Water Resources Research

Probabilistic Forecast of Single-Phase Flow in Porous Media With Uncertain Properties

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, USA

Abstract Uncertainty about geologic makeup and properties of the subsurface renders inadequate a unique quantitative prediction of flow and transport. Instead, multiple alternative scenarios have to be explored within the probabilistic framework, typically by means of Monte Carlo simulations (MCS). These can be computationally expensive, and often prohibitively so, especially when the goal is to compute the tails of a distribution, that is, probabilities of rare events, which are necessary for risk assessment and decision making under uncertainty. We deploy the method of distributions to derive a deterministic equation for the cumulative distribution function (CDF) of hydraulic head in an aquifer with uncertain (random) hydraulic conductivity. The CDF equation relies on a self-consistent closure approximation, which ensures that the resulting CDF of hydraulic head has the same mean and variance as those computed with statistical moment equations. We conduct a series of numerical experiments dealing with steady-state two-dimensional flow driven by either a natural hydraulic head gradient or a pumping well. These experiments reveal that the CDF method remains accurate and robust for highly heterogeneous formations with the variance of log conductivity as large as five. For the same accuracy, it is also up to four orders of magnitude faster than MCS in computing hydraulic head with a required degree of confidence (probability).

1. Introduction

Quantitative predictions of fluid flow in subsurface environments are compromised by multiscale heterogeneity and insufficient site characterization. These factors introduce uncertainty in input parameters (e.g., hydraulic conductivity and storage coefficient) and forcings (e.g., initial and boundary conditions and recharge rate), rendering model outputs uncertain as well. Quantification of predictive uncertainty is typically done within the probabilistic framework, which equates uncertainty with randomness. Thus, uncertain inputs and outputs of a given model, for example, the groundwater flow equation, are treated as spatiotemporal random fields characterized by corresponding probability density functions (PDFs) or cumulative distribution functions (CDFs). In other words, such a model has infinitely many solutions some of which are more likely than others; to assign probability to a particular solution, for example, hydraulic head $h(\mathbf{x})$ at any point $\mathbf{x}$ of a simulation domain, one has to compute hydraulic head's CDF $F_h(H; \mathbf{x}) \equiv \mathbb{P}[h \leq H; \mathbf{x}]$, the probability that an uncertain prediction of head h at point $\mathbf{x}$ does not exceed a value H.

Such information is required for risk assessment and decision making under uncertainty (e.g., Tartakovsky, 2007, 2013), yet it is absent in most stochastic analyses of subsurface flow and transport, which focus on the first two statistical moments of a system state, for example, on mean head $\bar{h}(\mathbf{x})$ as its "best" prediction and head variance $\sigma_h^2(\mathbf{x})$ as a measure of predictive uncertainty (e.g., among many others, Dagan & Neuman, 1997; Li et al., 2003; Neuman et al., 1996). Monte Carlo simulations (MCS) can be used to compute the CDF $F_h(H; \mathbf{x})$. However, this approach requires a large number of Monte Carlo (MC) realizations to estimate the tails of $F_h(H; \mathbf{x})$, considerably more than that required to estimate $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$ with the same accuracy. When a single model run is computationally expensive, the use of MCS to calculate F_h might become unfeasible.

Numerical strategies aiming to outperform MCS in terms of computational efficiency include quasi-MC (Caflisch, 1998), multilevel MC (Giles et al., 2015), and various stochastic finite element methods (Xiu, 2010). While widely used in practice, including for subsurface-related applications (e.g., Ciriello et al., 2017; Dodwell et al., 2015; Liodakis et al., 2018; and the references therein), under certain conditions such methods

can be slower than MCS. For example, multilevel MC might become slower than regular MC when estimating a system state's distribution to the same accuracy (Giles et al., 2015), and polynomial chaos-based techniques have been shown to underperform MC if random parameter fields in (nonlinear) models exhibit short correlation lengths and/or high variances (Barajas-Solano & Tartakovsky, 2016).

The method of distributions (Tartakovsky & Gremaud, 2016) provides another alternative to MCS by deriving a single deterministic equation for either PDF or CDF of a system state. It often treats nonlinearities in a governing equation exactly and remains robust and efficient for coefficients with short correlation length, including white noise. The method has been used extensively to quantify parametric uncertainty in hyperbolic problems, such as nonlinear advection-reaction transport (Boso et al., 2014; Lichtner & Tartakovsky, 2003; Shvidler & Karasaki, 2003; Tartakovsky & Broyda, 2011) and multiphase flow described by the Buckley-Leverett equation (Ibrahima et al., 2015, 2018; Wang et al., 2013). To the best of our knowledge, development of the method of distributions for elliptic problems with random coefficients (e.g., steady-state groundwater equation with uncertain hydraulic conductivity) remains an open challenge.

That is because the Laplace operator in parabolic and elliptic equations requires a closure approximation for the PDF or CDF equations. In turbulence and combustion literature, such a closure is obtained with the interaction by exchange with the mean (IEM) approximation (Villermaux & Falk, 1994) or its subsequent modifications (Raman et al., 2005). By construction, these closures preserve the mean of a state variable but have been shown to give incorrect estimates of its variance. The self-consistent closure of Boso & Tartakovsky (2016) ameliorates this deficiency by preserving both the mean and variance. It has been used to quantify uncertainty in advection-dispersion (Boso & Tartakovsky, 2016) and advection-dispersion-reaction (Boso et al., 2018) problems.

We develop the method of distributions for single-phase flow in subsurface environments with uncertain hydraulic conductivity and external forcings. Section 2 contains a formulation of groundwater flow problem with uncertain inputs and a derivation of the PDF and CDF equations for hydraulic head. In section 3, we compare numerical solutions of the CDF equation with MCS results in terms of their accuracy and computational efficiency. In this section we also demonstrate the robustness of the proposed method by analyzing its performance for different degrees of input uncertainty (variance of log hydraulc conductivity). Main findings and conclusions drawn from our study are summarized in section 4.

2. Problem Formulation and Method of Distributions

In this section we provide a probabilistic description of single-phase flow in a heterogeneous porous medium with uncertain hydraulic conductivity $K(\mathbf{x})$, and derive a deterministic equation for CDF $F_h(H; \mathbf{x})$ of hydraulic head $h(\mathbf{x})$.

2.1. Single-Phase Flow in Porous Media

Steady-state flow in a d-dimensional saturated heterogeneous porous medium $\Omega \subset \mathbb{R}^d$ is described by the groundwater flow equation

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

subject to boundary conditions

$$h(\mathbf{x}) = \Phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \quad \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N; \quad \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(x) + ah(\mathbf{x}) = \varphi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_R. \tag{2}$$

Here $g(\mathbf{x})$ represents point and/or distributed sources and sinks; $\Phi(\mathbf{x})$, $\psi(\mathbf{x})$ and $\varphi(\mathbf{x})$ are the hydraulic head, the normal component of the Darcy flux $\mathbf{q}(\mathbf{x}) = -K(\mathbf{x})\nabla h(\mathbf{x})$, and their linear combination prescribed, respectively, on the Dirichlet (Γ_D), Neumann (Γ_N), and Robin (Γ_R) segments of the boundary $\partial\Omega = \Gamma_D \cup \Gamma_N \cup \Gamma_R$ of the flow domain Ω; and $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to Γ_N.

The hydraulic conductivity $K(\mathbf{x})$ and boundary functions $\Phi(\mathbf{x})$, $\psi(\mathbf{x})$ and $\varphi(\mathbf{x})$ are uncertain and treated as random fields. Specifically, $K(\mathbf{x})$ is modeled as a second-order stationary multivariate lognormal field with constant mean $\bar{K}$, variance σ_K^2, correlation length ℓ_K, and correlation function $\rho_K(r/\ell_K)$ where $r = |\mathbf{x} - \mathbf{y}|$ is the distance between any two points $\mathbf{x}, \mathbf{y} \in \Omega$. The stationarity assumption precludes the presence of distinct geological units or hydrofacies; it can be relaxed by deploying the random domain decomposition (Winter et al., 2003) that treats individual facies as stationary. The boundary functions $\Phi(\mathbf{x})$, $\psi(\mathbf{x})$ and $\varphi(\mathbf{x})$

are characterized by single-point CDFs $F_\Phi(\Phi; \mathbf{x})$, $F_\psi(\Psi; \mathbf{x})$ and $F_\varphi(Y; \mathbf{x})$, respectively, and by arbitrary spatial correlation structures. These statistical properties of the inputs can be either estimated from spatially distributed data or assigned by experts.

A solution of (1) and (2) with random $K(\mathbf{x})$, $\Phi(\mathbf{x})$, $\psi(\mathbf{x})$, and $\varphi(\mathbf{x})$ is the one-point CDF of hydraulic head, $F_h(H; \mathbf{x}) = \mathbb{P}[h(\mathbf{x}) \leq H]$. Our goal is to derive a deterministic equation satisfied by $F_h(H; \mathbf{x})$.

