out}} - h_{\text{in}})/L = 0.1$, with the deterministic dimensionless hydraulic heads $h_{\text{in}} = 1.1$ and $h_{\text{out}} = 0.1$ (normalized with the reference hydraulic head h_{ref}) along $x_1 = 0$ and $x_1 = 1$, respectively. The radial flow is induced by a pumping well at the center of the domain, $(x_1, x_2) = (0.5, 0.5)$, which is controlled by a fixed dimensionless hydraulic head of $h_{\text{well}} = 0.1$; the boundary head h_D along $x_1 = 0$ and $x_1 = 1$ is now uncertain and modeled as a Gaussian field with the mean $\bar{h}_D = 1.0$ and variance $\sigma^2_{h_D} = 0.04$. For both flow scenarios, no-flow boundary conditions are applied at bottom and top boundaries ($x_2 = 0$ and $x_2 = 1$).

The computation domain $\tilde{\Omega} \equiv \Omega \times (H_{\min} = 0, H_{\max} = 1.2)$ is discretized using 45, 45, and 120 nodes in the x_1, x_2, and H directions, respectively. Thus, the total number of grids N_{fv} is 243,000.

We compare the performance of our CDF-RDD method with that of MCS. For each geological model, equiprobable MC realizations of $Y_i(\mathbf{x})$ are generated by the sequential Gaussian simulator (Deutsch & Journel, 1998). The convergence study of the two flow scenarios revealed that, for each geological model, it takes $N_{\text{MCS}} = 10^4$ realizations for MCS estimates of the exceedance probability $\mathbb{P}[h(\mathbf{x}) > H] = 1 - F_h(H;\mathbf{x})$ to stabilize with less than 0.01 of the coefficient of variation. Hence, the total number of MCS realizations is $N_{\text{MCS}} = 4 \cdot 10^4$, a prohibitively large number in most applications of practical significance.

3.1 Accuracy of the CDF method

Both the statistical moment equations (SME) and the CDF equation are derived via perturbation expansions in the variance of log-conductivity. If one were to treat the porous medium in fig. 1 as a single continuum, it would be characterized by the variance $\sigma_Y^2 \sim (\bar{Y}_1 - \bar{Y}_2)^2$ (Winter et al., 2003); for the parameters used in our experiments, $\sigma_Y^2 \approx 7$. Such a large variance is expected to undermine the accuracy of the moment and CDF equations derived without recourse to RDD. Figure 2 demonstrates this to be the case even for the mean, $\bar{h}(\mathbf{x})$, and variance, $\sigma_h^2(\mathbf{x})$, of the hydraulic head $h(\mathbf{x})$, let alone its CDF. This figure compares the results of MCS, which are treated as exact, to three alternative methods for computing these statistics: the RDD-enhanced SME, and evaluating the moments of F_h in the CDF equation with and without RDD. Figure 2 shows that the CDF method without RDD fails to predict the hydraulic head mean $\bar{h}$ and variance σ_h^2 with reasonable accuracy. Yet, the moments computed with SME-RDD and CDF-RDD are in close agreement with those computed via the reference MCS in both flow scenarios. By construction, F_h obtained from the CDF method is to have the same moments $\bar{h}$ and σ_h^2 as their counterparts computed with SME; a slight (about 0.2% for $\bar{h}$ and 4.7% for σ_h^2) disagreement between the two is due to the numerical error in computing the quadratures.

A natural interpretation of the hydraulic head CDF $F_h(H;\mathbf{x})$ is the probability $\mathbb{P}[h(\mathbf{x}) > H] = 1 - F_h(H;\mathbf{x})$ of hydraulic head $h(\mathbf{x})$ at any point $\mathbf{x}$ exceeding a mandated value H. Such exceedance probability maps are required for probabilistic risk assessment and delineation of, e.g., sustainable yield areas or well protection zones with a desired confidence level. The maps of $\mathbb{P}[h(\mathbf{x}) > 0.85]$ obtained with CDF-RDD and MCS are virtually indistinguishable (by the "eyeball measure") from each other, and appreciably different from the map constructed via the CDF approach without RDD (fig. 3). Figure 4 elaborates this point further by presenting the CDF estimates $F_h(H;\mathbf{x})$, obtained with MCS, CDF-RDD, and the CDF method, at several points $\mathbf{x}$ in the computational domain.

A more quantitative comparison between the alternative CDF (or exceedance probability) estimates is provided by the first Wasserstein distance (aka Earth Mover's metric) (Yang et al., 2019; Boso & Tartakovsky, 2016),

$$\mathcal{D}(\mathbf{x}) \equiv \int_{H_{\min}}^{H_{\max}} \mid F_h(H;\mathbf{x}) - F_h^{\text{MCS}}(H;\mathbf{x}) \mid \mathrm{d}H, \qquad \mathcal{D}_{\text{ave}} = \frac{1}{V}\int_{\Omega} \mathcal{D}(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{14}$$

where V is the volume of the simulation domain Ω, $F_h^{\text{MCS}}(H;\mathbf{x})$ is the "exact" CDF computed via MCS, and $F_h(H;\mathbf{x})$ is its approximation obtained from either the CDF-RDD method or the original CDF equation. In the two flow regimes considered, $\mathcal{D}(\mathbf{x})$ of the original CDF method is relatively large throughout the simulation domain, being larger than 0.15 for the mean uniform flow and 0.09 for the convergent flow (fig. 5). The Wasserstein distance for the proposed CDF-RDD approach is an order of magnitude smaller, not exceeding 0.035 for mean uniform flow and 0.016 for convergent flow. These results demonstrate that RDD extends the CDF method to statistically inhomogeneous formations with complex and uncertain geology.

