DISTRIBUTION-BASED FRAMEWORK FOR UNCERTAINTY PROPAGATION IN HETEROGENEOUS POROUS MEDIA

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ENERGY RESOURCES ENGINERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Hyung Jun Yang
March 2022

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Daniel Tartakovsky) Principal Co-Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Hamdi Tchelepi) Principal Co-Advisor

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Francesca Boso)

Approved for the Stanford University Committee on Graduate Studies

Abstract

Quantitative predictions of fluid flow and transport in porous media are often compromised by multi-scale heterogeneity and insufficient site characterization. These factors introduce uncertainty on input and output of physical systems which are generally expressed as partial differential equations (PDEs). The characterization of this predictive uncertainty is typically done with forward propagation of input uncertainty as well as inverse modeling for the dynamic data integration. The main challenges of forward uncertainty propagation arise from the slow convergence of Monte Carlo Simulations (MCS) especially when the goal is to compute the probability distribution which is necessary for risk assessment and decision making under uncertainty. On the other hand, reliable inverse modeling is often hampered by ill-posedness of the problem, thus the incorporation of geological constraints becomes increasingly important. In the thesis, four significant contributions are made to alleviate these outstanding issues underlying on forward and inverse problems.

First, the method of distributions for steady state flow problem is developed to yield a full probabilistic description of outputs via probability distribution function (PDF) or cumulative distribution (CDF). The derivation of deterministic equation for CDF relies on stochastic averaging techniques and self-consistent closure approximation which ensures the resulting CDF has the same mean and variance as those computed with moment equations or MCS. 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 proposed 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 with a required degree of confidence.

The second contribution of this work is the extension of distribution-based method to account for uncertainty in the geologic makeup of a subsurface environment and non-stationary cases. Our CDF-RDD framework provides probabilistic assessment of uncertainty in highly heterogeneous subsurface formations by combining the method of distributions and the random domain decomposition (RDD). Our numerical experiments reveal that the CDF-RDD remains accurate for two-dimensional flow in a porous material composed of two heterogeneous geofacies, a setting in which the original distribution method fails. For the same accuracy, the CDF-RDD is an order of magnitude faster than MCS.

Next, we develop a complete distribution-based method for the probabilistic forecast of two-phase flow in porous media. The CDF equation for travel time is derived within the efficient streamline-based framework to replacing the MCS in previous FROST method. For getting fast and stable results, we employ the numerical techniques including pseudo-time integration, flux-limited scheme, and exponential grid spacing. Our CDF-FROST framework uses the results of method of distributions for travel time as an input of FROST method. The proposed method provides probability distribution of saturation without any assist of sampling-based methods. The numerical tests demonstrate that the CDF-FROST shows good accuracy in estimating the probability distributions of both saturation and travel time. For the same accuracy, it is about 5 and 10 times faster than previous FROST method and naive MCS, respectively.

Lastly, we propose a consensus equilibrium (CE) framework to reconstruct the realistic geological model by the inverse modeling of sparse dynamic data. The optimization-based inversion techniques are integrated with recent machine learning-based methods (e.g., variational auto-encoder and convolutional neural network) by the proposed CE algorithm to capture the complicated geological features. The numerical examples verify that the proposed method well preserves the geological realism, and it efficiently quantifies the uncertainty conditioned on dynamic information.

Acknowledgments

Foremost, I would like to thank to my advisors Daniel Tartakovsky and Hamdi Tchelepi. Daniel set an example of great advisor and principal investigator. He encouraged me to explore many different ideas. His optimistic spirit and dedicative support make me to finish this long research journey. Under his supervision, I could truly grow as researcher. I am equally grateful that I was advised by Hamdi. Hamdi gave a lot of freedom to explore a wide variety of research problems. His thoughtful advice has always leaded my research on the right path and motivated me a lot. Thanks to his guidance, I could learn how to think and solve problems as researcher as well as engineer.

I would like to express sincere thanks to Dr. Francesca Boso for many great support and serving for reading committee. Large part of my research is inspired by her great suggestions, discussions, and mathematical background. It was a great honor for me to work with Francesca. I would also like to thank to Dr. Brendt Wohlberg at Los Alamos National Laboratory. The last chapter of this thesis could be completed under his supervision. Brendt gave me a lot of fruitful suggestions especially related to machine learning and computer vision techniques while I was working at Los Alamos and even after. I feel fortunate to work with him and I could grow more from the collaboration with him.

Next, I am so grateful to my committee members Professor Eric Darve, his productive comments on this work, and Professor Hai Wang for being chair during my defense. Thanks are also due to all ERE faculty and staff for their great support during my PhD. I'm especially grateful for all the members of Daniel's research group (Data-driven Modeling and Simulation), Dong Song, Hannah Lu, Dr. Zitong Zhou, Dr. Søren Taverniers, Dr. Arnout Boelens, Weiyu Li, Lama El Halabi, and Livia Fulchignoni. I appreciate for your comments, discussions, and a lot of great dinners during my PhD. I also thank SUPRI-B group and all members. A lot of great seminars and discussions helped to move forward.

Now, I would like to thank to my Korean friends in ERE and School of Earth, Timothy Yeo, Jaewoo An, Yongdo Kim, YOngsoo Park, and Jaehong Jung. Your support and meaningful friendship enriched my PhD life. I am thankful to all KBSK members who helped me a lot physically and mentally since I arrived Stanford.

Most importantly, I would thank my family - my parents Jinhee and Hyesuk and my younger

brother Kijun. They gave me endless love and encouragement to pursue my dream and goals. I feel so lucky to have these thoughtful and dedicated parents, and my journey at Stanford was all possible due to them. Above all, I appreciate you, heavenly father.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Quantitative prediction of fluid flow and transport in subsurface porous media is a key component of proper risk assessment and decision-making in many applications including water resources management, extraction of fossil fuels, geologic carbon sequestration, and contaminant management. Given a proper description of the physical properties, the dynamic processes in subsurface formations can be predicted using the numerical solution of governing equations derived from fundamental conservation laws and Darcy's law [6]. However, the reliable characterizations of subsurface properties are notoriously elusive due to insufficient site data. The available site information including rock and fluid samples and geology information from seismic data or outcrops of the formation (see Figure 1.1) often shows a very sparse sampling density in the subsurface area of interest. Moreover, the problem is further complicated by multi-scale heterogeneity of physical properties (see Figure 1.2).

The combination of high spatial variability and scarce site data introduces uncertainty in input parameters (e.g., permeability and porosity) and forcings (e.g., initial and boundary conditions and production rate), rendering model outputs uncertain as well. In the last several decades, many predictive models have been developed to quantify this uncertainty within the probabilistic framework, which equates uncertainty with randomness. Two main research areas actively being pursed for the characterization of uncertainty are forward propagation of input uncertainty (forward problem) and inversion for conditioning on observed dynamic data (inverse problem). This thesis makes a contribution to both of these two problems.

In this introduction, we first formulate the flow and transport problems in porous media that

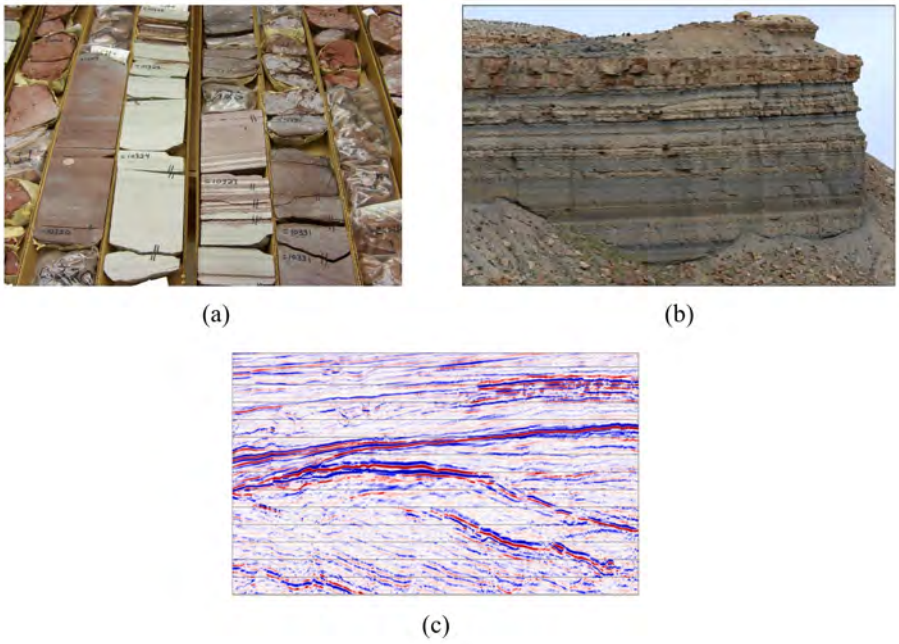

Figure 1.1: Various sources of site information: (a) core sample [100], (b) outcrop data [15], and (c) seismic data [69].