2.2. CDF Equation for Hydraulic Head

The main result of our study is the derivation of a $(d + 1)$-dimensional deterministic equation for the CDF $F_h(H; \mathbf{x})$ of the hydraulic head $h(\mathbf{x})$. Let us consider a functional $\Pi(H, h(\mathbf{x})) = \mathcal{H}(H - h(\mathbf{x}))$, where $\mathcal{H}(\cdot)$ is the Heaviside function and H is the coordinate in the event space for the random hydraulic head $h(\mathbf{x})$. The ensemble mean of Π over all possible values of the random variable h at any point $\mathbf{x}$ is the single-point CDF of h,

$$F_h(H; \mathbf{x}) = \langle \Pi(H, h(\mathbf{x})) \rangle. \tag{3}$$

Multiplying (1) with $-\partial\Pi/\partial H$ and accounting for the equality $\nabla\Pi = -(\partial\Pi/\partial H)\nabla h$ yields a stochastic $(d + 1)$-dimensional advection-diffusion equation for Π,

$$\nabla \cdot [K(\mathbf{x})\nabla\Pi] - K(\mathbf{x})\frac{\partial^2\Pi}{\partial H^2}\nabla h(\mathbf{x}) \cdot \nabla h(\mathbf{x}) = -g(\mathbf{x})\frac{\partial\Pi}{\partial H}. \tag{4}$$

We use the Reynolds decomposition to represent the random functions in (4) as the sum of their ensemble means and zero mean fluctuations around these means, $K = \langle K \rangle + K'$ and $\Pi = \langle \Pi \rangle + \Pi'$. The ensemble average of the resulting equation yields an unclosed equation for the CDF $F_h(H; \mathbf{x}, t)$,

$$\bar{K}\nabla^2 F_h + M = -g(\mathbf{x})\frac{\partial F_h}{\partial H}, \qquad M \equiv \nabla \cdot \langle K'(\mathbf{x})\nabla\Pi' \rangle - \langle K(\mathbf{x})\frac{\partial^2\Pi}{\partial H^2}\nabla h(\mathbf{x}) \cdot \nabla h(\mathbf{x}) \rangle. \tag{5}$$

This equation is unsolvable, since the mixed moments in the definition of M are unknown. Several approximations (closures) can be used to express these moments, which account for diffusion and dissipation of uncertainty, in terms of the known quantities. We generalize the classic IEM approach (Villermaux & Falk, 1994) by postulating a closure

$$M \approx [\alpha(\mathbf{x})(H - \bar{h}(\mathbf{x})) + \beta(\mathbf{x})]\frac{\partial F_h}{\partial H}, \tag{6}$$

where $\bar{h}$ is the mean hydraulic head and α and β are the closure variables. The IEM closure has been formulated in the context of diffusive processes, wherein it takes advantage of the fact that diffusion drives probable states to the mean. Our use of this approximation is guided by the functional similarity between (5) and the (steady-state) advection-diffusion equation. Substitution of (6) into (5) gives a closed CDF equation

$$\bar{K}\nabla^2 F_h + [\alpha(\mathbf{x})(H - \bar{h}(\mathbf{x})) + \beta(\mathbf{x}) + g(\mathbf{x})]\frac{\partial F_h}{\partial H} = 0, \qquad (\mathbf{x}, H) \in \Omega \times (H_{\min}, H_{\max}). \tag{7}$$

Empirical or phenomenological selection of the closure variables (Haworth, 2010; Pope, 2001; Raman et al., 2005) does not automatically guarantee an accurate reproduction of the first and second statistical moment of the distribution, that is, mean $\bar{h}(\mathbf{x})$ and variance $\sigma_h^2(\mathbf{x})$.

Following Boso and Tartakovsky (2016) and Boso et al. (2018), we construct the closure variables α and β in a way that ensures that the CDF equation (7) gives rise to the moment equations satisfied by $\bar{h}$ and σ_h^2. We start by recalling that if a random variable h is defined on an interval $[H_{\min}, H_{\max}]$, then the mean and variance of the CDF $F_h(H)$ are

$$\bar{h}(\mathbf{x}) = H_{\max} - \int_{H_{\min}}^{H_{\max}} F_h(H; \mathbf{x})\mathrm{d}H, \qquad \sigma_h^2(\mathbf{x}) = H_{\max}^2 - 2\int_{H_{\min}}^{H_{\max}} H F_h(H; \mathbf{x})\mathrm{d}H - \bar{h}(\mathbf{x})^2. \tag{8}$$

Hence, since $F_h(H_{\min}; \mathbf{x}) = 0$ and $F_h(H_{\max}; \mathbf{x}) = 1$, integrating (7) over H yields

$$\bar{K}\nabla^2\bar{h} - \beta(\mathbf{x}) - g(\mathbf{x}) = 0. \tag{9}$$

By the same token, multiplying both sides of (7) by H and integrating the resulting equation over H yields

$$\bar{K}\nabla^2\sigma_h^2 + 2\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - 2\alpha(\mathbf{x})\sigma_h^2 + 2\bar{h}[\bar{K}\nabla^2\bar{h} - \beta(\mathbf{x}) - g(\mathbf{x})] = 0$$

or, accounting for (9),

$$\bar{K}\nabla^2\sigma_h^2 + 2\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - 2\alpha(\mathbf{x})\sigma_h^2 = 0. \tag{10}$$

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 moment equations (Appendix A)

$$\bar{K}\nabla^2\tilde{h} + \rho(\mathbf{x}) - g(\mathbf{x}) = 0, \qquad \rho \equiv \bar{K}\lim_{\boldsymbol{\chi}\to\mathbf{x}}[\nabla_\mathbf{x}\cdot\nabla_{\boldsymbol{\chi}}C_{Yh}(\mathbf{x},\boldsymbol{\chi})] \tag{11}$$

and

$$\bar{K}\nabla^2\tilde{\sigma}_h^2 + 2V(\mathbf{x}) = 0 \tag{12a}$$

with

$$V \equiv \frac{1}{2}\bar{K}\lim_{\boldsymbol{\chi}\to\mathbf{x}}[\nabla_\mathbf{x}h^{(0)}\cdot\nabla_\mathbf{x}C_{Yh}(\mathbf{x},\boldsymbol{\chi}) - \nabla_{\boldsymbol{\chi}}\cdot\nabla_\mathbf{x}C_h(\mathbf{x},\boldsymbol{\chi})] + \left(1 + \frac{\sigma_Y^2}{2}\right)g(x)C_{Yh}(x,x). \tag{12b}$$

The moment equations are derived via perturbation expansions in the variance σ_Y^2 of log conductivity $Y(\mathbf{x}) = \ln K(\mathbf{x})$, and are accurate up to the first order in σ_Y^2. In these equations, $h^0(\mathbf{x})$ is the zeroth-order approximation of $\bar{h}(\mathbf{x})$; the mean head $\bar{h}$ is approximated with $\tilde{h} = h^{(0)} + h^{(1)} + \mathcal{O}(\sigma_Y^4)$, and the variance σ_h^2 with $\tilde{\sigma}_h^2 = [\sigma_h^2]^{(1)} + \mathcal{O}(\sigma_Y^4)$; $C_{Yh}(\mathbf{x},\chi)$ is the first-order approximation of the cross covariance $\langle Y'(\mathbf{x})h'(\chi)\rangle$; and $C_h(\mathbf{x},\chi)$ is the first-order approximation of the hydraulic head's autocovariance $\langle h'(\mathbf{x})h'(\chi)\rangle$.

Imposition of the equivalency between the mean ($\bar{h}$) and variance (σ_h^2) computed with the CDF method, (9) and (10), and the moment equations, (11) and (13), yields expressions for the closure variables $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$. Specifically, the equations for the mean, (9) and (11), are equivalent (up to the first order in σ_Y^2) if $\beta \equiv -\rho$; and the equations for the variance, (10) and (13), are the same (up to the first order in σ_Y^2) if $\alpha \equiv (\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - V)/\sigma_h^2$. These conditions yield

$$\alpha(\mathbf{x}) = \frac{\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - V}{\sigma_h^2}, \qquad \beta(\mathbf{x}) = \bar{K}\nabla^2\bar{h} + g(\mathbf{x}), \qquad V(\mathbf{x}) = -\frac{1}{2}\bar{K}\nabla^2\sigma_h^2. \tag{13}$$

These terms can be computed with various methods, including MCS. In that case, the computational advantage of using this CDF equation to compute F_h stems from the fact that it takes many fewer MC realizations to estimate $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$ than $F_h(H;\mathbf{x})$. In our implementation, we accelerate the computation further by deploying deterministic moment equations (Appendix A) to compute $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$.

The CDF equation (7) is subject to boundary conditions that reflect both possible uncertainty about the boundary functions $\Phi(\mathbf{x})$ and $\psi(\mathbf{x})$ in (2) and general properties of CDFs. To be specific, we consider $\Gamma = \Gamma_D \cup \Gamma_N$, the random boundary function $\Phi(\mathbf{x})$ to be characterized by a single-point CDF $F_\Phi(\Phi;\mathbf{x})$, and set $\psi \equiv 0$. Then (7) is subject to boundary conditions

$$F_h(H;\mathbf{x}) = F_\Phi(H;\mathbf{x}), \quad \mathbf{x}\in\Gamma_D; \qquad \nabla F_h(H;\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) = 0, \quad \mathbf{x}\in\Gamma_N. \tag{14a}$$

(As discussed in Appendix B, derivation of boundary conditions for inhomogeneous Neumann and/or Robin boundary segments is more evolved and omitted here for the sake of brevity.) If the boundary head Φ is known with certainty, that is, is deterministic, then its CDF is the Heaviside function, $F_\Phi(H;\mathbf{x}) = \mathcal{H}[H - \Phi(\mathbf{x})]$. The general property of a CDF provides the remaining boundary conditions in the H space,

$$F(H = H_{\min};\mathbf{x}) = 0, \qquad F(H = H_{\max};\mathbf{x}) = 1. \tag{14b}$$

This straightforward formulation for 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.