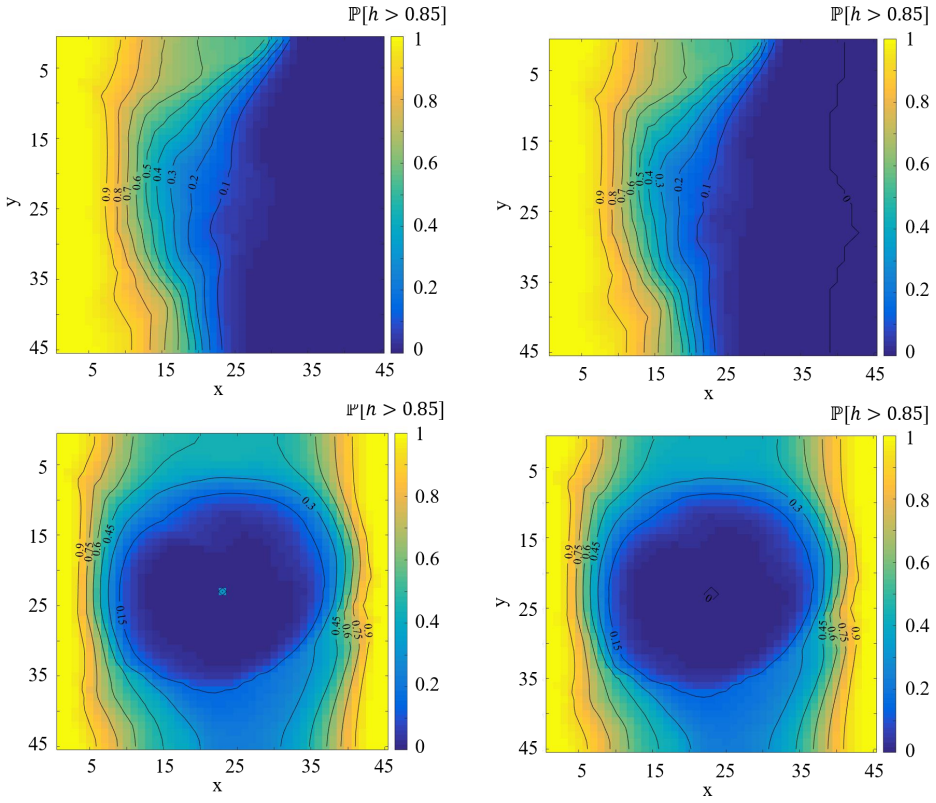

Figure 3: Spatial maps of exceedance probability $\mathbb{P}[h(\mathbf{x}) > H = 0.85] = 1 - F_h(H = 0.85; \mathbf{x})$ obtained with MCS (top row), the CDF method (center row), and CDF-RDD (bottom row) for mean uniform flow (left column) and convergent flow (right column).

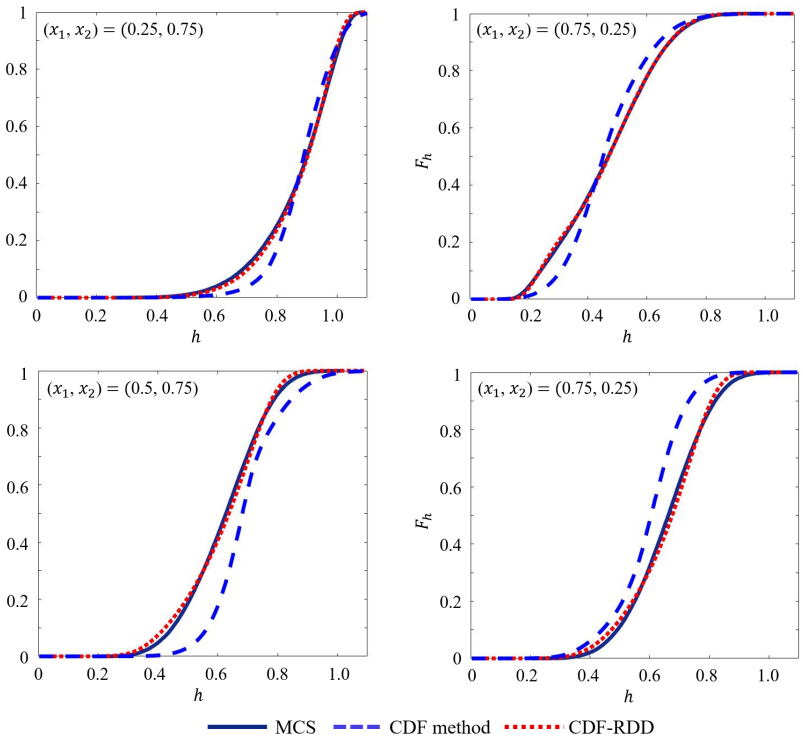

Figure 4: Hydraulic head CDFs F_h computed with MCS, CDF method, and CDF-RDD at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the simulation domain for the mean uniform flow (top row) and the convergent flow (bottom row).