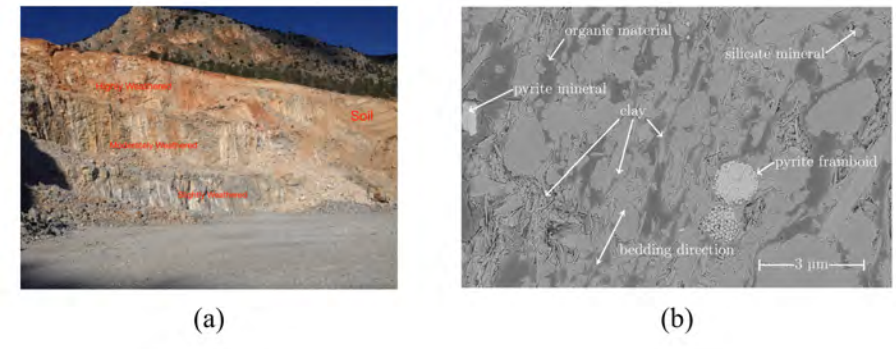

Figure 1.2: Multi-scale heterogeneity of subsurface formation: (a) large-scale [77], (b) small-scale[14].

will be considered through the thesis. Then, the brief description of forward and inverse problems and their main challenges are presented. Finally, we present the thesis outline with the main ideas and contributions in each of the following chapters.

1.2 Flow and transport in porous media

In this thesis, we study two different problems. First, we consider the single phase flow problem in heterogeneous porous media. The QoI of this flow problem is hydraulic head h or pressure p. Second, the two-phase flow in porous media would be considered. The QoI of this two-phase problem is pressure p as well as saturation S.

1.2.1 Single phase flow in porous media

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

$$\nabla \cdot [K(\mathbf{x})\nabla h(\mathbf{x})] = f(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{1.1}$$

subject to boundary conditions

$$h(\mathbf{x}) = \phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \quad \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N; \tag{1.2}$$

Here $f(\mathbf{x})$ represents point and/or distributed sources and sinks; $\phi(\mathbf{x})$ and $\psi(\mathbf{x})$ are the hydraulic head and the normal component of the Darcy flux $\mathbf{q}(\mathbf{x}) = -K(\mathbf{x})\nabla h(\mathbf{x})$, on the Dirichlet ($\Gamma_D$) and Neumann ($\Gamma_N$) of the flow domain Ω; $K(\mathbf{x})$ is a hydraulic conductivity; and $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to Γ_N. (1.1) can be rewritten in terms of pressure $p = \rho g h$ and permeability $k = (K\mu)/(\rho g)$ as following:

$$\nabla \cdot [\frac{k(\mathbf{x})}{\mu}\nabla p(\mathbf{x})] = f(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{1.3}$$

where μ is viscosity, ρ is fluid density, and g is the accelaration of gravity. When fluid or rock is compressible, flow equation is changed to transient equation as

$$S_s \frac{\partial h}{\partial t} = \nabla \cdot (K\nabla h) - f(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \quad t > 0, \tag{1.4}$$

subject to the boundary condition (1.2).

1.2.2 Two-phase flow in porous media

Next, we consider incompressible immiscible displacement of non-wetting phase (e.g., oil) by wetting phase (e.g., water) in a d-dimensional heterogeneous porous medium $\Omega \subset \mathbb{R}^d$. Neglecting capillary pressure and gravity, the Darcy fluxes for water $\mathbf{q}_w$ and non-wetting phase $\mathbf{q}_{nw}$ are described by

$$\mathbf{q_w} = -\lambda_w \nabla p, \qquad \mathbf{q}_{nw} = -\lambda_{nw} \nabla p. \tag{1.5}$$

where p represents the pressure, and λ_w and λ_{nw}, mobility of each phases, are given by

$$\lambda_w = \frac{K(\mathbf{x}) k_{rw}(S_w)}{\mu_w}, \qquad \lambda_{nw} = \frac{K(\mathbf{x}) k_{rnw}(S_{nw})}{\mu_{nw}} \tag{1.6}$$

Here, μ_{nw} and μ_w are viscosities of each phases; $k_{rw}(S_w)$ and $k_{rnw}(S_{nw})$ are relative permeability of wetting and non-wetting phases; and $K(\mathbf{x})$ is absolute permeability.

Since the total Darcy flux $\mathbf{q}_{tot} = \mathbf{q}_w + \mathbf{q}_{nw}$ satisfies the continuity condition, the steady-state flow equation can be written as

$$-\nabla \cdot (\lambda_{tot} \nabla p) = g(\mathbf{x}). \tag{1.7}$$

subject to boundary conditions

$$p(\mathbf{x}) = \phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad \mathbf{q}_{tot}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N, \tag{1.8}$$

where $\lambda_{tot} = \lambda_w + \lambda_{nw}$ is the total mobility. It can be easily shown that (1.7) is analogous to (1.3) when the porous media is fully saturated (i.e., $S_w = 1$ or $S_{nw} = 1$).

With the condition $S_w + S_{nw} = 1$, the conservation of mass can be reduced to following saturation equation for the wetting phase,

$$\phi(\mathbf{x}) \frac{\partial S_w}{\partial t} + \mathbf{q}_{tot} \cdot \nabla f_w(S_w) = 0. \tag{1.9}$$

subject to initial and boundary conditions

$$S_w(\mathbf{x}, t) = 1 - S_{nwi}, \quad \mathbf{x} \in \Gamma_{inj}, \quad t > 0; \qquad S_w(\mathbf{x}, 0) = 1 - S_{wi}, \quad \mathbf{x} \in \Omega. \tag{1.10}$$

where Γ_{inj} is the boundary of the domain where the wetting phase is injected from the injector, and S_{wi} and S_{nwi} are irreducible saturation of water and oil, respectively.

1.3 Forward uncertainty propagation

The heterogeneous nature of the porous media and incomplete knowledge about their properties, facilitates modeling of the parameters as random fields [27]. When the permeability or other input properties, the coefficients of governing equations in Section 1.2, are also treated as random fields, then the equations become stochastic PDEs whose solution is probabilistic. The objective of the forward uncertainty propagation is to quantify the impact of the input uncertainties on model outputs. To formulate the uncertainty propagation more specifically, we assume uncertain input η follows a known probability distribution $p(\eta)$. Let $Q(\mathbf{x}, t, \eta)$ be our quantity of interest (QoI) which is the function of space $\mathbf{x}$, time t, and random input η. The Q therefore will be random as well. Then, the uncertainty propagation problem is to characterize the statistical properties of the output QoI $Q(\mathbf{x}, t, \eta)$, like mean $\bar{Q}$ or variance σ_Q^2,

$$\bar{Q} = \int Q(\mathbf{x}, t, \eta) p(\eta) d\eta, \tag{1.11}$$

$$\sigma_Q^2 = \int (Q - \bar{Q})^2 p(\eta) d\eta. \tag{1.12}$$

However, the full probabilistic descriptions of the output i.e. the probability density function (PDF) f_Q, or cumulative distribution function (CDF) F_Q are often desired for the probabilistic risk assessment and decision making which typically requires the estimates of probability of rare events. For instance, estimating the CDF $F_Q(q, \mathbf{x}, t)$ is equivalent to the prediction of how likely the model output $Q(\mathbf{x}, t)$ at the location $\mathbf{x}$ and t stays below a safe limit q. The PDF $f_Q(q, \mathbf{x}, t)$ and CDF $F_Q(q, \mathbf{x}, t)$ for the given value q are formulated as

$$f_Q(q, \mathbf{x}, t) = \int \delta(q - Q) p(\eta) d\eta, \tag{1.13}$$

$$F_Q(q, \mathbf{x}, t) = \int \mathcal{H}(q - Q) p(\eta) d\eta, \tag{1.14}$$

where $\delta(\cdot)$ and $\mathcal{H}(\cdot)$ are delta and Heaviside functions.