A plethora of efficient numerical schemes have been developed to solve linear advection-diffusion equations like (7). Since the coefficients of the CDF equation (7) are ensemble averages (e.g., $\bar{K}$), they are significantly smoother than their randomly fluctuating counterparts (e.g., K). Consequently, this equation, and the corresponding moment equations, can be solved on coarser grids than the underlying stochastic flow equation to achieve the same accuracy. We use this fact to further speed up the computations.

Our numerical solution of the boundary value problem (7)–(14) comprises two modules. The first module provides finite-volume solutions of the statistical moment equations (SME) (A4)–(A12) and yields numerical approximations of the statistical moments of head, $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$. It utilizes the research code developed by Likanapaisal et al. (2012). The second module computes the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ in (7), and solves the latter in nonconservative form by employing a finite-difference scheme.

If needed, PDF of the hydraulic head, $f_h(H; \mathbf{x})$, can be obtained either by differentiating $F_h(H; \mathbf{x})$ or by deriving a $(d + 1)$-dimensional PDF equation (Tartakovsky & Gremaud, 2016).

3. Numerical Experiments

We use two sets of numerical experiments to demonstrate the accuracy, robustness, and versatility of the proposed approach. These experiments involve mean uniform flow driven by externally imposed hydraulic head gradient and convergent flow toward a pumping well.

In both cases, the two-dimensional flow domain Ω is a square of dimensionless (normalized with the domain size L) length 1. The log hydraulic conductivity (transmissivity) $Y(\mathbf{x}) = \ln K(\mathbf{x})$ is modeled as a second-order stationary multivariate Gaussian field with zero mean ($\bar{Y} = 0$), variance σ_Y^2, an isotropic exponential covariance function $C_Y(r) = \sigma_Y^2 \exp(-r/\ell_Y)$, and dimensionless (normalized with the domain size L) correlation length ℓ_Y. The position vector $\mathbf{x} = (x_1, x_2)^\top$ and the distance $r = |\mathbf{x} - \mathbf{y}|$ between any two points $\mathbf{x}$ and $\mathbf{y}$ in the flow domain Ω are normalized with the domain size L. The flow domain boundaries $x_2 = 0$ and $x_2 = 1$, are impermeable; the deterministic (known with certainty) hydraulic heads h_{in} and h_{out} are imposed along the boundaries $x_1 = 0$ and $x_1 = 1$, respectively.

The mean uniform flow is driven by a hydraulic head gradient $J \equiv (h_{\text{out}} - h_{\text{in}})/L = 0.1$, with the dimensionless hydraulic heads $h_{\text{in}} = 1.1$ and $h_{\text{out}} = 0.1$ (normalized with the reference hydraulic head h_{ref}). The spatial domain Ω is discretized with a staggered 99×99 grid, and the number of grid points along the H coordinate is set to 55. The radial flow is induced by a pumping well located at the center of the domain, $(x_1 = 1/2, x_2 = 1/2)$, and operated at a fixed dimensionless hydraulic head of $h_{\text{well}} = 0.1$; the dimensionless hydraulic heads at the boundaries $x_1 = 0$ and $x_1 = 1$ are $h_{\text{in}} = h_{\text{out}} = 1$. In our implementation, a pumping well is represented by the source term $g(\mathbf{x}) = T_{\text{well}}(h(\mathbf{x}) - h_{\text{well}})$ in (1), where T_{well} is the prescribed well transmissibility. In this case, Ω is discretized with a 105×105 grid, and 60 grid points are used to discretize the H coordinate.

For both flow scenarios, we compare our estimates of the hydraulic head CDF $F_h(H; \mathbf{x})$ with those computed via MCS. Equiprobable MC realizations were generated by the sequential Gaussian simulator (Deutsch & Journel, 1998). Our convergence study of the exceedance probability for a given hydraulic head value H, $\mathbb{P}[h(\mathbf{x}) > H] = 1 - F_h(H; \mathbf{x})$, in the mean uniform flow case with $\sigma_Y^2 = 1$ and $\ell_Y = 0.3$, revealed that an MC estimate of $\mathbb{P}[h(\mathbf{x}) > H]$ stabilizes after about $N_{\text{MCS}} = 7,000$ MC realizations. To use the MCS estimates of $F_h(H; \mathbf{x})$ as a yardstick for ascertaining the accuracy of our CDF method for all experiments, we therefore rely on a conservative number of realizations $N_{\text{MCS}} = 10,000$.

3.1. Accuracy of the CDF Method

Since the coefficients in the CDF equation (7) are given in terms of the mean and variance of the hydraulic head $h(\mathbf{x})$, we start by analyzing the ability of the SME (A4)–(A12) to accurately approximate $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$. Figure 1 exhibits these statistical moments along the cross section $x_2 = 0.5$ for $\sigma_Y^2 = 1$ and $\ell_Y = 0.3$ in the case of mean uniform flow, and for $\sigma_Y^2 = 2.0$ and $\ell_Y = 0.2$ in the case of convergent flow. These profiles $\bar{h}(x_1, \cdot)$ and $\sigma_h^2(x_1, \cdot)$ are alternatively computed with MCS, the SME, and the CDF method.

By construction, the CDF $F_h(H; \mathbf{x})$ in (7) must have the same moments $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$ as their counterparts computed with the SME. Figure 1 reveals a slight discrepancy between these two sets of moments, as quantified by the average errors

$$\bar{\epsilon}_{\text{mean}} = \frac{1}{\|\Omega\|} \int_\Omega |\bar{h}_{\text{SME}} - \bar{h}_{\text{CDF}}| d\mathbf{x}, \qquad \bar{\epsilon}_{\text{var}} = \frac{1}{\|\Omega\|} \int_\Omega |\sigma_{h,\text{SME}}^2 - \sigma_{h,\text{CDF}}^2| d\mathbf{x}, \tag{15}$$
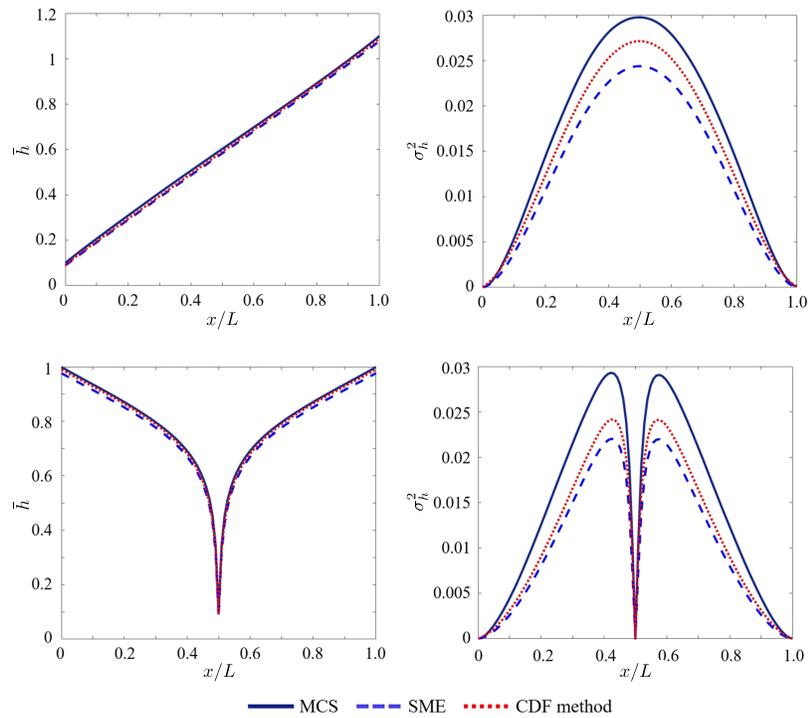
Figure 1. Mean (left column) and variance (right column) of hydraulic head, $\bar{h}(x_1, x_2 = 1/2)$ and $\sigma_h^2(x_1, x_2 = 1/2)$, for mean uniform flow (top row) and flow to a well located at the middle of the domain (bottom row). These moments are alternatively computed with Monte Carlo simulations (MCS), the statistical moment equations (SME), and the CDF method. The statistical properties of log conductivity are $\bar{Y} = 0$; and $\sigma_Y^2 = 1$ and $\ell_Y = 0.3$ in the case of mean uniform flow, and $\sigma_Y^2 = 2.0$ and $\ell_Y = 0.2$ in the case of convergent flow.