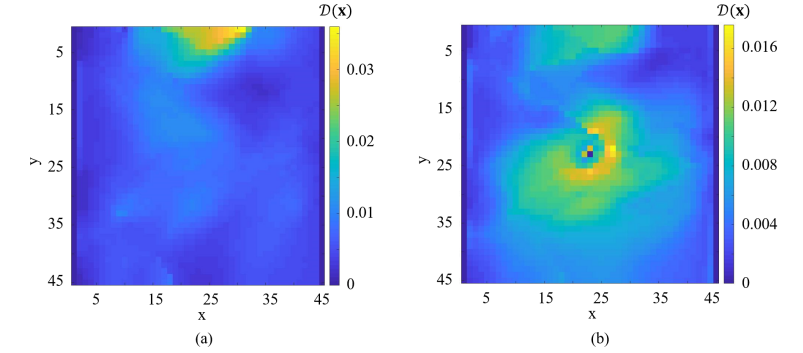

Figure 5: Spatial maps of the Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the "exact" MCS estimate of the hydraulic head CDF F_h^{MCS} and its approximations provided by either the CDF method (top row) or CDF-RDD (bottom row), for the mean uniform flow (left column) and the convergent flow (right column).

3.2 Computational Efficiency of CDF-RDD

The CDF method has been shown to be an order of magnitude faster than MCS in statistically homogeneous media (Yang et al., 2019). Instead of multiple MC solves of the d-dimensional groundwater flow equation (1), it solves a single $(d+1)$-dimensional CDF equation (5). The relative smoothness of the coefficients in the CDF and moment equations allows for the use of coarser meshes and increases the efficiency of a linear solver. The proposed CDF-RDD method retains these features and, hence, one should expect it to be computationally more efficient than MCS in complex geologic settings as well.

These general considerations are confirmed in table 1, which collates the computational costs and accuracy (expressed in terms of the average Wassertein distance $\mathcal{D}_{\text{ave}}$) of the CDF method and CDF-RDD methods. It takes $N_{\text{MCS}} = 6,040$ and $N_{\text{MCS}} = 7,320$ Monte Carlo realizations to obtain the discrepancy levels of CDF-RDD in the mean uniform flow and convergent flow, respectively. For the same discrepancy level, CDF-RDD is an order of magnitude faster than MCS. The discrepancy level of the CDF method without RDD is relatively large, $D_{\text{ave}} \approx 0.03$, but it provides a four-fold speedup relative to CDF-RDD.

Table 1: Computational times and accuracy of MCS and the CDF and CDF-RDD methods.

Flow regime	Method	Error $\mathcal{D}_{\text{ave}}$	CPU time (min)
Mean uniform flow	CDF method	$3.46 \cdot 10^{-2}$	$2.95 \cdot 10^{0}$
	CDF-RDD	$6.56 \cdot 10^{-3}$	$9.07 \cdot 10^{0}$
	MCS with $N_{\text{MCS}} = 6040$	$6.56 \cdot 10^{-3}$	$1.07 \cdot 10^{2}$
	MCS with $N_{\text{MCS}} = 4 \cdot 10^4$	0	$8.43 \cdot 10^{2}$
Convergent flow	CDF method	$3.2 \cdot 10^{-2}$	$3.08 \cdot 10^{0}$
	CDF-RDD	$4.90 \cdot 10^{-3}$	$1.13 \cdot 10^{1}$
	MCS with $N_{\text{MCS}} = 7320$	$4.90 \cdot 10^{-3}$	$1.35 \cdot 10^{2}$
	MCS with $N_{\text{MCS}} = 4 \cdot 10^4$	0	$8.61 \cdot 10^{2}$

4 Summary and Conclusions

We proposed the integrated CDF-RDD framework to quantify geologic and parametric uncertainty in groundwater flow models. The original CDF method for groundwater modeling (Yang et al., 2019) provides a computationally efficient alternative to MCS, but its applicability is limited to statistically homogeneous fields. This limitation has been overcome by deploying RDD (Winter & Tartakovsky, 2000). A key component of CDF-RDD is the derivation of a deterministic equation satisfied by a conditional CDF $F_{h|\alpha}$, the CDF of hydraulic head $h(\mathbf{x})$ conditioned on a realization (labeled by α) of the site geology. The sample average, over alternative geological maps (multiples values of α), of the solutions of this CDF equation yields the hydraulic head CDF F_h. We performed a series of numerical experiments to demonstrate the accuracy and computational efficiency of the CDF-RDD method. Our study leads to the following conclusions.

- The CDF-RDD method yields accurate estimates of the hydraulic head CDF (exceedance probability) for statistically inhomogeneous porous media in both linear and radial flow regimes.
- Unlike its original incarnation, the CDF-RDD method accounts for geologic uncertainty and is applicable to highly heterogeneous subsurface environments.
- For the same accuracy, the CDF-RDD method is an order of magnitude faster MCS in both radial and linear flow regimes.

- The CDF-RDD method provides information that is necessary for probabilistic risk assessment and rare event analysis.