Generally, multiple alternative scenarios have to be explored within the probabilistic framework to obtain PDF/CDF, typically by means of Monte Carlo simulations (MCS). These can be computationally expensive, and often prohibitively so, especially when the goal is to accurately compute the tails of a distribution. Among a plethora of methods aiming for uncertainty quantification, the method of distributions [96] provides such information by solving the deterministic equations for

PDF/CDF of the QoI. The computational efficiency of this method over other approaches has been proved in previous works [45, 118].

1.4 Inverse problem

Inversion techniques (aka history matching or data assimilation), which minimize the mismatch between observations and model predictions, have become an essential tool of uncertainty characterization. For Gaussian subsurface model, application of the Bayesian framework leads the inverse problem to the minimization of maximum-a-posteriori (MAP) estimates defined as follows:

$$S(\mathbf{m}) = \frac{1}{2}(d - d^*)^T C_d^{-1}(d - d^*) + \frac{1}{2}(m - m^*)^T C_m^{-1}(m - m^*), \tag{1.15}$$

where S is the misfit function we seek to minimize, C_m is the spatial covariance matrix of the prior model $\mathbf{d}$ represents the predicted dynamic data obtained by solving flow and transport problems in Section 1.2 with geological model $\mathbf{m}$, and $\mathbf{d}^*$ designates the measured dynamic data. The measurements errors associated with $\mathbf{d}^*$ are often assumed to be Gaussian, with zero mean and covariance matrix C_d. For the MAP estimate, we set $\mathbf{m}^* = \bar{m}$ (the prior mean) and $\mathbf{d}^* = \mathbf{d}_{obs}$, where $\mathbf{d}_{obs}$ is the observed dynamic data. The MAP estimate does not solely assess the uncertainty. One of the most widely used approaches for uncertainty quantification is the randomized maximum likelihood (RML) which minimizes (1.15) repeatedly.

The minimization of (1.15) provides good estimates of uncertainty when it is well-posed problem and the geological model is stationary Gaussian field. However, it is well known that subsurface inversion is ill-posed problem with non-unique solutions, and additional prior information in form of geological constraints is required to obtain realistic subsurface models that have good predictive capability [17]. In other words, the resulting model of inversion process should be consistent to the prior geostatistical description.

Most optimization-based approaches fail to incorporate various geological priors and capture the geological constraints. One alternative is to employ a parameterization procedure aiming to represent geological maps in terms of a small number of parameters. Methods of this class include principal component analysis [80, 104] and deep learning-based techniques [20, 53, 52, 64]. Though these approaches have shown good performance in many subsurface applications, their results vary with the subjectively defined number of parameters and parameterization methods, i.e., require significant fine-tuning.

1.5 Thesis outline

In Chapter 2, We deploy the method of distributions to derive a deterministic equation for the CDF of hydraulic head in an aquifer with uncertain (random) hydraulic conductivity. The proposed 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 either statistical moment equation or MCS. 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 verify the accuracy of the proposed method over Monte-Carlo approach. The robustness of the proposed method in terms of spatial variability is also demonstrated. These findings were communicated in [45].

The distribution based method proposed in Chapter 2, is only applicable to stationary, mildly-to-moderately heterogeneous porous media. In Chapter 3, we combines the random domain decomposition (RDD) and the method of distributions to alleviates these limitations. The proposed CDF-RDD scheme 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 saturated flow. The solutions of this equation are then averaged over realizations of the geological maps to compute the hydraulic head CDF. We present the numerical implementation of the proposed CDF-RDD scheme and its computational gains over other approaches. These contributions were also communicated in [118].

Next, the distribution-based approaches for flow problem developed in Chapters 2 and 3 are coupled with the stochastic methods for transport problem. We consider two-phase immisicible flow and transport in heterogeneous formation. The previous studies [44, 43] have shown that the distribution of saturation can be achieved efficiently by considering the distribution of travel time and the statistics of equivalent injection time (EIT). This method, so-called FROST method, is robust for geostatistical models with high permeability variances ($\sigma_K^2 \geq 1$), but still requires a lot of MC realizations (>5000) to compute the distribution of travel time accurately.

In Chapter 4, we extend the previous FROST method to compute both saturation and travel time without employing the MCS method. We first derive the deterministic CDF equation for travel time. The derived high dimensional CDF equation is transformed into multiple two-dimensional equations using the mean streamline coordinate for computational accuracy. A new numerical scheme using pseudo-time integration and flux-limited scheme is developed for getting an accurate and cost-effective numerical solution in the presence of discontinuities on solutions and non-smooth coefficients. The proposed method which are referred to as CDF-FROST method couples the CDF method for travel with the FROST method. The numerical experiments in incompressible and

immiscible displacement settings are performed to reveal the accuracy and computational efficiency of CDF-FROST over previous approaches.

In Chapter 5, we are changing the subject from forward problem to inverse problem. We alleviate the ill-posedness of subsurface inversion by deploying the plug and play (PnP) [101] and consensus equilibrium (CE) strategies [16], which provide a flexible framework for image reconstruction. The proposed framework integrates conventional optimization-based methods with the state-of-the-art machine learning-based methods.

Our CE methodology for spatial delineation of geologic formations consists of an image denoiser [72, 26, 121] and a variational auto-encoder (deep learning-based emulator) [48]. The former ameliorates the reconstruction noise, yielding well-defined geological structures; its mathematical equivalence with the proximal operator allows the deployment of advanced denoisers (e.g., CNN-based denoiser) that do not correspond to a regularization objective. The latter defines a geology prior that imposes a geological constraint, e.g., continuity and shape of geological features, onto the reconstructed image. We conduct a series of numerical experiments dealing with transient two-dimensional flow driven by a pumping well and natural hydraulic head gradient to demonstrate the CE framework's ability to delineate, both probabilistically and deterministically, complex subsurface environments. These results were communicated in [119].

Chapter 2

Method of distributions for steady state flow in heterogeneous porous media

The concept of this chapter is adopted from a previously published paper:

H. Yang, F. Boso, H. A. Tchelepi, & D. M. Tartakovsky, Probabilistic forecast of single-phase flow in porous media with uncertain properties. Water Resources Research, 55(11), 8631-8645, 2019.

Most of figures and equations are identical to corresponding sections of the paper.

2.1 Introduction

The probability distribution information including CDFs and PDFs is required for risk assessment and decision-making under uncertainty [90, 91], 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, e.g., 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 [68, 27, 56]. Monte Carlo simulations (MCS) can be used to compute the CDF $F_h(H;\mathbf{x})$. However, this approach requires a large number of 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 [18], multilevel MC [38] and various stochastic finite element methods [113]. While widely used in practice, including for subsurface-related applications [63, 31, 23], under certain conditions such methods can be slower than MCS. For example, MLMC might become slower than regular MC when estimating a system state's distribution to the same accuracy [38], 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 [3].

The method of distributions [96] 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 [60, 84, 95, 7] and multiphase flow described by the Buckley-Leverett equation [105, 44, 43]. 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 [103] or its subsequent modifications [76]. 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 [9] ameliorates this deficiency by preserving both the mean and variance. It has been used to quantify uncertainty in advection-dispersion [9] and advection-dispersion-reaction [8] problems.