where $\|\Omega\|$ is the volume of the flow domain Ω. The errors $\bar{\epsilon}_{mean}$ and $\bar{\epsilon}_{var}$ decrease as the grid size along the H coordinate, ΔH, decreases (Figure 2). This result confirms that the discrepancy is solely due to numerical solution of the CDF equation and the subsequent evaluation of the quadratures required to compute the first two moments of a CDF.
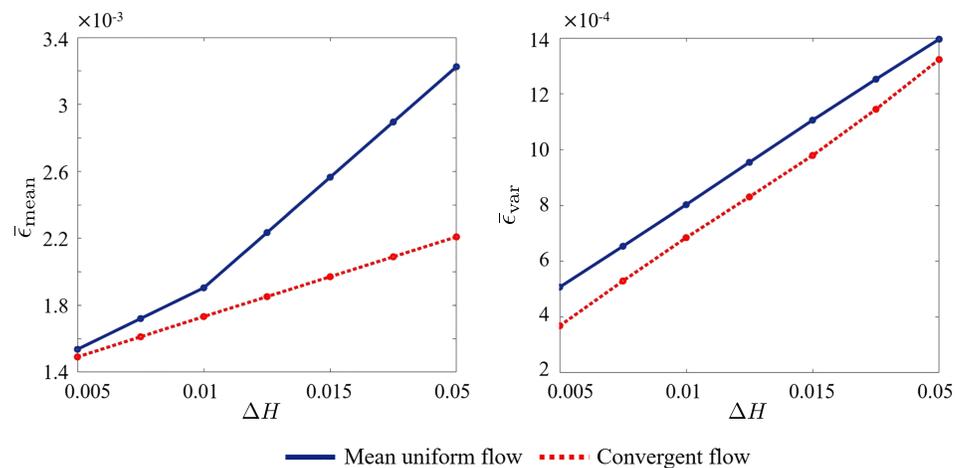
Figure 2. Average discrepancies $\bar{\epsilon}_{mean}$ (left) and $\bar{\epsilon}_{var}$ (right) between the mean and variance of hydraulic head $h(\mathbf{x})$, alternatively computed as quadratures of the CDF $F(H; \mathbf{x})$ in (7) or by solving the SME. The discrepancies decay as the grid size along the H coordinate, ΔH, becomes smaller. The statistical properties of log conductivity are $\bar{Y} = 0$; and $\sigma_Y^2 = 1$ and $\ell_Y = 0.3$ in the case of mean uniform flow, and $\sigma_Y^2 = 2.0$ and $\ell_Y = 0.2$ in the case of convergent flow.
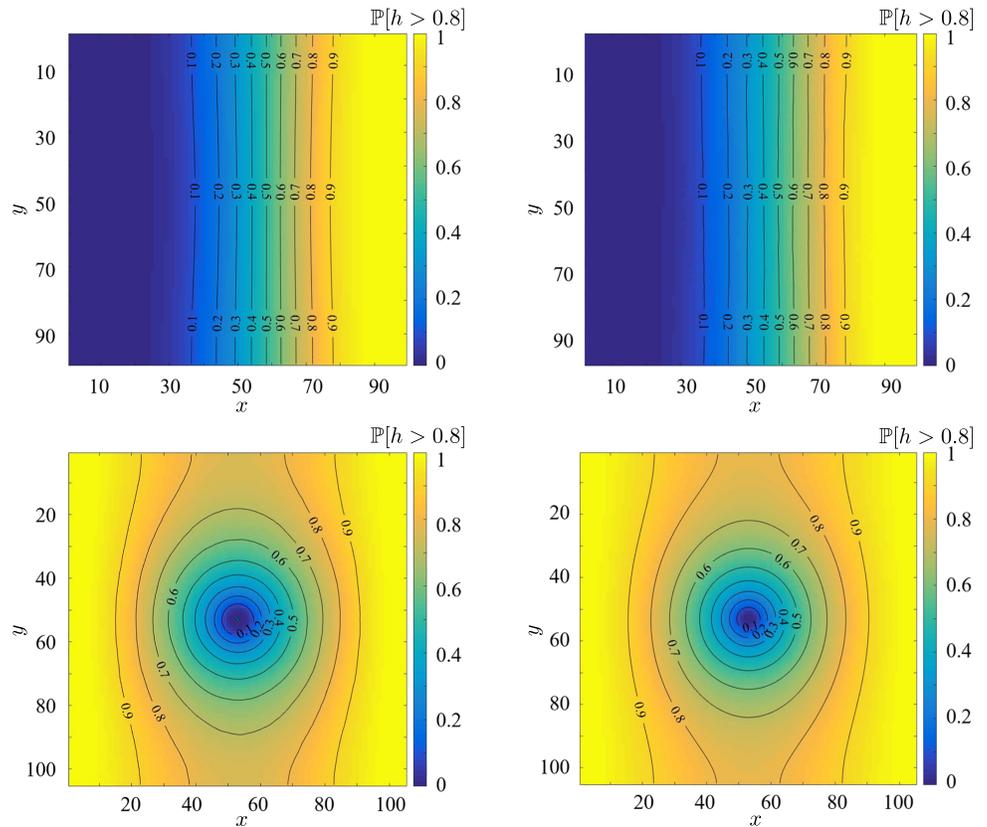
Figure 3. Spatial maps of exceedance probability $\mathbb{P}[h(\mathbf{x}) > H = 0.8] = 1 - F_h(H = 0.8; \mathbf{x})$ obtained with MCS (left column) and CDF method (right column) for mean uniform flow (top row) and convergent flow (bottom row). The statistical properties of log conductivity are $\bar{Y} = 0$; and $\sigma_Y^2 = 1$ and $\ell_Y = 0.3$ in the case of mean uniform flow, and $\sigma_Y^2 = 2.0$ and $\ell_Y = 0.2$ in the case of convergent flow.

Consistent with the previous SME-focused studies (e.g., Li et al., 2003; Neuman et al., 1996; Likanapaisal et al., 2012; Severino & De Bartolo, 2015; Tartakovsky & Neuman, 1998a, 1998b; among many others), the mean and variance of hydraulic head computed with the SME are in agreement with those inferred from MCS, regardless of the flow regime. The discrepancy between the two approaches is larger for the variance than for the mean. It also increases with the variance of log conductivity (σ_Y^2), which is used as a small perturbation parameter to derive the SME: $\sigma_Y^2 = 1$ for the mean uniform flow, and $\sigma_Y^2 = 2$ for the convergent flow.

Spatial maps of exceedance/nonexceedance probabilities ($\mathbb{P}[h(\mathbf{x}) > H] = 1 - F_h(H; \mathbf{x})$ and $\mathbb{P}[h(\mathbf{x}) \leq H] = F_h(H; \mathbf{x})$, respectively) for a selected hydraulic head threshold H are required to identify regions in the spatial domain where the corresponding risk is higher than desired. Figure 3 exhibits such maps of the probability of $h(\mathbf{x})$ exceeding $H = 0.8$ for mean uniform and convergent flows. With some degree of abstraction, these can be used to delineate the coastal regions in risk of seawater intrusion due to rising sea levels (the mean uniform flow scenario) or identify well capture zones with a prescribed level of confidence (the convergent flow scenario). These probabilities are alternatively computed with the reference MCS and as a solution of the CDF equation (7). Visual inspection of the two sets of map, as well as the CDFs $F_h(H; \mathbf{x})$ presented in Figure 4 for several points $\mathbf{x} \in \Omega$, demonstrates a close agreement between the two methods.

A more quantitative assessment of the agreement between the CDFs computed with the CDF method (F_h) and the reference MCS comprising $N_{\mathrm{MCS}} = 10000$ realizations (F_h^{MCS}) is provided by the first Wasserstein distance between two distributions (also known as Earth Mover's metric)

$$\mathcal{D}(\mathbf{x}) \equiv \int_{H_{\min}}^{H_{\max}} |F_h(H', \mathbf{x}) - F_h^{\mathrm{MCS}}(H', \mathbf{x})| \mathrm{d}H'. \tag{16}$$
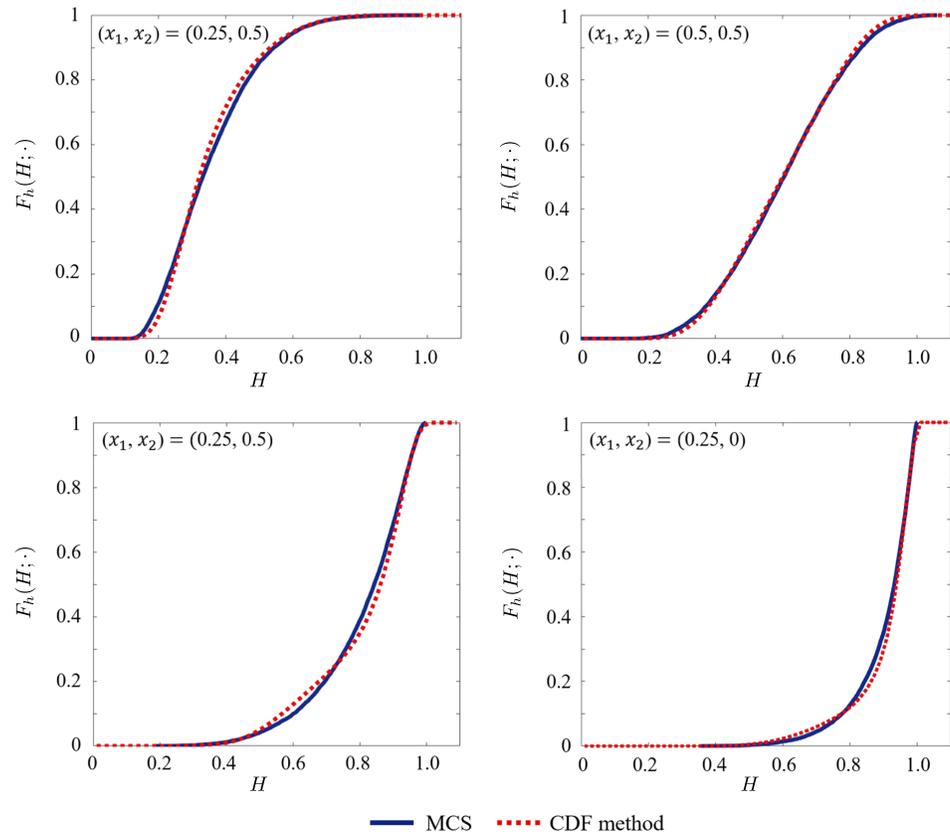
Figure 4. Hydraulic head CDFs F_h computed with MCS and the CDF method at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the flow domain for mean uniform flow (top row) and convergent flow (bottom row).