Appendix A CDF Equation for Flow in Composite Porous Media

Our goal is to compute, for each geological map α, $\mathbb{P}[h(\mathbf{x}) \geq H] \equiv F_{h|\alpha}(H;\mathbf{x})$, the probability of an uncertain model prediction of the hydraulic head $h(\mathbf{x})$, at any point $\mathbf{x}$ in the subdomain Ω, exceeding a value H. To simplify the notation, we omit the reference to α in the following formulations. Let us consider a functional $\Pi[H, h(\mathbf{x})] \equiv \mathcal{H}[H - h(\mathbf{x})]$ defined in terms of the Heaviside function $\mathcal{H}(\cdot)$. By definition, the single-point CDF of h, F_h, is computed as the ensemble mean of Π, over all possible values of the random variable h at any point $\mathbf{x} \in \Omega$,

$$F_h(H;\mathbf{x}) = \langle \Pi[H, h(\mathbf{x})] \rangle. \tag{A1}$$

Multiplying (1) with $-\partial\Pi/\partial H$ and accounting for the equality $\nabla\Pi = -(\partial\Pi/\partial H)\nabla h$, we obtain a stochastic $(d+1)$-dimensional PDE for Π,

$$\nabla \cdot (K\nabla\Pi) - K\frac{\partial^2\Pi}{\partial H^2}\nabla h \cdot \nabla h = -g\frac{\partial\Pi}{\partial H}. \tag{A2}$$

Next, we consider a perturbation expansion for the random variables $K(\mathbf{x})$ and $\Pi(U;\mathbf{x})$ in (A2), which are expressed as the sum of their ensemble means and zero-mean fluctuations around these means, e.g., $K = \bar{K} + K'$ and $\Pi = F_h + \Pi'$. Ensemble averaging of the resulting equation yields

$$\nabla \cdot (\bar{K}(\mathbf{x})\nabla F_h) + M = -g\frac{\partial F_h}{\partial H}, \qquad M \equiv \nabla \cdot \langle K'\nabla\Pi' \rangle - \langle K\frac{\partial^2\Pi}{\partial H^2}\nabla h \cdot \nabla h \rangle, \tag{A3}$$

which requires closure approximations to render the mixed moments computable. We use the modified Interaction-by-Exchange-with-the-Mean closure (Boso & Tartakovsky, 2016; Yang et al., 2019), for the unknown mixed moments in (A3)

$$M \approx [\upsilon(\mathbf{x})(H - \bar{h}) + \zeta(\mathbf{x})]\frac{\partial F_h}{\partial H}, \tag{A4}$$

where $\bar{h}(\mathbf{x})$ is the mean hydraulic head, and $\upsilon(\mathbf{x})$ and $\zeta(\mathbf{x})$ are the closure variables determined below. Finally, we obtain a closed CDF equation

$$\nabla \cdot (\bar{K}(\mathbf{x})\nabla F_h) + [\upsilon(\mathbf{x})(H - \bar{h}) + \zeta(\mathbf{x})]\frac{\partial F_h}{\partial H} = -g\frac{\partial F_h}{\partial H}, \qquad \mathbf{x} \in \Omega. \tag{A5}$$

Rearranging the terms, (A5) yields (5) for each subdomain in any given geological realization. Expressions for $\upsilon_i(\mathbf{x})$ and $\zeta_i(\mathbf{x})$ are determined by enforcing consistency between the moments obtained by integration of F_{h_i} obeying the CDF equation

$$\bar{h} = H_{\max} - \int_{H_{\min}}^{H_{\max}} F_h(H;\mathbf{x})\mathrm{d}H, \qquad \sigma_h^2 = H_{\max}^2 - 2\int_{H_{\min}}^{H_{\max}} HF_h(H;\mathbf{x})\mathrm{d}H - \bar{h}^2, \tag{A6}$$

and the moment equations satisfied by $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$, as in (Boso & Tartakovsky, 2016; Boso et al., 2018; Yang et al., 2019). These expressions assume the random variable $h(\mathbf{x})$ and the correspondong CDF $F_h(U;\mathbf{x})$ to be defined on the interval $[H_{\min}, H_{\max}]$.

Since $F_h(H_{\min};\mathbf{x}) = 0$ and $F_h(H_{\max};\mathbf{x}) = 1$, integrating (A5) over H yields

$$\nabla \cdot (\bar{K}(\mathbf{x})\nabla\bar{h}) - \zeta(\mathbf{x}) = g(\mathbf{x}). \tag{A7}$$

Multiplying both sides of (A5) by H, integrating the resulting equation over H and accounting for (A7) yields

$$\nabla \cdot (\bar{K}(\mathbf{x})\nabla\sigma_h^2) + 2\bar{K}(\mathbf{x})\nabla\bar{h} \cdot \nabla\bar{h} - 2\upsilon(\mathbf{x})\sigma_h^2 = 0. \tag{A8}$$

We obtain first-order approximations for $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$, respectively $\tilde{h}(\mathbf{x})$ and $\tilde{\sigma}_h^2$, by solving

$$\bar{K}_i\nabla^2\tilde{h} + \rho_i = g, \quad \rho_i(\mathbf{x}) = \bar{K}_i\nabla \cdot [\lim_{\boldsymbol{\chi}\to\mathbf{x}} \nabla_{\mathbf{x}} C_{Yh(\boldsymbol{\chi},\mathbf{x})}]; \qquad \mathbf{x} \in \Omega_i, \quad \boldsymbol{\chi} \in \Omega \tag{A9}$$

and

$$\bar{K}_i \nabla^2 \tilde{\sigma}_h^2 + 2V(\mathbf{x}) = 0, \qquad \mathbf{x} \in \Omega_i. \tag{A10}$$