We develop the method of distributions for steady-state saturated flow in subsurface environments with uncertain hydraulic conductivity and external forcings. Section 2.2 contains a formulation of groundwater flow problem with uncertain inputs and a derivation of the PDF and CDF equations for hydraulic head. In section 2.3, 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 the input uncertainty (variance of log hydraulc conductivity). Main findings and conclusions drawn from our study are summarized in section 2.4.

2.2 Problem Formulation and Method of Distributions

In this section we provide a probabilistic description of steady-state 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 $h(\mathbf{x})$.

2.2.1 Steady-State Flow in Porous Media

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

$$\nabla \cdot [K(\mathbf{x})\nabla h(\mathbf{x})] = g(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{2.1}$$

subject to boundary conditions

$$\begin{aligned} h(\mathbf{x}) = \phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \quad \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N; \\ \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(x) + ah(\mathbf{x}) = \varphi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_R. \end{aligned} \tag{2.2}$$

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 a linear combination of the former (with given a) 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 log-normal 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 [110] 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 $F_\varphi(\Upsilon;\mathbf{x})$, 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 (2.1) and (2.2) with random $K(\mathbf{x})$, $\phi(\mathbf{x})$, $\psi(\mathbf{x})$ and $\varphi(\mathbf{x})$ is the one-point CDF of hydraulic head, $F_h(H;\mathbf{x}) = \mathbb{P}[h(\mathbf{x}) \le H]$. Our goal is to derive a deterministic equation satisfied by $F_h(H;\mathbf{x})$.

2.2.2 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,

$$F_h(H;\mathbf{x}) = \langle \Pi(H,h(\mathbf{x}))\rangle. \tag{2.3}$$

Multiplying (2.1) 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,

$$\nabla\cdot[K(\mathbf{x})\nabla\Pi] - K(\mathbf{x})\frac{\partial^2\Pi}{\partial H^2}\nabla h(\mathbf{x})\cdot\nabla h(\mathbf{x}) = -g(\mathbf{x})\frac{\partial\Pi}{\partial H}. \tag{2.4}$$

We use the Reynolds decomposition to represent the random functions in (2.4) 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 $F_h(H;\mathbf{x},t)$,

$$\bar{K}\nabla^2 F_h + M = -g(\mathbf{x})\frac{\partial F_h}{\partial H}, \qquad M \equiv \nabla\cdot\langle K'(\mathbf{x})\nabla\Pi'\rangle - \langle K(\mathbf{x})\frac{\partial^2\Pi}{\partial H^2}\nabla h(\mathbf{x})\cdot\nabla h(\mathbf{x})\rangle. \tag{2.5}$$

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 Interaction-by-Exchange-with-the-Mean (IEM) approach [103] by postulating a closure

$$M \approx [\alpha(\mathbf{x})(H-\bar{h}(\mathbf{x})) + \beta(\mathbf{x})]\frac{\partial F_h}{\partial H}, \tag{2.6}$$

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 (2.5) and the (steady-state) advection-diffusion equation. Substitution of (2.6) into (2.5) gives a closed CDF equation

$$\bar{K}\nabla^2 F_h + [\alpha(\mathbf{x})(H-\bar{h}(\mathbf{x})) + \beta(\mathbf{x}) + g(\mathbf{x})]\frac{\partial F_h}{\partial H} = 0, \qquad (\mathbf{x},H)\in\Omega\times(H_{\min},H_{\max}). \tag{2.7}$$

Empirical or phenomenological selection of the closure variables [75, 76, 42] does not automatically guarantee an accurate reproduction of the first and second statistical moment of the distribution, i.e., mean $\bar{h}(\mathbf{x})$ and variance $\sigma_h^2(\mathbf{x})$.

Following [9] and [8], we construct the closure variables α and β in a way that ensures that the CDF equation (2.7) 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

$$\bar{h}(\mathbf{x}) = H_{\max} - \int_{H_{\min}}^{H_{\max}} F_h(H;\mathbf{x})\mathrm{d}H, \qquad \sigma_h^2(\mathbf{x}) = H_{\max}^2 - 2\int_{H_{\min}}^{H_{\max}} HF_h(H;\mathbf{x})\mathrm{d}H - \bar{h}(\mathbf{x})^2. \quad (2.8)$$

Hence, since $F_h(H_{\min};\mathbf{x}) = 0$ and $F_h(H_{\max};\mathbf{x}) = 1$, integrating (2.7) over H yields

$$\bar{K}\nabla^2\bar{h} - \beta(\mathbf{x}) - g(\mathbf{x}) = 0. \quad (2.9)$$

By the same token, multiplying both sides of (2.7) 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 (2.9),

$$\bar{K}\nabla^2\sigma_h^2 + 2\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - 2\alpha(\mathbf{x})\sigma_h^2 = 0. \quad (2.10)$$

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)

$$\bar{K}\nabla^2\tilde{h} + \rho(\mathbf{x}) - g(\mathbf{x}) = 0, \qquad \rho \equiv \bar{K}\lim_{\boldsymbol{\chi}\to\mathbf{x}}[\nabla_{\mathbf{x}}\cdot\nabla_{\boldsymbol{\chi}}C_{Yh}(\mathbf{x},\boldsymbol{\chi})] \quad (2.11)$$

and

$$\bar{K}\nabla^2\tilde{\sigma}_h^2 + 2V(\mathbf{x}) = 0 \quad (2.12a)$$

with

$$V \equiv \frac{1}{2}\bar{K}\lim_{\boldsymbol{\chi}\to\mathbf{x}}[\nabla_{\mathbf{x}}h^{(0)}\cdot\nabla_{\mathbf{x}}C_{Yh}(\mathbf{x},\boldsymbol{\chi}) - \nabla_{\chi}\cdot\nabla_{\mathbf{x}}C_h(\mathbf{x},\boldsymbol{\chi})] + \left(1+\frac{\sigma_Y^2}{2}\right)g(x)C_{Yh}(x,x). \quad (2.12b)$$

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 $\tilde{\sigma}_h^2 = [\sigma_h^2]^{(1)} + \mathcal{O}(\sigma_Y^4)$; $C_{Yh}(\mathbf{x},\boldsymbol{\chi})$ is the first-order approximation of the cross-covariance $\langle Y'(\mathbf{x})h'(\boldsymbol{\chi})\rangle$; and $C_h(\mathbf{x},\boldsymbol{\chi})$ is the first-order approximation of the hydraulic head's

auto-covariance $\langle h'(\mathbf{x})h'(\boldsymbol{\chi})\rangle$.

Imposition of the equivalency between the mean ($\bar{h}$) and variance (σ_h^2) computed with the CDF method, (2.9) and (2.10), and the moment equations, (2.11) and (2.12), yields expressions for the closure variables $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$. Specifically, the equations for the mean, (2.9) and (2.11), are equivalent (up to the first order in σ_Y^2) if $\beta \equiv -\rho$; and the equations for the variance, (2.10) and (2.12), 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

$$\alpha(\mathbf{x}) = \frac{\bar{K}\nabla\bar{h}\cdot\nabla\bar{h} - V}{\sigma_h^2}, \qquad \beta(\mathbf{x}) = \bar{K}\nabla^2\bar{h} + g(\mathbf{x}), \qquad V(\mathbf{x}) = -\frac{1}{2}\bar{K}\nabla^2\sigma_h^2. \tag{2.13}$$

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 $\sigma_h^2(\mathbf{x})$.

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 (2.7) is subject to boundary conditions

$$F_{h|\alpha} = F_\phi, \quad \mathbf{x}\in\Gamma_D; \qquad -\nabla F_{h|\alpha}\cdot\mathbf{n} = [\gamma(\mathbf{x})(H-\bar{h}) + \eta(\mathbf{x})]\frac{\partial F_{h|\alpha}}{\partial H}, \quad \mathbf{x}\in\Gamma_N, \tag{2.14a}$$

where

$$\gamma(\mathbf{x}) = \frac{\bar{K}\sigma_h^2\cdot\mathbf{n}}{2\sigma_h^2 - 4\bar{h}^2}, \qquad \eta(\mathbf{x}) = \bar{K}\nabla\bar{h}\cdot\mathbf{n} - \bar{\psi}, \tag{2.14b}$$

$\bar{\psi}(\mathbf{x})$ 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 (The detailed derivations of boundary conditions are discussed in Appendix B). The general property of a CDF provides the remaining boundary conditions in the H space,

$$F(H = H_{\min};\mathbf{x}) = 0, \qquad F(H = H_{\max};\mathbf{x}) = 1. \tag{2.14c}$$

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 (2.7). Since the coefficients of the CDF equation (2.7) 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 (2.7)–(2.14) comprises two modules. The first module provides finite-volume solutions of the SMEs (A.4)–(A.12) 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 [61]. The second module computes the coefficients $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ in (2.7), and solves the latter in non-conservative 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 [96].