The numerical integration is carried out with the Gauss-Legendre quadrature rule. The resulting contour plots of $\mathcal{D}(\mathbf{x})$ are shown in Figure 5. The error metric $\mathcal{D}(\mathbf{x})$ is smallest close to locations where the hydraulic head h is known with certainty (the prescribed head boundaries in the case of mean uniform flow, and the prescribed head boundaries and the well in the case of convergent flow), and increase with distance from those locations. The behavior of $\mathcal{D}(\mathbf{x})$ mirrors that of the hydraulic head variance σ_h^2 and reflects the error in the perturbation-based estimation of the latter. In both flow scenarios, $\mathcal{D}(\mathbf{x})$ remains small, not exceeding 0.011 for mean uniform flow and 0.023 for convergent flow. This performance is remarkable, given relatively large values of the perturbation parameter σ_Y^2 used in these simulations ($\sigma_Y^2 = 1$ and 2 for mean uniform flow and convergent flow, respectively).
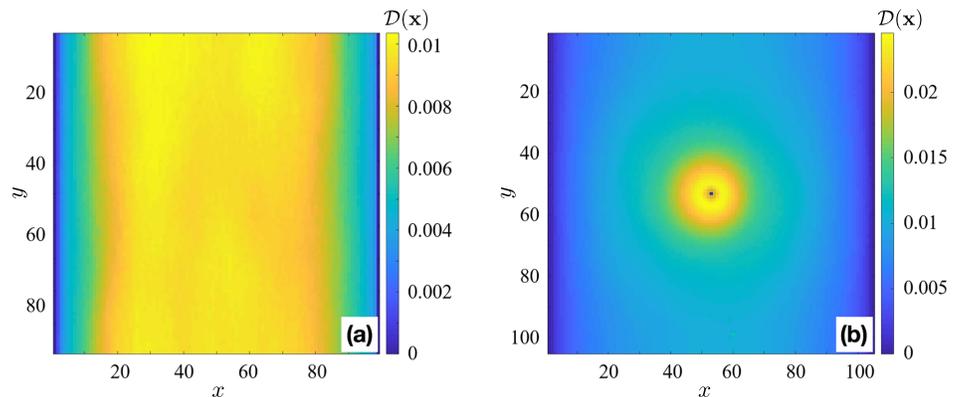
Figure 5. Spatial maps of the Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the hydraulic head CDFs computed with the CDF method and Monte Carlo simulations for mean uniform flow (left) and convergent flow (right).

Table 1
Computational Time of the CDF Method and MCS in the Case of Mean Uniform Flow

Method	Grid size	Error $\mathcal{D}_{ave}$	CPU time
CDF method	33×33	1.02×10^{-2}	4.25×10^{-1} min
	99×99	8.14×10^{-3}	1.77×10^{1} min
MCS with 1,240 realizations	99×99	8.14×10^{-3}	1.33×10^{2} min
MCS with 10,000 realizations	99×99	0	1.07×10^{3} min

3.2. Computational Efficiency of the CDF Method

As mentioned in section 1, a raison d'être for the development of the method of distributions and other uncertainty quantification techniques is the need to outperform MCS in terms of computational efficiency. While the CDF method calls for solving a $(d + 1)$-dimensional linear partial differential equation, like (7), MCS consist of repeated solves of a large number of d-dimensional (possibly nonlinear) PDEs like the flow equation (1). In addition, coefficients in the CDF and moment equations are smooth functions (ensemble averages), whereas coefficients in the original equations fluctuate randomly in space. For example, the average conductivity $\bar{K}$ in (7) is constant, even though $K(\mathbf{x})$ can vary by orders of magnitude from one cell of a numerical grid to the next. The spatial homogeneity of $\bar{K}$ not only increases the efficiency of the linear solver used to solve the SME (A4)–(A12) but also allows us to solve these equations on coarser grids without any averaging of cell properties.

The resulting computational gains provided by our CDF method are reported in Tables 1 and 2 for mean uniform flow and convergent flow, respectively. The computation times are reported for an Intel Xeon e5-2660 machine running at 2.2 GHz. The CPU comparison is carried out for the same discrepancy level, defined by the average Wasserstein distance between the CDFs computed with our method and MCS, $\mathcal{D}_{ave} = \|\Omega\|^{-1} \int_{\Omega} D(\mathbf{x})d\mathbf{x}$. Specifically, the discrepancy level $\mathcal{D}_{ave} \approx 0.01$ of the CDF method is achieved by MCS with $N_{MCS} = 1,240$ realizations in the mean uniform flow regime and with $N_{MCS} = 1,470$ realizations in the convergent flow regime. For the same discrepancy level of $\mathcal{D}_{ave} \approx 0.01$, the CDF method is about an order of magnitude faster than MCS when the same numerical grid is used. Coarsening the mesh used to solve the SME by a factor of 3 results in the similar discrepancy level while speeding up the computation by another order of magnitude.

3.3. Robustness of the CDF Method

The accuracy of the CDF method is expected to depend on the degree of uncertainty/regularity in the hydraulic conductivity, as characterized by σ_Y^2 and ℓ_Y. We perform a series of numerical experiments to analyze the robustness of the CDF method to the magnitude of these statistical parameters.

Impact of conductivity's variance. The closure approximations for the SME (A4)–(A12) are obtained via the perturbation expansion in the variance of log hydraulic conductivity σ_Y^2. Consequently, one would expect the accuracy of the CDF method to deteriorate as σ_Y^2 increases. Yet the average Wasserstein distance between our CDF solution and its MCS estimate does not appreciably change ($\mathcal{D}_{ave}$ increases by about a factor of 2) as σ_Y^2 increases from 1 to 5 (for fixed $\ell_Y = 0.1$); that is, the spatial variability of conductivity $K(\mathbf{x})$ increases by about 5 orders of magnitude (Figure 6a).

Impact of conductivity's correlation length. The correlation length ℓ_Y controls the degree of regularity of the (log) conductivity field. The dependence of $\mathcal{D}_{ave}$ on ℓ_Y (for fixed $\sigma_Y^2 = 1.0$) is shown in Figure 6b. In both flow configurations, $\mathcal{D}_{ave}$ increases with ℓ_Y as long as $\ell_Y \leq 0.7$ and decreases when $\ell_Y \geq 0.7$. The maximum values of $\mathcal{D}_{ave}$ are 0.014 and 0.0086 for mean uniform flow and convergent flow, respectively. The heterogeneous structures of the hydraulic conductivity field do not appear when the correlation length

Table 2
Computational Time of the CDF Method and MCS in the Case of Convergent Flow

Methods	Grid size	Error $\mathcal{D}_{ave}$	CPU time
CDF method	35×35	1.09×10^{-2}	5.15×10^{-1} min
	105×105	9.20×10^{-3}	2.12×10^{1} min
MCS with 1,470 realizations	105×105	9.20×10^{-3}	2.23×10^{2} min
MCS with 10,000 realizations	105×105	0	1.52×10^{3} min
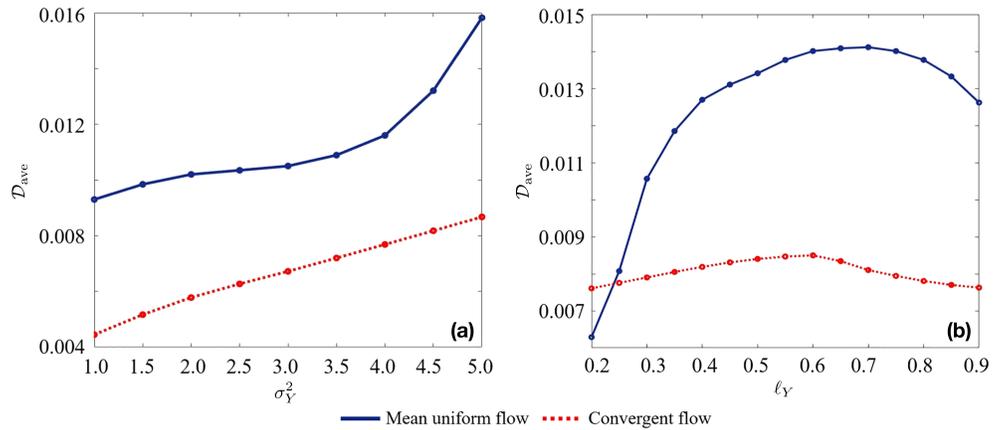
Figure 6. Dependence of the average Wasserstein distance between our CDF solution and its MCS estimate, $\mathcal{D}_{\mathrm{ave}} = \mathcal{D}_{\mathrm{ave}}(\sigma_Y^2, \ell_Y)$, on the variance ($\sigma_Y^2$ for fixed $\ell_Y = 0.1$, left) and correlation length (ℓ_Y for fixed $\sigma_Y^2 = 1.0$, right) of log hydraulic conductivity $Y = \ln K$ for mean uniform flow and convergent flow.

is extremely small. The hydraulic conductivity field becomes homogeneous when the correlation length approaches the size of a computational domain. Therefore, the heterogeneity and the approximation error of the SME and closures are maximum when the correlation length is intermediate (Li et al., 2003). Regardless, the average discrepancy $\mathcal{D}_{\mathrm{ave}}$ remains small regardless of ℓ_Y, which demonstrates that the CDF method is robust to the magnitude of ℓ_Y.