Here $C_{Yh}(\mathbf{x}, \boldsymbol{\chi})$ is the first-order approximation (in σ_Y^2) of the cross-covariance $\langle Y'(\mathbf{x}) h'(\boldsymbol{\chi}) \rangle$ and

$$V \equiv \bar{K}_i \lim_{\boldsymbol{\chi} \to \mathbf{x}} [\nabla_{\mathbf{x}} h^{(0)} \cdot \nabla_{\mathbf{x}} C_{Yh}(\mathbf{x}, \boldsymbol{\chi}) - \nabla_{\chi} \cdot \nabla_{\mathbf{x}} C_h(\mathbf{x}, \boldsymbol{\chi})] + \mathrm{e}^{\sigma_{Y_i}^2/2} g(\mathbf{x}) C_{Yh}(\mathbf{x}, \mathbf{x}), \tag{A11}$$

with $\boldsymbol{\chi} \in \Omega$, and $C_h(\mathbf{x}, \boldsymbol{\chi})$ denoting the first-order approximation of auto-covariance $\langle h'(\mathbf{x}) h'(\boldsymbol{\chi}) \rangle$ of the hydraulic head $h(\mathbf{x})$. The derivation via perturbation expansion of (A9) and (A10) is illustrated in Appendix B, where we also list the corresponding boundary and interface conditions. On the other hand, approximations of $\bar{h}(\mathbf{x})$ and σ_h^2, denoted respectively by $\tilde{h}(\mathbf{x})$ and $\tilde{\sigma}_h^2$, satisfy the moment equations, which are obtained via a perturbation expansion (Appendix B).

We impose equivalency (up to the first order in σ_Y^2) between the equations for the mean, (A7) and (A9), and between the equations for the variance, (A8) and (A10). This results in the following expressions for the closing terms $\upsilon(\mathbf{x})$ and $\zeta(\mathbf{x})$

$$\upsilon \equiv (\bar{K} \nabla \bar{h} \cdot \nabla \bar{h} - V)/\sigma_h^2, \quad \zeta \equiv -\rho. \tag{A12}$$

Substituting these expressions in (A5) yields (5) and (6).

The derivation of boundary conditions for the CDF equation is consistent with the derivation of the CDF equation. Using the definition of Π and ensemble averaging leads to the boundary condition along Γ_D for the CDF equation without any closure approximation

$$F_h = F_\phi(U; \mathbf{x}), \quad \mathbf{x} \in \Gamma_D, \tag{A13}$$

where $F_\phi(U; \mathbf{x})$ is the single-point CDF for the random boundary head $\phi(\mathbf{x})$. For the Neumann boundary condition in (2), we multiply by $\partial\Pi/\partial H$ to obtain

$$-K(\mathbf{x}) \nabla \Pi \cdot \mathbf{n}(\mathbf{x}) = -\varphi \frac{\partial \Pi}{\partial H}, \quad \mathbf{x} \in \Gamma_N. \tag{A14}$$

Ensemble averaging of (A14) yields

$$-\bar{K}(\mathbf{x}) \nabla F_h \cdot \mathbf{n}(\mathbf{x}) = -\bar{\varphi} \frac{\partial F_h}{\partial H} + \langle K' \frac{\partial \Pi'}{\partial H} \rangle - \langle \varphi' \frac{\partial \Pi'}{\partial H} \rangle, \quad \mathbf{x} \in \Gamma_N. \tag{A15}$$

Like with the closure developed for (3), we impose

$$-\bar{K}(\mathbf{x}) \nabla F_h \cdot \mathbf{n}(\mathbf{x}) = -\bar{\varphi} \frac{\partial F_h}{\partial H} + \big(\gamma(\mathbf{x})(H - \bar{h}(\mathbf{x})) + \eta(\mathbf{x})\big) \frac{\partial F_h}{\partial H}, \quad \mathbf{x} \in \Gamma_N, \tag{A16}$$

and express $\gamma(\mathbf{x})$ and $\eta(\mathbf{x})$ as in (7) to ensure consistency with the boundary conditions of the moment equation (Appendix B).

Appendix B Moment Differential Equations

The moment equations (MDEs) for highly heterogeneous media composed of distinct geological facies has been widely investigated in the hydrology community. Here, we presents the combination of random domain decomposition (RDD) and SMEs implemented by (Tchelepi & Li, 2004). The formulation of SMEs and their boundary conditions for each facies is analogous to the case of statistically homogeneous porous media, and it is summarized below. These formulations are based on perturbation expansion of the log-conductivity field in each facies, and are formally valid for $\sigma_{Y_i}^2/2 < 1$, although robust for $\sigma_{Y_i}^2$ as large as 4.

In order to obtain an equation the ensemble mean of the hydraulic head in a given geological map α, we introduce log hydraulic conductivity $Y(\mathbf{x}) = \ln K(\mathbf{x})$ where $Y(\mathbf{x}) = Y_i$ when $x \in \Omega_i$. In each subdomain i, the log-conductivity is considered as second-order stationary multivariate Gaussian, with constant mean $\bar{Y}_i$ and variance $\sigma_{Y_i}^2 = \langle Y_i'^2 \rangle$. Hence, if the number of subdomain is greater than 1, the log-conductivity $Y(\mathbf{x})$ has spatially varying mean $\bar{Y}(\mathbf{x})$ and variance $\sigma_Y^2(\mathbf{x})$.