2.3 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 towards 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 multi-variate 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 $g(x) = T_{\text{well}}(h(x) - h_{\text{well}})$ in (2.1), 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 Monte Carlo simulations (MCS). Equiprobable MC realizations were generated by the sequential Gaussian simulator [29]. 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 $N_{\text{MCS}} = 10,000$.

2.3.1 Accuracy of the CDF method

Since the coefficients in the CDF equation (2.7) are given in terms of the mean and variance of the hydraulic head $h(\mathbf{x})$, we start by analyzing the ability of the SMEs (A.4)–(A.12) to accurately approximate $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$. Figure 2.1 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 SMEs, 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 2.1 reveals a slight discrepancy between these two sets of moments, as quantified by the average errors

$$\bar{\epsilon}_{\text{mean}} = \frac{1}{\|\Omega\|}\int_\Omega \mid \bar{h}_{\text{SME}} - \bar{h}_{\text{CDF}} \mid \mathrm{d}\mathbf{x}, \qquad \bar{\epsilon}_{\text{var}} = \frac{1}{\|\Omega\|}\int_\Omega \mid \sigma^2_{h,\text{SME}} - \sigma^2_{h,\text{CDF}} \mid \mathrm{d}\mathbf{x}, \tag{2.15}$$

where $\|\Omega\|$ is the volume of the flow domain Ω. The errors $\bar{\epsilon}_{\text{ave}}$ and $\bar{\epsilon}_{\text{var}}$ decrease with the grid size along the H coordinate, ΔH, decreases (Figure 2.2). This result verifies that the discrepancy is 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.

Consistent with the previous SME-focused studies [68, 92, 93, 56, 61, 82], the mean and variance of hydraulic head computed with SMEs 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 SMEs: $\sigma_Y^2 = 1$ for the mean uniform flow, and $\sigma_Y^2 = 2$ for the convergent flow.

Spatial maps of exceedance/non-exceedance 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

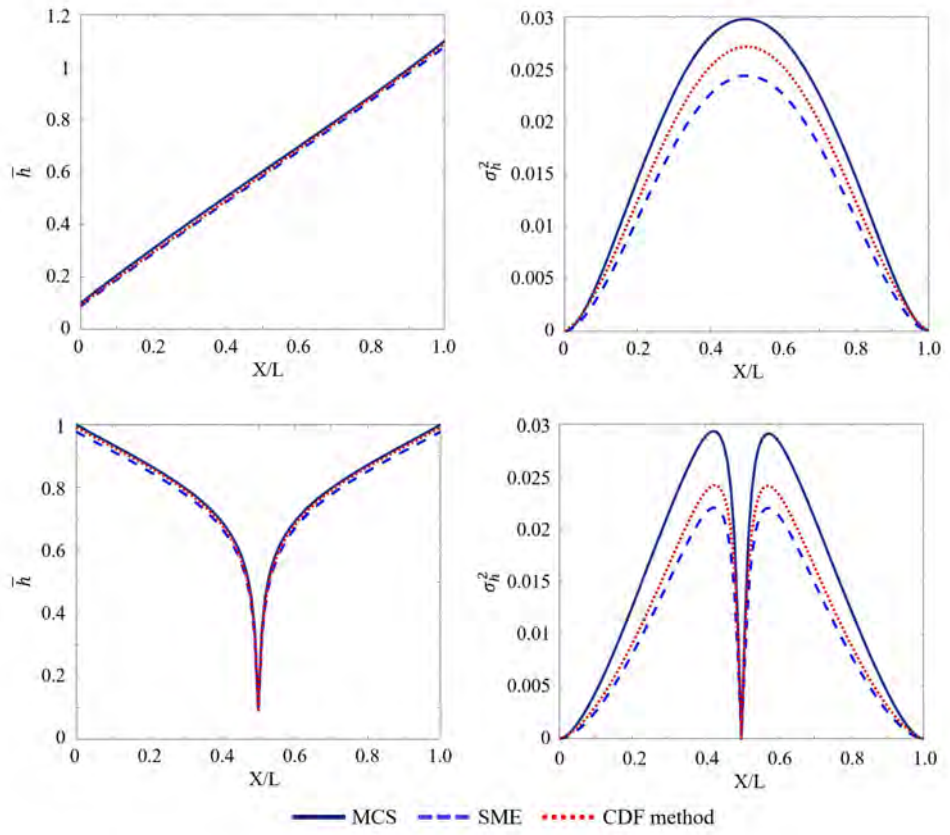

Figure 2.1: Mean (left column) and variance (right column) of hydraulic head, $\bar{h}(x_1, x_2 = 1/2)$ and $\sigma_h^2(x_1, x_2 = 1/2)$, for mean uniform flow (top row) and flow to a well located at the middle of the domain (bottom row). These moments are alternatively computed with Monte Carlo simulations (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.

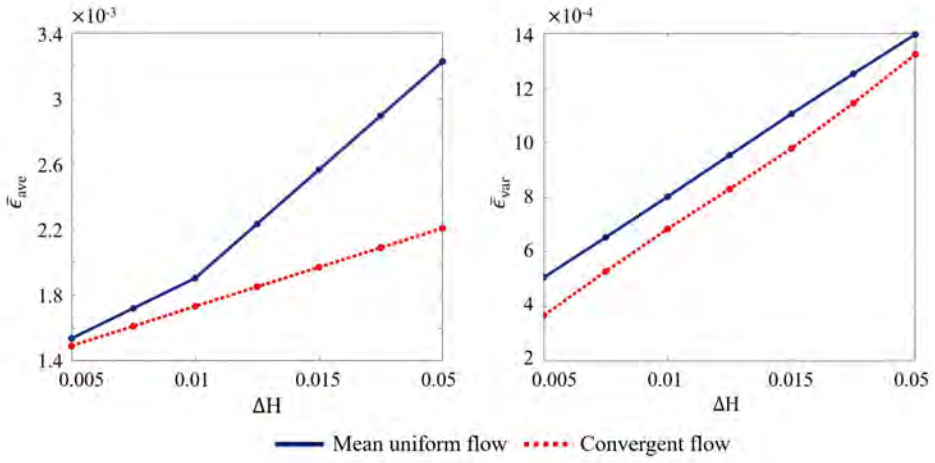

Figure 2.2: Average discrepancies $\bar{\epsilon}_{\text{mean}}$ (left) and $\bar{\epsilon}_{\text{var}}$ (right) between the mean and variance of hydraulic head $h(\mathbf{x})$, alternatively computed as quadratures of the CDF $F(H;\mathbf{x})$ in (2.7) or by solving the SME. The discrepancies decay as the 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.

identify regions in the spatial domain where the corresponding risk is higher than desired. Figure 2.3 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 (2.7). Visual inspection of the two sets of map, as well as the CDFs $F_h(H;\mathbf{x})$ presented in Figure 2.4 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 (F_h^{MCS}, computed using $N_{\text{MCS}} = 10000$ realizations) is provided by the first Wasserstein distance between two distributions (a.k.a. Earth Mover's metric)

$$\mathcal{D}(\mathbf{x}) \equiv \int_{H_{\min}}^{H_{\max}} \mid F_h(H',\mathbf{x}) - F_h^{\text{MCS}}(H',\mathbf{x}) \mid \mathrm{d}H'. \tag{2.16}$$