3.4. Impact of Moments' Approximation

Two types of approximations underpin the derivation of the CDF equation: the moment-preserving closure leading to (7) and the perturbation approximation used to close the moment equations. The latter affects the coefficients α and β in the CDF equation (7), which depend on the hydraulic head moments $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$. To eliminate the second source of error or, equivalently, to isolate its impact, we compare the CDFs $F_h(H; \mathbf{x})$ obtained by solving the CDF equation (7), whose coefficients are alternatively computed with the SME and the reference MCS. Since the moments computed with the reference MCS are treated as exact, their use in the CDF equation (7) isolates the impact of the moment-preserving closure.

For both flow configurations, we set $\sigma_Y^2 = 3.0$ and $\ell_Y = 0.4$. Table 3 shows the relative minor impact of the perturbation closures of the SME (A4)–(A12) on the average Wasserstein distance $\mathcal{D}_{\mathrm{ave}}$ between the CDFs resulted from the two strategies for computing the coefficients α and β. However, this integral metric of accuracy tells only part of the story. Figure 7 reveals that the CDF $F_h(H; \cdot)$ computed with the reference MCS moments is closer to the reference solution than the CDF $F_h(H; \cdot)$ computed with SME moments. This demonstrates that the performance of the CDF method relies on the accuracy of moments. It also increases confidence in the moment-preserving closure leading to the CDF equation (7).

4. Summary and Conclusions

We developed the method of distributions to probabilistically predict single-phase flow in porous media with uncertain hydraulic conductivity and/or uncertain boundary functions. The method results in a deterministic partial-differential equation for the CDF of hydraulic head. The derivation of this equation relies on a novel moment-preserving closure approximation, which expresses the coefficients of the CDF equation in terms of the mean and variance of hydraulic head. These hydraulic head statistics can be computed either

Table 3
Average Wasserstein Discrepancy $\mathcal{D}_{ave}$ Between the CDF Method With MCS Moments and the CDF Method With SME Moments for Two Flow Configurations

Methods	Error $\mathcal{D}_{\mathrm{ave}}$	
	Mean uniform flow	Convergent flow
CDF method with MCS moments	2.26×10^{-2}	1.72×10^{-2}
CDF method with SME moments	2.78×10^{-2}	2.02×10^{-2}
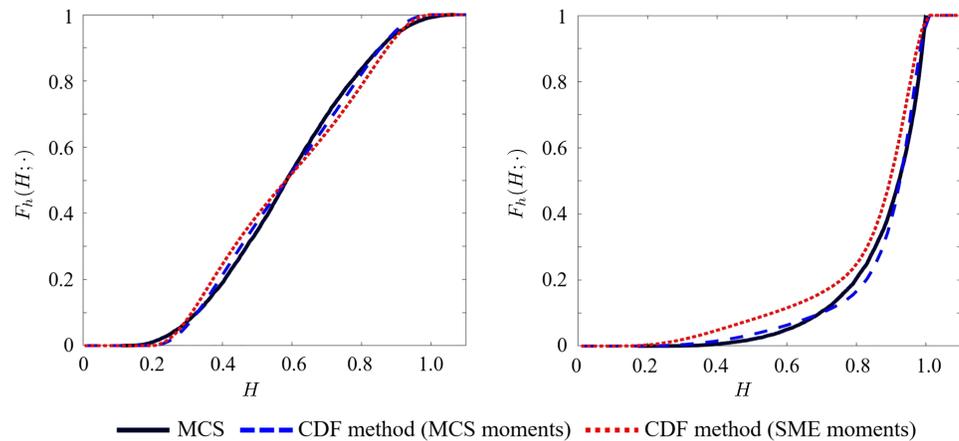
Figure 7. Hydraulic head CDFs $F_h(H; \mathbf{x})$ computed with MCS, the CDF method with MCS moments, and the CDF method SME moments at $\mathbf{x} = (0.5, 0.5)^\top$ for mean uniform flow (left) and at $\mathbf{x} = (0.5, 0.25)^\top$ for convergent flow (right).

with MCS or by solving the corresponding SMEs. The latter require an addition closure approximation, such as perturbation expansions in variance of log hydraulic conductivity. We performed a series of numerical experiments to evaluate the accuracy, robustness, and computational efficiency of the CDF method. Our study leads to the following conclusions.

- The CDF method yields spatial maps of the exceedance probability for hydraulic head. This information is required for probabilistic risk assessment, for example, for probabilistic delineation of well capture zones.
- The CDFs obtained with the CDF method are in good agreement with the reference MCS for a wide range of statistical properties of hydraulic conductivity (its variance and correlation length). The CDF equations remain robust for the conductivity variance as large as 5.
- The accuracy of the CDF method depends on the approximation of moments. Employing the exact MCS moments instead of their SME-based counterparts as an input for the CDF equation increases the accuracy of the solution.
- The CDF method is two orders of magnitude more efficient than MCS. This computational speed up stems from the smoothness of the coefficients in the SME and CDF equation, for example, from replacing randomly fluctuating hydraulic conductivity with its ensemble mean counterpart. This not only speeds up the linear solver but also facilitates the use of low-cost coarse-scale solutions.

Our CDF equation is derived by using perturbation expansions in the variance σ_Y^2 of log conductivity $Y(\mathbf{x})$. The CDF method can readily accommodate other stationary, unimodal distributions of conductivity $K(\mathbf{x})$ by using a Taylor expansion around their respective means $\bar{K}$. It can also handle nonstationary and multimodal distributions of K, which are indicative of subsurface environments composed of multiple hydrofacies. A follow-up study will deal with this setting by deploying the random domain decomposition (Winter & Tartakovsky, 2002; Winter et al., 2003), as discussed in section 2. We also plan to deploy the CDF method to characterize parametric uncertainty in realistic three-dimensional problems with complex boundary conditions and source/sink terms (e.g., injection or production wells and spatially extended sources representing recharge).

Appendix A: Derivation of Moment Equations

Derivation and analysis of the moment equations (MDEs) for the hydraulic head h have been a subject of intensive research in stochastic hydrogeology for several decades (e.g., Li et al., 2003; Neuman et al., 1996; Tartakovsky & Neuman, 1998a, 1998b; Severino & De Bartolo, 2015; among many others). A brief derivation of the MDEs implemented numerically by Likanapaisal et al. (2012) is presented below for completeness.

The steady-state groundwater flow equation (1) is rewritten in terms of log hydraulic conductivity $Y(\mathbf{x}) = \ln K(\mathbf{x})$ as

$$\nabla^2 h + \nabla Y \cdot \nabla h = g(\mathbf{x}) e^{-Y}. \tag{A1}$$

Using the Reynolds decomposition $Y(\mathbf{x}) = \bar{Y} + Y'(\mathbf{x})$, recalling that $Y(\mathbf{x})$ is second-order stationary multivariate Gaussian, that is, that its mean $\bar{Y}$ and variance $\sigma_Y^2 = \langle Y'^2 \rangle$ are constant, defining by $K_G = \exp(\bar{Y})$ the geometric mean of the hydraulic conductivity K, expanding $\exp(Y')$ into a Taylor series around $Y' = 0$, and taking the ensemble mean of the resulting equation leads to

$$\nabla^2 \bar{h} + \nabla \cdot \langle Y' \nabla h \rangle = \frac{g}{K_G} \sum_{n=0}^{\infty} \frac{1}{(2n)!} \sigma_Y^{2n}. \tag{A2}$$

Here the notation $\bar{\mathcal{A}}$ and $\langle \mathcal{A} \rangle$ is used interchangeably to denote the ensemble mean of any random quantity $\mathcal{A}$. The right-hand side is derived by taking advantage of the fact that all odd moments of a Gaussian Y' are 0. The unknown ensemble moments $\bar{h}$ and $\langle Y' \nabla h \rangle = \langle Y' \nabla h' \rangle$ are expanded into asymptotic series in the powers of σ_Y^2,

$$\bar{h} = \bar{h}^{(0)} + \bar{h}^{(1)} + \dots , \qquad \langle Y' \nabla h' \rangle = \langle Y' \nabla h' \rangle^{(1)} + \langle Y' \nabla h' \rangle^{(2)} + \dots , \tag{A3}$$

where the superscript $^{(n)}$ indicates that the corresponding quantity is of order σ_Y^{2n}. The use of these expansions formally limits the applicability of the resulting solutions to $\sigma_Y^2/2 < 1$, but has been shown to remain robust for σ_Y^2 as large as 4.