Here and in the following, both $\bar{\mathcal{A}}$ and $\langle\mathcal{A}\rangle$ indicate the ensemble mean of the random quantity $\mathcal{A}$. In (B1), we employ the Reynolds decomposition for the log-conductivity $Y(\mathbf{x}) = \bar{Y}(\mathbf{x}) + Y'(\mathbf{x})$, so that $K(\mathbf{x}) = \exp(Y) = K_{\mathrm{G}}(\mathbf{x})\exp(Y')$, where $K_{\mathrm{G}}(\mathbf{x}) = \exp(\bar{Y})$ is the geometric mean of the hydraulic conductivity K. Then, we rewrite (1) as

$$\nabla \cdot (K_{\mathrm{G}}(\mathbf{x})\mathrm{e}^{Y'}\nabla h) = g(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{B1}$$

where $K_{\mathrm{G}}(\mathbf{x}) = K_{\mathrm{G},i}$ for $\mathbf{x} \in \Omega_i$. We expand $\exp(Y')$ into a Taylor series around $Y'(\mathbf{x}) = 0$. Taking the ensemble average of the resulting equation, and recalling that all odd moments of a Gaussian $Y'(\mathbf{x})$ are zero, yields

$$\nabla \cdot (K_{\mathrm{G}}(\mathbf{x})\nabla\bar{h}) + \nabla \cdot (K_{\mathrm{G}}\langle Y'\nabla h'\rangle) + \text{h.o.t.} = g(\mathbf{x}), \tag{B2}$$

where h.o.t. refers to the terms of order higher than σ_Y^2. We expand $\bar{h}$ and $\langle Y'\nabla h'\rangle$ into asymptotic series in the powers of σ_Y^2,

$$\bar{h} = \bar{h}^{(0)} + \bar{h}^{(1)} + \cdots, \qquad \langle Y'\nabla h'\rangle = \langle Y'\nabla h'\rangle^{(1)} + \langle Y'\nabla h'\rangle^{(2)} + \cdots, \tag{B3}$$

where the superscript $^{(n)}$ indicates the order of the term with respect to σ_Y^2.

An approximation of the ensemble mean is then computed recursively, solving differential equations obtained by retaining the terms of equal powers of σ_Y^{2n}:

$$\nabla \cdot (K_{\mathrm{G}}\nabla\bar{h}^{(0)}) = g, \qquad \nabla \cdot (K_{\mathrm{G}}\nabla\bar{h}^{(1)}) + \nabla \cdot (K_{\mathrm{G}}\langle Y'\nabla h\rangle^{(1)}) + \frac{1}{2}\nabla \cdot (K_{\mathrm{G}}\sigma_Y^2\bar{h}^{(0)}) = 0. \tag{B4}$$

The unknown term $\langle Y'(\mathbf{x})\nabla h(\mathbf{x})\rangle^{(1)}$ is computed as $\langle Y'\nabla h\rangle^{(1)} = \lim_{\boldsymbol{\chi}\to\mathbf{x}}[\nabla_{\mathbf{x}} C_{Yh}(\boldsymbol{\chi},\mathbf{x})]$, where $C_{Yh}(\boldsymbol{\chi},\mathbf{x}) = \langle Y'(\boldsymbol{\chi})h(\mathbf{x})\rangle^{(1)}$ is the first-order approximation of the cross-correlation between log-conductivity and hydraulic head. The latter satisfies (Yang et al., 2019)

$$\nabla_{\mathbf{x}} \cdot (K_{\mathrm{G}}\nabla_{\mathbf{x}} C_{Yh}) + \nabla_{\mathbf{x}}(K_{\mathrm{G}} C_Y \nabla_{\mathbf{x}}\bar{h}^{(0)}) = 0, \quad \boldsymbol{\chi}, \mathbf{x} \in \Omega, \tag{B5}$$

subject to the boundary conditions

$$C_{Yh} = 0, \quad \mathbf{x} \in \Gamma_D; \qquad K_{\mathrm{G}}\nabla_{\mathbf{x}} C_{Yh} \cdot \mathbf{n}(\mathbf{x}) = \bar{\psi} C_Y, \quad \mathbf{x} \in \Gamma_N. \tag{B6}$$

Here $C_Y(\boldsymbol{\chi},\mathbf{x}) = \langle Y'(\boldsymbol{\chi})Y'(\mathbf{x})\rangle$ is an auto-correlation function of the log-condunctivity $Y(\mathbf{x})$. In most applications, the conductivities of two different subdomains Ω_i and Ω_j are uncorrelated, i.e., $C_Y(\boldsymbol{\chi},\mathbf{x}) = 0$ when $\boldsymbol{\chi} \in \Omega_i$ and $\mathbf{x} \in \Omega_j$ $(i \neq j)$. The relative importance of cross-correlations between conductivities of different geological units was investigated in (Winter et al., 2006). Once $C_{Yh}(\boldsymbol{\chi},\mathbf{x})$ is evaluated, we compute $\nabla_{\mathbf{x}} C_{Yh}(\boldsymbol{\chi},\mathbf{x})$ and then evaluate $\langle Y'\nabla h\rangle^{(1)} = \lim_{\boldsymbol{\chi}\to\mathbf{x}} \nabla_{\mathbf{x}} C_{Yh(\boldsymbol{\chi},\mathbf{x})}$.