The numerical integration is computed with the Gauss-Legendre quadrature rule. The resulting contour plots of $\mathcal{D}(\mathbf{x})$ are shown in Figure 2.5. 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

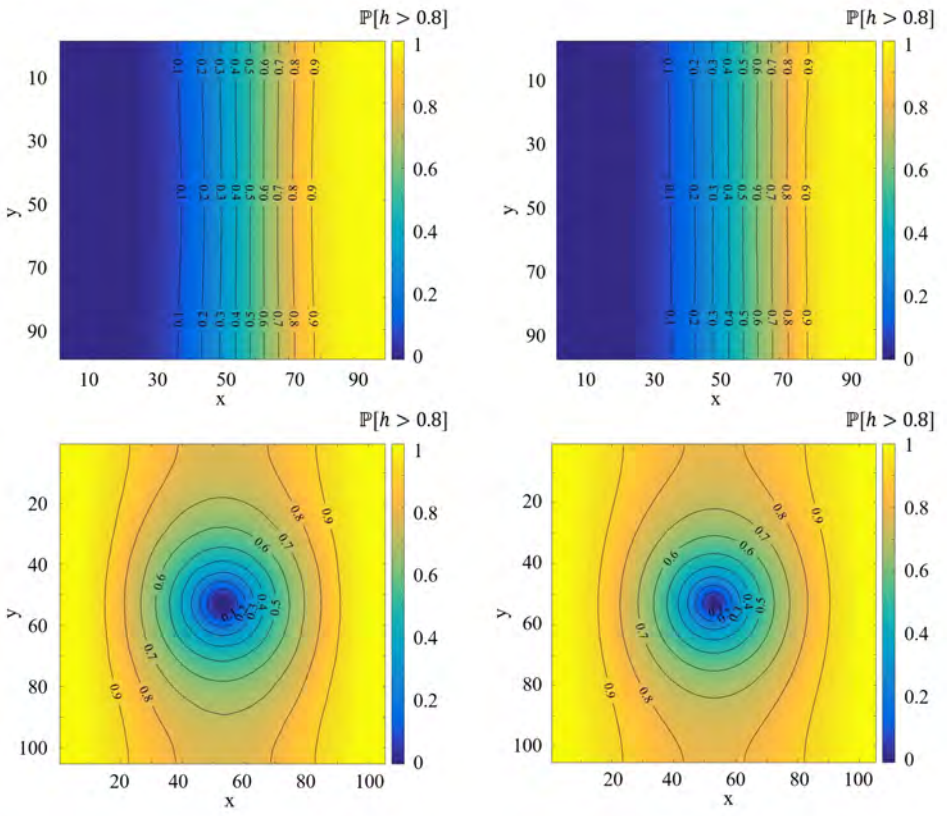

Figure 2.3: Spatial maps of exceedance probability $\mathbb{P}[h(\mathbf{x}) > H = 0.8] = 1 - F_h(H = 0.8; \mathbf{x})$ obtained with MCS (left column) and CDF method (right column) for mean uniform flow (top row) and convergent flow (bottom row). The statistical properties of log-conductivity are $\bar{Y} = 0$; 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.

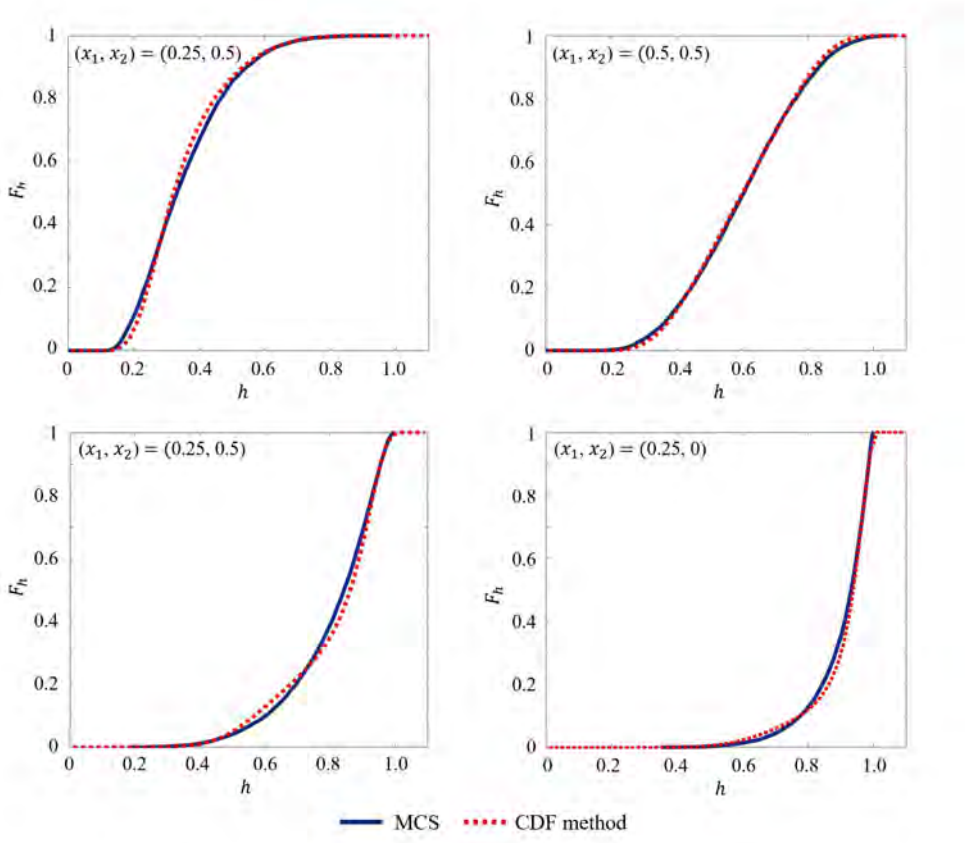

Figure 2.4: Hydraulic head CDFs F_h computed with MCS and the CDF method at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the flow domain for mean uniform flow (top row) and convergent flow (bottom row).

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).

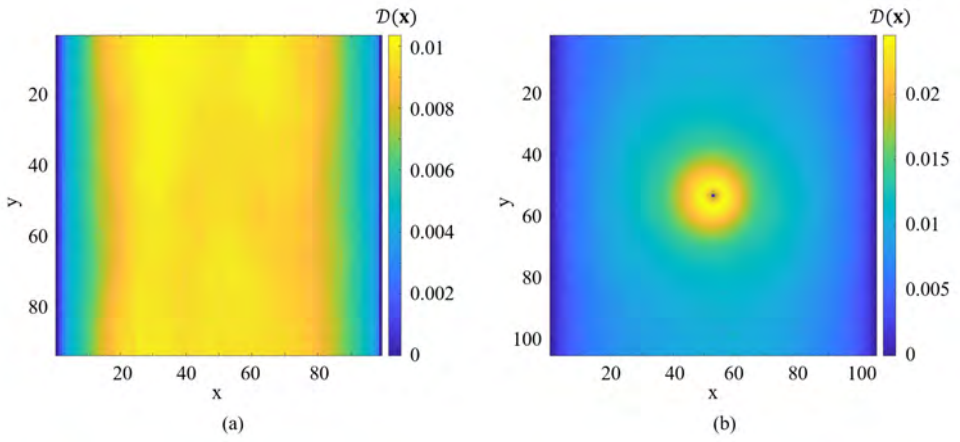

Figure 2.5: Spatial maps of the Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the hydraulic head CDFs computed with the CDF method and Monte Carlo simulations for mean uniform flow (left) and convergent flow (right).

2.3.2 Computational Efficiency of the CDF Method

As mentioned in the Introduction, a raison d'être for the development of the method of distribution 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 PDE (2.7), MCS consist of repeated solves of a large number of d-dimensional (possibly nonlinear) PDEs like the flow equation (2.1). Figure 2.6 visualizes the difference of MCS and distribution-based methods on computing probability distributions.