Collecting the terms of equal powers of σ_Y^{2n} in (A2) yields a recursive set of partial-differential equations

$$\nabla^2 \bar{h}^{(0)} = \frac{g}{K_G}, \qquad \nabla^2 \bar{h}^{(n)} + \nabla \cdot \langle Y' \nabla h \rangle^{(n)} = \frac{g}{2K_G} \sigma_Y^{2n}, \quad n \geq 1. \tag{A4}$$

The boundary conditions for these equations are obtained from (2) by following a similar procedure:

$$\bar{h}^{(0)} = \bar{\Phi}(\mathbf{x}), \quad \bar{h}^{(n)} = 0, \quad n \geq 1, \qquad \mathbf{x} \in \Gamma_D; \tag{A5a}$$

$$-K_G \nabla \bar{h}^{(n)} \cdot \mathbf{n}(\mathbf{x}) = \frac{1}{(2n)!} \bar{\psi}(\mathbf{x}) \sigma_Y^{2n}, \quad n \geq 0, \quad \mathbf{x} \in \Gamma_N; \tag{A5b}$$

$$-K_G \nabla \bar{h}^{(n)} \cdot \mathbf{n}(\mathbf{x}) + \frac{1}{(2n)!} a h^{(n)} \sigma_Y^{2n} = \frac{1}{(2n)!} \bar{\varphi}(\mathbf{x}) \sigma_Y^{2n}, \quad n \geq 0, \quad \mathbf{x} \in \Gamma_R. \tag{A5c}$$

The latter results rely on a reasonable assumption that the hydraulic conductivity K is not correlated with both ψ and φ.

Apart from $n = 0$, the equations in (A4) are unclosed since each of them contains two unknowns, $\bar{h}^{(n)}$ and $\langle Y' \nabla h \rangle^{(n)}$. To remediate this problem, we derive an equation for the first-order approximation of cross-correlation $C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) = \langle Y'(\boldsymbol{\chi}) h(\mathbf{x}) \rangle^{(1)}$ by multiplying (A1) with $Y'(\boldsymbol{\chi})$, taking the ensemble mean, and retaining the terms of order σ_Y^2,

$$\nabla_{\mathbf{x}}^2 C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) + \nabla_{\mathbf{x}} C_Y(\mathbf{x}, \boldsymbol{\chi}) \cdot \nabla_{\mathbf{x}} \bar{h}^{(0)} = -\frac{g}{K_G} C_Y(\mathbf{x}, \boldsymbol{\chi}), \tag{A6}$$

where $C_Y(\mathbf{x}, \boldsymbol{\chi}) = \langle Y'(\mathbf{x}) Y'(\boldsymbol{\chi}) \rangle$ is the autocorrelation of $Y(\mathbf{x})$. Accounting for the lack of correlation between Y and the boundary functions Φ and ψ, it follows from (2) that the moment equation (A6) is subject to boundary conditions

$$C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) = 0, \quad \mathbf{x} \in \Gamma_D; \qquad \nabla_{\mathbf{x}} C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \frac{\bar{\psi}}{K_G} C_Y(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_N. \tag{A7}$$

Once this boundary value problem is solved and $C_{Yh}(\boldsymbol{\chi}, \mathbf{x})$ 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 use the first-order (in σ_Y^2) approximations of the statistical

moments, that is, approximate the mean head $\bar{h}$ with $\tilde{h} = \bar{h}^{(0)} + \bar{h}^{(1)}$. Multiplying (A4) with $\bar{K}$ and summing the equations for $\bar{h}^{(0)}$ and $\bar{h}^{(1)}$ yields subject to boundary conditions

$$\tilde{h} = \bar{\Phi}(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; -K_G \nabla\tilde{h} \cdot \mathbf{n}(\mathbf{x}) = \bar{\psi}(\mathbf{x})\left(1 + \frac{\sigma_Y^2}{2}\right), \quad \mathbf{x} \in \Gamma_N;$$

$$K_G \nabla\tilde{h} \cdot \mathbf{n}(\mathbf{x}) + a\tilde{h}(\mathbf{x})\left(1 + \frac{\sigma_Y^2}{2}\right) = \bar{\varphi}(\mathbf{x})\left(1 + \frac{\sigma_Y^2}{2}\right), \quad \mathbf{x} \in \Gamma_R. \tag{A8}$$

An equation for the first-order approximation of the head variance, $\tilde{\sigma}_h^2$, is derived by subtracting (11) from (A4), multiplying the resulting equation with $h'(\mathbf{x})$, and taking the ensemble mean:

$$\nabla_{\mathbf{x}}^2 \sigma_h^2(\mathbf{x}) - 2\langle \nabla h \cdot \nabla h' \rangle^{(1)} + 2\nabla_{\mathbf{x}}\bar{h}^{(0)} \cdot \langle \nabla Y' h \rangle^{(1)} = -\frac{2g(\mathbf{x})}{K_G}C_{Yh}(\mathbf{x}, \mathbf{x}). \tag{A9}$$

Similar to the equation for $\tilde{h}$, we compute $\nabla_{\mathbf{x}}C_{Yh}(\mathbf{x}, \boldsymbol{\chi})$ to evaluate $\langle \nabla Y' h \rangle^{(1)} = \lim_{\boldsymbol{\chi} \to \mathbf{x}}[\nabla_{\mathbf{x}}C_{Yh}(\mathbf{x}, \boldsymbol{\chi})]$. To obtain the workable expression for the unknown term $\langle \nabla h \cdot \nabla h' \rangle^{(1)}$, we solve the equation for the first-order approximation of the hydraulic head's autocovariance function, $C_h(\mathbf{x}, \boldsymbol{\chi}) = \langle h(\mathbf{x})h'(\boldsymbol{\chi}) \rangle^{(1)}$. The equation for $C_h(\mathbf{x}, \boldsymbol{\chi})$ is derived by multiplying (A1) with $h'(\boldsymbol{\chi})$, taking the ensemble mean, and retaining the terms of order σ_Y^2,

$$\nabla_{\mathbf{x}}^2 C_h(\mathbf{x}, \boldsymbol{\chi}) + \nabla_{\mathbf{x}}C_{Yh}(\mathbf{x}, \boldsymbol{\chi}) \cdot \nabla_{\mathbf{x}}\bar{h}^{(0)} = -\frac{g(\mathbf{x})}{K_G}C_{Yh}(\mathbf{x}, \boldsymbol{\chi}). \tag{A10}$$

The boundary conditions for this equation are obtained by multiplying (2) with $h'(\boldsymbol{\chi})$, taking the ensemble average, and retaining the terms of order σ_Y^2,

$$C_h(\mathbf{x}, \boldsymbol{\chi}) = C_{\Phi h}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_D; \qquad K_G \nabla_{\mathbf{x}}C_h(\mathbf{x}, \boldsymbol{\chi}) \cdot \mathbf{n}(\mathbf{x}) = C_{\psi h}(\mathbf{x}, \boldsymbol{\chi}) - \bar{\psi}(\mathbf{x})C_{Yh}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_N;$$

$$K_G \nabla_{\mathbf{x}}C_h(\mathbf{x}, \boldsymbol{\chi}) \cdot \mathbf{n}(\mathbf{x}) - aC_h(\mathbf{x}, \boldsymbol{\chi}) = C_{\varphi h}(\mathbf{x}, \boldsymbol{\chi}) - [\bar{\varphi}(\mathbf{x}) - a\tilde{h}(\mathbf{x})]C_{Yh}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_R. \tag{A11}$$

The boundary cross covariances $C_{\Phi h}(\mathbf{x}, \boldsymbol{\chi})$, $C_{\psi h}(\mathbf{x}, \boldsymbol{\chi})$ and $C_{\varphi h}(\mathbf{x}, \boldsymbol{\chi})$ are computed by multiplying (2) with $h'(\boldsymbol{\chi})$ and taking the ensemble average. If the boundary functions Φ and ψ are deterministic, as is the case in our numerical experiments, then $C_{\Phi h}(\mathbf{x}, \boldsymbol{\chi}) = 0$ and $C_{\psi h}(\mathbf{x}, \boldsymbol{\chi}) = 0$.