We approximate the mean head $\bar{h}$ to first order in σ_Y^2 as $\tilde{h} = \bar{h}^{(0)} + \bar{h}^{(1)}$. Since in each subdomain Ω_i the conductivity field is second-order stationary, (B4) gives rise to

$$K_{\mathrm{G},i}\nabla^2\bar{h}^{(0)} = g, \qquad K_{\mathrm{G},i}\nabla^2\bar{h}^{(1)} + K_{\mathrm{G},i}\nabla \cdot [\lim_{\boldsymbol{\chi}\to\mathbf{x}} \nabla_{\mathbf{x}} C_{Yh(\boldsymbol{\chi},\mathbf{x})}] + \frac{\sigma_{Y_i}^2}{2} g = 0, \qquad \mathbf{x} \in \Omega_i. \tag{B7}$$

Summing up these two equations,

$$K_{\mathrm{G},i}\nabla^2\tilde{h} + K_{\mathrm{G},i}\nabla \cdot [\lim_{\boldsymbol{\chi}\to\mathbf{x}} \nabla_{\mathbf{x}} C_{Yh(\boldsymbol{\chi},\mathbf{x})}] = (1 - \sigma_{Y_i}^2/2)g, \qquad \mathbf{x} \in \Omega_i, \quad \boldsymbol{\chi} \in \Omega. \tag{B8}$$

Next, we use the fact that $(1 - \sigma_{Y_i}^2/2) \approx \exp(-\sigma_{Y_i}^2/2)$ as long as $\sigma_{Y_i}^2/2 \ll 1$. With this approximation, and since $\bar{K}_i = K_{\mathrm{G},i}\exp(\sigma_{Y_i}^2/2)$, (B8) yields (A9). A similar procedure leads to the boundary conditions

$$\tilde{h} = \bar{\phi}(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad -\bar{K}\nabla\tilde{h} \cdot \mathbf{n}(\mathbf{x}) = \bar{\psi}(\mathbf{x}). \quad \mathbf{x} \in \Gamma_N, \tag{B9}$$

Subtracting the sum of the two equations in (B4) from (B1) yields an equation for the head fluctuations $h' \approx h - \tilde{h}$. Multiplying the latter with $h'(\mathbf{x})$ and taking the ensemble mean yield an equation for the first-order approximation of the head variance, $\tilde{\sigma}_h^2(\mathbf{x})$,

$$\nabla \cdot (K_\mathrm{G} \nabla \tilde{\sigma}_h^2) - 2K_\mathrm{G} \langle \nabla h' \cdot \nabla h' \rangle^{(1)} + 2K_\mathrm{G} \nabla \bar{h}^{(0)} \cdot \langle h' \nabla Y' \rangle^{(1)} = -2gC_{Yh}(\mathbf{x}, \mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{B10}$$

Using the already evaluated cross-correlation $C_{Yh}(\boldsymbol{\chi}, \mathbf{x})$, we compute $\langle h' \nabla Y' \rangle^{(1)}$ as the limit $\langle h' \nabla Y' \rangle^{(1)} = \lim_{\boldsymbol{\chi} \to \mathbf{x}} [\nabla_{\boldsymbol{\chi}} C_{Yh}(\boldsymbol{\chi}, \mathbf{x})]$. To obtain a workable expression for the unknown term $\langle \nabla h' \cdot \nabla h' \rangle^{(1)}$, we solve the following equation for the first-order approximation of the hydraulic head's auto-covariance function, $C_h(\mathbf{x}, \boldsymbol{\chi}) = \langle h'(\mathbf{x}) h'(\boldsymbol{\chi}) \rangle^{(1)}$,

$$\nabla_\mathbf{x} \cdot (K_\mathrm{G} \nabla_\mathbf{x} C_h) + K_\mathrm{G} \nabla_\mathbf{x} C_{Yh}(\mathbf{x}, \boldsymbol{\chi}) \cdot \nabla_\mathbf{x} \bar{h}^{(0)} = -gC_{Yh}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x}, \boldsymbol{\chi} \in \Omega, \tag{B11}$$

subject to the following boundary conditions

$$C_h = C_{\phi h}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_D; \qquad K_\mathrm{G} \nabla_\mathbf{x} C_h \cdot \mathbf{n} = C_{\psi h}(\mathbf{x}, \boldsymbol{\chi}) - \bar{\psi} C_{Yh}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_N. \tag{B12}$$

Once $\tilde{C}_h(\mathbf{x}, \boldsymbol{\chi})$ is computed, we evaluate $\langle \nabla h' \cdot \nabla h' \rangle^{(1)} = \lim_{\boldsymbol{\chi} \to \mathbf{x}} [\nabla_\mathbf{x} \cdot \nabla_\chi C_h(\boldsymbol{\chi}, \mathbf{x})]$. Then, writing (B10) for individual subdomains Ω_i and multiplying the resulting equations with $\exp(\sigma_{Y_i}^2/2)$ leads to the closed equations (A10) for an approximation of the head variance, subject to boundary conditions

$$\tilde{\sigma}_h^2 = \sigma_\phi^2, \quad \mathbf{x} \in \Gamma_D; \qquad K_\mathrm{G} \nabla_\mathbf{x} \tilde{\sigma}_h^2 \cdot \mathbf{n} = 2C_{\psi h}(\mathbf{x}, \mathbf{x}) - 2\bar{\psi} C_{Yh}(\mathbf{x}, \mathbf{x}), \quad \mathbf{x} \in \Gamma_N. \tag{B13}$$

The cross-covariances $C_{\phi h}(\mathbf{x}, \boldsymbol{\chi})$ and $C_{\psi h}(\mathbf{x}, \boldsymbol{\chi})$ in (B12) and (B13) are derived by multiplying the equation for the head fluctuations $h'(\mathbf{x})$ with ϕ' and ψ', respectively (Neuman et al., 1996).