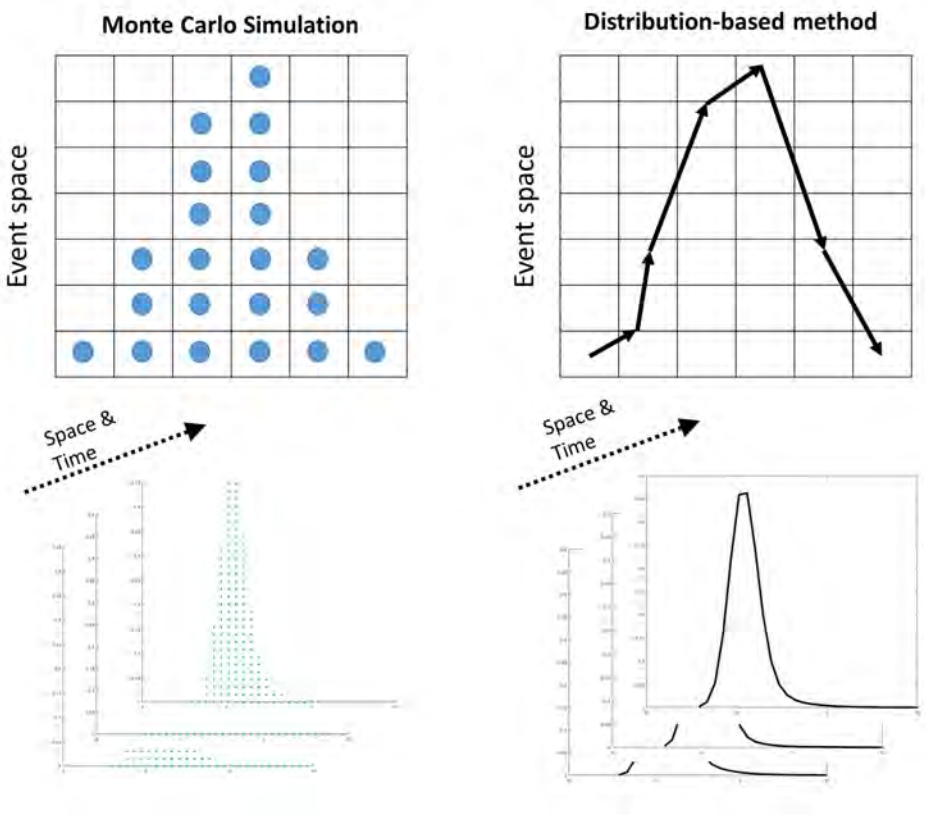

Figure 2.6: Visualization of difference between MCS (left) and distribution-based methods.

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 (2.7) 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 SMEs (A.4)–(A.12), 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 2.1 and 2.2 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, $\mathcal{D}_{\text{ave}} = \|\Omega\|^{-1} \int_{\Omega} \mathcal{D}(\mathbf{x}) \mathrm{d}\mathbf{x}$ where $\|\Omega\|$ is the volume of the flow domain Ω (in our simulation, the unit area of the square). Specifically, the discrepancy level $\mathcal{D}_{\text{ave}} \approx 0.01$ of the CDF method is achieved by MCS with $N_{\text{MCS}} = 1,240$ realizations in the mean uniform flow regime and with $N_{\text{MCS}} = 1,470$ realizations in the convergent flow regime. For the same discrepancy level of $\mathcal{D}_{\text{ave}} \approx 0.01$, 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 SMEs by a factor of 3 results in the similar discrepancy level, while speeding up the computation by another order of magnitude.

Table 2.1: Computational time of the CDF method and MCS in the case of mean uniform flow.

Method	Grid size	Error $\mathcal{D}_{\text{ave}}$	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

Table 2.2: Computational time of the CDF method and MCS in the case of convergent flow.

Methods	Grid size	Error $\mathcal{D}_{\text{ave}}$	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

2.3.3 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 SMEs (A.4)–(A.12) 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 ($\mathcal{D}_{\text{ave}}$ increases by about a factor of 2) as σ_Y^2 increases from 1 to 5 (for fixed $\ell_Y = 0.1$), i.e., the spatial variability of conductivity $K(\mathbf{x})$ increases by about five orders of magnitude (Figure

2.7a).

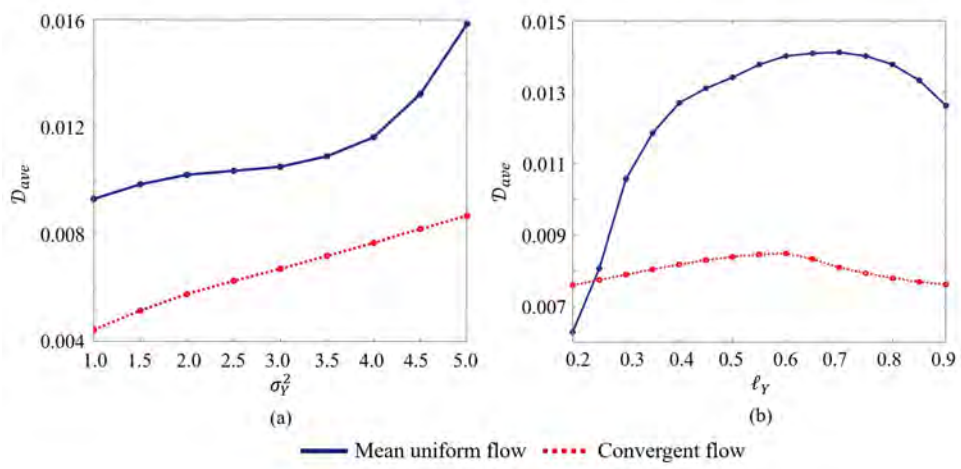

Figure 2.7: Dependence of the average Wasserstein distance between our CDF solution and its MCS estimate, $\mathcal{D}_{\text{ave}} = \mathcal{D}_{\text{ave}}(\sigma_Y^2, \ell_Y)$, on the variance ($\sigma_Y^2$ for fixed $\ell_Y = 0.1$, left) and correlation length (ℓ_Y for fixed $\sigma_Y^2 = 1.0$, right) of log hydraulic conductivity $Y = \ln K$ for mean uniform flow and convergent flow .

Impact of Conductivity's Correlation Length. The correlation length ℓ_Y controls the degree of regularity of the (log) conductivity field. The dependence of $\mathcal{D}_{\text{ave}}$ on ℓ_Y (for fixed $\sigma_Y^2 = 1.0$) is shown in Figure 2.7b. In both flow configurations, $\mathcal{D}_{\text{ave}}$ increases with ℓ_Y as long as $\ell_Y \leq 0.7$, and decreases when $\ell_Y \geq 0.7$. The maximum values of $\mathcal{D}_{\text{ave}}$ 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 is extremely small. Obviously, 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 SME and closures are maximum when the correlation length is intermediate [56]. Regardless, the average discrepancy $\mathcal{D}_{\text{ave}}$ remains small regardless of ℓ_Y, which demonstrates that the CDF method is robust to the magnitude of ℓ_Y.

2.3.4 Impact of Moments' Approximation

Two types of approximations underpin the derivation of the CDF equation: the moment-preserving closure leading to (2.7) and the perturbation approximation used to close the moment equations. The latter affects the coefficients α and β in the CDF equation (2.7), 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 (2.7) whose coefficients are alternatively computed with the SMEs and the reference MCS. Since the moments

computed with the reference MCS are treated as exact, their use in the CDF equation (2.7) isolates the impact of the moment-preserving closure.

Table 2.3: Average Wasserstein discrepancy $\mathcal{D}_{\text{ave}}$ between the CDF method with MCS moments and the CDF method with SME moments for two flow configurations.