Once this boundary value problem is solved, that is, $C_h(\mathbf{x}, \boldsymbol{\chi})$ is computed, we evaluate $\langle \nabla h \cdot \nabla h' \rangle^{(1)} = \lim_{\boldsymbol{\chi} \to \mathbf{x}}[\nabla_{\mathbf{x}} \cdot \nabla_{\boldsymbol{\chi}}C_h(\boldsymbol{\chi}, \mathbf{x})]$. Multiplying (A9) with $\bar{K}$ and evaluating $\langle \nabla h \cdot \nabla h' \rangle^{(1)}$ and $\langle \nabla h \cdot \nabla h' \rangle^{(1)}$ leads to the closed equations (12) for the first-order approximation of the head variance subject to boundary conditions

$$\tilde{\sigma}_h^2(\mathbf{x}) = C_{\Phi h}(\mathbf{x}, \mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad K_G \nabla_{\mathbf{x}}\tilde{\sigma}_h^2(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = 2C_{\psi h}(\mathbf{x}, \mathbf{x}) - 2\bar{\psi}(\mathbf{x})C_{Yh}(\mathbf{x}, \mathbf{x}), \quad \mathbf{x} \in \Gamma_N;$$

$$K_G \nabla_{\mathbf{x}}\tilde{\sigma}_h^2(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) - a\tilde{\sigma}_h^2(\mathbf{x}) = 2C_{\varphi h}(\mathbf{x}, \boldsymbol{\chi}) - 2[\bar{\varphi}(\mathbf{x}) - a\tilde{h}(\mathbf{x})]C_{Yh}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_R. \tag{A12}$$

Alternatively, $\tilde{\sigma}_h^2$ can be obtained by taking the limit of the head's autocovariance function $C_h(\boldsymbol{\chi}, \mathbf{x})$, that is, $\tilde{\sigma}_h^2 = \lim_{\boldsymbol{\chi} \to \mathbf{x}}C_h(\boldsymbol{\chi}, \mathbf{x})$. The limit can be computed from the numerical solution of (A10), C_h, between the grid point $\mathbf{x}$ and all grid points in the domain.

Appendix B: Boundary Conditions for the CDF Equation

Boundary conditions for the CDF equation along the physical boundaries Γ_N and Γ_R are obtained from (2) in three steps. We show here the derivation for mixed type boundary conditions along Γ_R; conditions along Γ_N are identical by imposing $a = 0$ and substituting φ with ψ. First, we multiply (2) along Γ_R by $\partial\Pi/\partial H$ to obtain

$$-K(\mathbf{x})\nabla\Pi \cdot \mathbf{n}(\mathbf{x}) - aH\frac{\partial\Pi}{\partial H} = -\varphi\frac{\partial\Pi}{\partial H}. \tag{B1}$$

Ensemble averaging of (B1) yields

$$-\bar{K}(\mathbf{x})\nabla F \cdot \mathbf{n}(\mathbf{x}) = aH\frac{\partial F}{\partial H} - \bar{\varphi}\frac{\partial F}{\partial H} + \langle K'\frac{\partial\Pi'}{\partial H}\rangle - \langle \varphi'\frac{\partial\Pi'}{\partial H}\rangle \tag{B2}$$

which requires closure. Consistently with the IEM closure developed for (1), we impose

$$-\bar{K}(\mathbf{x})\nabla F \cdot \mathbf{n}(\mathbf{x}) = aH\frac{\partial F}{\partial H} - \bar{\varphi}\frac{\partial F}{\partial H} + \left(\Gamma(\mathbf{x})(H - \bar{h}(\mathbf{x})) + \eta(\mathbf{x})\right)\frac{\partial F}{\partial H}, \quad \mathbf{x} \in \Gamma_R, \tag{B3}$$

where $\Gamma(\mathbf{x})$ and $\eta(\mathbf{x})$ are required to guarantee consistency with the boundary conditions for the moment equation (Appendix A). Upon integration, this yields

$$\Gamma(\mathbf{x}) = \frac{(\bar{K}/2)\nabla\sigma_h^2 \cdot \mathbf{n}(\mathbf{x}) - a\sigma_h^2(\mathbf{x}) + 2a\bar{h}^2(\mathbf{x})}{\sigma_h^2(\mathbf{x}) - 2\bar{h}^2(\mathbf{x})}, \qquad \eta(\mathbf{x}) = \bar{K}\nabla\bar{h} \cdot \mathbf{n}(\mathbf{x}) - a\bar{h}(\mathbf{x}) + \bar{\varphi}(\mathbf{x}). \tag{B4}$$

References

Barajas-Solano, D. A., & Tartakovsky, D. M. (2016). Stochastic collocation methods for nonlinear parabolic equations with random coefficients. *SIAM / ASA Journal on Uncertainty Quantification*, *4*(1), 475–494.

Boso, F., Broyda, S. V., & Tartakovsky, D. M. (2014). Cumulative distribution function solutions of advection-reaction equations with uncertain parameters. *Proceedings of the Royal Society A*, *470*(2166), 20140189. https://doi.org/10.1098/rspa.2014.0189

Boso, F., Marzadri, A., & Tartakovsky, D. M. (2018). Probabilistic forecasting of nitrogen dynamics in hyporheic zone. *Water Resources Research*, *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 Resources Research*, *52*, 4700–4712. https://doi.org/10.1002/2016WR018745

Caflisch, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, *7*, 1–49.

Ciriello, V., Lauriola, I., Bonvicini, S., Cozzani, V., Federico, V. D., & Tartakovsky, D. M. (2017). Impact of hydrogeological uncertainty on estimation of environmental risks posed by hydrocarbon transportation networks. *Water Resources Research*, *53*, 8686–8697.

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

Deutsch, C., & Journel, A. (1998). *GSLIB: Geostatistics Software Library and User's Guide*. New York: Oxford University Press.

Dodwell, T. J., Ketelsen, C., Scheichl, R., & Teckentrup, A. L. (2015). A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM-ASA Journal on Uncertainty Quantification*, *3*, 1075–1108.

Giles, M. B., Nagapetyan, T., & Ritter, K. (2015). Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM / ASA Journal on Uncertainty Quantification*, *3*, 267–295.

Haworth, D. C. (2010). Progress in probability density function methods for turbulent reacting flows. *Progress in Energy and Combustion Science*, *36*(2), 168–259.

Ibrahima, F., Meyer, D. W., & Tchelepi, H. A. (2015). Distribution functions of saturation for stochastic nonlinear two-phase flow. *Transport in Porous Media*, *109*(1), 81–107.

Ibrahima, F., Tchelepi, H. A., & Meyer, D. W. (2018). An efficient distribution method for nonlinear two-phase flow in highly heterogeneous multidimensional stochastic porous media. *Computers & Geosciences*, *22*, 389–412.

Li, L., Tchelepi, H. A., & Zhang, D. (2003). Perturbation-based moment equation approach for flow in heterogeneous porous media: Applicability range and analysis of high-order terms. *Journal of Computational Physics*, *108*, 296–317.

Lichtner, P. C., & Tartakovsky, D. (2003). Stochastic analysis of effective rate constant for heterogeneous reactions. *Stochastic Environmental Research and Risk Assessment*, *17*(6), 419–429.

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

Liodakis, S., Kyriakidis, P., & Gaganis, P. (2018). Conditional Latin hypercube simulation of (log)Gaussian random fields. *Mathematical Geosciences*, *50*, 127–146.

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 Resources Research*, *32*(5), 1479–1480.

Pope, S. B. (2001). Turbulent flows.

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

Severino, G., & De Bartolo, S. (2015). Stochastic analysis of steady seepage underneath a water-retaining wall through highly anisotropic porous media. *Journal of Fluid Mechanics*, *778*, 253–272.

Shvidler, M., & Karasaki, K. (2003). Probability density functions for solute transport in random field. *Transport in porous media*, *50*(3), 243–266.

Tartakovsky, D. M. (2007). Probabilistic risk analysis in subsurface hydrology. *Geophysical Research Letters*, *34*, L05404. https://doi.org/10.1029/2007GL029245

Tartakovsky, D. M. (2013). Assessment and management of risk in subsurface hydrology: A review and perspective. *Advances in Water Resources*, *51*, 247–260. https://doi.org/10.1016/j.advwatres.2012.04.007

Tartakovsky, D. M., & Broyda, S. (2011). Pdf equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of contaminant hydrology*, *120*, 129–140.

Tartakovsky, D. M., & Gremaud, P. A. (2016). Method of distributions for uncertainty quantification, *Handbook of uncertainty quantification* pp. 1–22): Springer, Cham.

Tartakovsky, D. M., & Neuman, S. P. (1998a). Transient flow in bounded randomly heterogeneous domains 1. Exact conditional moment equations and recursive approximations. *Water Resources Research*, *34*(1), 1–12.

Tartakovsky, D. M., & Neuman, S. P. (1998b). Transient flow in bounded randomly heterogeneous domains 2. Localization of conditional mean equations and temporal nonlocality effects. *Water Resources Research*, *34*(1), 13–20.

Villermaux, J., & Falk, L. (1994). A generalized mixing model for initial contacting of reactive fluids. *Chemical Engineering Science*, *49*(24), 5127–5140.

Wang, P., Tartakovsky, D. M., Jarman, J. K. D., & Tartakovsky, A. M. (2013). CDF solutions of Buckley-Leverett equation with uncertain parameters. *Multiscale Modeling and Simulation*, *11*(1), 118–133. https://doi.org/10.1137/120865574

Winter, C. L., & Tartakovsky, D. M. (2002). Groundwater flow in heterogeneous composite aquifers. *Water Resources Research*, *38*(8), 1148. https://doi.org/10.1029/2001WR000450

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

Xiu, D. (2010). *Numerical methods for stochastic computations: A spectral method approach*. New York: Princeton University Press.