As an alternative, $\tilde{\sigma}_h^2$ can be computed by taking the limit of the head's auto-covariance function $C_h(\mathbf{x}, \boldsymbol{\chi})$, i.e., $\tilde{\sigma}_h^2 = \lim_{\boldsymbol{\chi} \to \mathbf{x}} C_h(\mathbf{x}, \boldsymbol{\chi})$.

Acknowledgments

This work was supported in part by Air Force Office of Scientific Research under award number FA9550-17-1-0417 and by a gift from Total. There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

References

Boso, F., Marzadri, A., & Tartakovsky, D. M. (2018). Probabilistic forecasting of nitrogen dynamics in hyporheic zone. *Water Resour. Res.*, *54*(7), 4417-4431.

Boso, F., & Tartakovsky, D. M. (2016). The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resour. Res.*, *52*(6), 4700-4712. doi: 10.1002/2016WR018745

Dagan, G., & Neuman, S. P. (Eds.). (1997). *Subsurface flow and transport: A stochastic approach*. New York: Cambridge.

Deutsch, C. V., & Journel, A. (1998). *GSLIB: Geostatistics software library and user's guide*. Oxford Univ. Press.

Deutsch, C. V., & Tran, T. T. (2002). FLUVSIM: a program for object-based stochastic modeling of fluvial depositional systems. *Computers and Geosciences*, *28*(4), 525-535.

Guadagnini, A., Guadagnini, L., Tartakovsky, D. M., & Winter, C. L. (2003). Random domain decomposition for flow in heterogeneous stratified aquifers. *Stochastic Environmental Research and Risk Assessment*, *17*(6), 394-407.

Guadagnini, L., Guadagnini, A., & Tartakovsky, D. M. (2004). Probabilistic reconstruction of geologic facies. *J. Hydrol.*, *294*, 57-67.

Haworth, D. C. (2010). Progress in probability density function methods for turbulent reacting flows. *Progr. Energy Combust. Sci.*, *36*(2), 168-259.

Likanapaisal, P., Li, L., & Tchelepi, H. A. (2012). Dynamic data integration and quantification of prediction uncertainty using statistical-moment equations. *SPE J.*, *17*(1), 98–111.

Lin, G., Tartakovsky, A. M., & Tartakovsky, D. M. (2010). Uncertainty quantification via random domain decomposition and probabilistic collocation on sparse grids. *Journal of Computational Physics*, *229*(19), 6995–7012.

Neuman, S. P., Tartakovsky, D., Wallstrom, T. C., & Winter, C. (1996). Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resour. Res.*, *32*(5), 1479-1480.

Pope, S. B. (2001). *Turbulent flows.* IOP Publishing.

Raman, V., Pitsch, H., & Fox, R. O. (2005). Hybrid large-eddy simulation/Lagrangian filtered-density-function approach for simulating turbulent combustion. *Combust. Flame*, *143*, 56-78.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, *34*(1), 1-21.

Tartakovsky, D. M., Wohlberg, B. E., & Guadagnini, A. (2007). Nearest neighbor classification for facies delineation. *Water Resour. Res.*, *34*, L05404. doi: 10.1029/2007GL029245

Tchelepi, H., & Li, L. (2004). Statistical moment equations for flow in composite heterogeneous porous media. In *Ecmor ix-9th european conference on the mathematics of oil recovery.*

Winter, C. L., Guadagnini, A., Nychka, D., & Tartakovsky, D. M. (2006). Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers. *J. Comput. Phys.*, *217*(1), 166-175.

Winter, C. L., & Tartakovsky, D. M. (2000). Mean flow in composite porous media. *Geophysical Research Letters*, *27*(12), 1759-1762.

Winter, C. L., & Tartakovsky, D. M. (2002). Groundwater flow in heterogeneous composite aquifers. *Water Resour. Res.*, *38*(8). doi: 10.1029/2001WR000450

Winter, C. L., Tartakovsky, D. M., & Guadagnini, A. (2003). Moment equations for flow in highly heterogeneous porous media. *Surv. Geophys.*, *24*(1), 81-106.

Wohlberg, B., Tartakovsky, D. M., & Guadagnini, A. (2005). Subsurface characterization with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(1), 47-57.

Xiu, D., & Tartakovsky, D. M. (2004). A two-scale non-perturbative approach to uncertainty analysis of diffusion in random composites. *Multiscale Modeling & Simulation*, *2*(4), 662–674.

Yang, H.-J., Boso, F., Tchelepi, H. A., & Tartakovsky, D. M. (2019). Probabilistic forecast of single-phase flow in porous media with uncertain properties. *Water Resour. Res.*, *55*(11), 8631-8645. doi: 10.1029/2019WR026090