Methods	Error $\mathcal{D}_{\text{ave}}$	
	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}

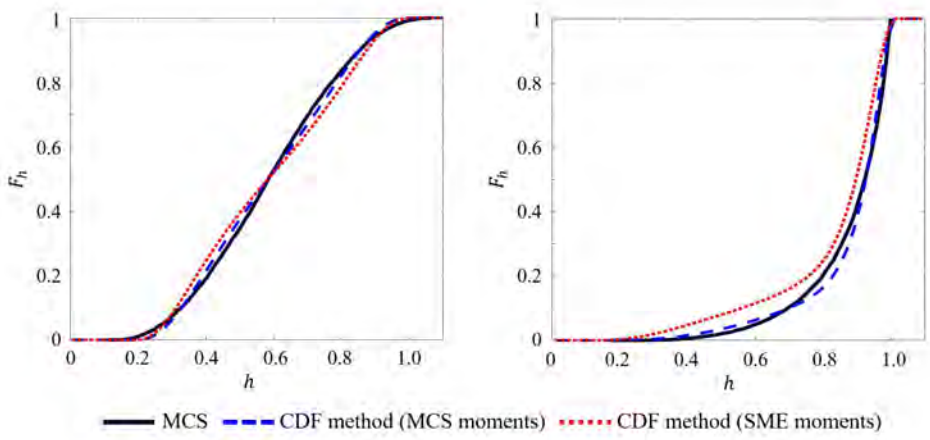

Figure 2.8: Hydraulic head CDF $F_h(H;\mathbf{x})$ computed with MCS, the CDF method with MCS moments, and the CDF method SME moments at $\mathbf{x} = (0.5, 0.5)^\top$ for mean uniform flow (left) and at $\mathbf{x} = (0.5, 0.25)^\top$ for convergent flow (right).

For both flow configurations, we set $\sigma_Y^2 = 3.0$ and $\ell_Y = 0.4$. Table 2.3 shows the relative minor impact of the perturbation closures of the SMEs (A.4)–(A.12) on the average Wasserstein distance $\mathcal{D}_{\text{ave}}$ 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 2.8 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 (2.7) .

2.4 Summary and Conclusions

We developed the method of distributions to probabilistically predict saturated flow in porous media with uncertain hydraulic conductivity and/or uncertain boundary functions. The method results in a deterministic partial-differential equation for the cumulative distribution function (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 with Monte Carlo simulations (MCS) or by solving the corresponding statistical moment equations (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, e.g., 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 SMEs and CDF equation, e.g., 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.

Chapter 3

Method of distributions for quantification of geologic uncertainty in flow simulations

The concept of this chapter is adopted from a previously published paper:

H. Yang, F. Boso, H. A. Tchelepi, & D. M. Tartakovsky, Method of distributions for quantification of geologic uncertainty in flow simulations. Water Resources Research, 56(7), e2020WR027643.

Most of figures and equations are identical to corresponding sections of the paper.

3.1 Introduction

The method of distributions in Chapter 2 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 [68, 61, 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) [111, 109] 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 [110], generalized polynomial chaos expansions [114], and stochastic collocation methods [62]. 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 3.2. It combines the CDF method for flow in statistically homogeneous porous media [116] with RDD that accounts for geologic uncertainty [40]. In Section 3.3, 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 3.4.

3.2 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 [41], object-based geostatistics [30] and multi-point geostatistcs [87]—or machine learning tools such as support vector machines [112] or nearest-neighbor estimators [94]. 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

$$\nabla \cdot [K(\mathbf{x})\nabla h(\mathbf{x})] = g(\mathbf{x}), \qquad \mathbf{x} = (x_1, \ldots, x_d)^\top \in \Omega, \tag{3.1}$$

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 (3.1) is subject to boundary conditions

$$h(\mathbf{x}) = \phi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad -K(\mathbf{x})\nabla h(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \psi(\mathbf{x}), \quad \mathbf{x} \in \Gamma_N. \tag{3.2}$$

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 $\mathbf{x}_n \in \Omega$ ($n = 1, \ldots, N_{\text{meas}}$). 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 [61] and PDF/CDF equations [116], 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 3.2.1. It is deployed in Section 3.2.2 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 3.2.3.

Distribution based framework for computation of flow and transport

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 ($i = 1, \ldots, N_{\text{gu}}$). For instance, the random α can be parameterized as random contact point for the stratified reservoir or random field for the complicated facies geometry. 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 [108].

With these preliminaries, we replace (3.1) with

$$\nabla\cdot[K_i(\mathbf{x})\nabla h_i(\mathbf{x})] = g(\mathbf{x}), \qquad \mathbf{x}\in\Omega_i, \qquad i=1,\dots,N_{\mathrm{gu}}, \tag{3.3}$$

which is subject to boundary conditions (3.2)

and the continuity conditions

$$h_i(\mathbf{x}) = h_j(\mathbf{x}), \qquad K_i(\mathbf{x})\nabla h_i(\mathbf{x})\cdot\mathbf{n}_i(\mathbf{x}) = K_j(\mathbf{x})\nabla h_j(\mathbf{x})\cdot\mathbf{n}_j(\mathbf{x}), \qquad \mathbf{x}\in\Gamma_{ij}. \tag{3.4}$$

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 (3.3) and (3.4), 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 [110]; here, we use it to derive a deterministic equation for the full CDF of hydraulic head.

3.2.2 Combined CDF-RDD Approach

For a given geological map, defined by the label (realization) $\alpha = a$, we show in Section 2.2 that the conditional CDF $F_{h|\alpha}(H;\mathbf{x}|\alpha=a)$ of hydraulic head $h(\mathbf{x})$ in (3.3) satisfies a deterministic $(d+1)$-dimensional differential equation

$$\nabla\cdot(\bar{K}(\mathbf{x})\nabla F_{h|\alpha}) + \frac{\partial(UF_{h|\alpha})}{\partial H} = \upsilon F_{h|\alpha}, \qquad \tilde{\mathbf{x}} = (x_1,\dots,x_d,H)^\top\in\tilde{\Omega}. \tag{3.5}$$

The (ensemble) averaged hydraulic conductivity $\bar{K}(\mathbf{x})$ takes the constant value of $\bar{K}_i$ for $\mathbf{x}\in\Omega_i$, where $i=1,\dots,N_{\mathrm{gu}}$. The PDF equation (3.5) 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

$$U = \upsilon(H - \bar{h}) + \nabla \cdot (\bar{K}\nabla\bar{h}) + 2g, \quad \upsilon = \frac{\bar{K}\nabla\bar{h} \cdot \nabla\bar{h} - V}{\sigma_h^2}, \quad V = -\frac{1}{2}\bar{K}\nabla^2\sigma_h^2. \tag{3.6}$$

The coefficients (3.6) 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 the solutions of deterministic statistical moment equations (SME), defined in A, which proved to be more computationally efficient than MCS.

The CDF equation (3.5) is derived by deploying the self-consistent closure approximation in Section 2.2 that ensures that the differential equations for the moments $\bar{h}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$, obtained by integrating (3.5), are identical to the SME derived in Appendix A. This is in contrast to the interaction-with-the-mean closures [75, 76, 42] used in turbulence and combustion. The latter fail to preserve a system state's variance and, under certain conditions, its mean [9, 116].

As discussed in Section 2.2.2, (3.5) is subject to boundary conditions

$$F_{h|\alpha} = F_\phi, \quad \mathbf{x} \in \Gamma_D; \qquad -\nabla F_{h|\alpha} \cdot \mathbf{n} = [\gamma(\mathbf{x})(H - \bar{h}) + \eta(\mathbf{x})]\frac{\partial F_{h|\alpha}}{\partial H}, \quad \mathbf{x} \in \Gamma_N, \tag{3.7}$$

where

$$\gamma(\mathbf{x}) = \frac{\bar{K}\sigma_h^2 \cdot \mathbf{n}}{2\sigma_h^2 - 4\bar{h}^2}, \qquad \eta(\mathbf{x}) = \bar{K}\nabla\bar{h} \cdot \mathbf{n} - \bar{\psi}, \tag{3.8}$$

$\bar{\psi}(\mathbf{x})$ 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.

Here, we present the derivation of interface conditions for CDF equation. Due to the continuity of h at the interface, the CDF F_h always satisfies the following continuity condition

$$F_{h_i}(\mathbf{x}) = F_{h_j}(\mathbf{x}), \qquad \mathbf{x} \in \Gamma_{ij}.$$

To derive the flux continuity condition at the interface between facies i and j, Γ_{ij}, we introduce the spatial domain Ω_{ij}, centered at any point $\mathbf{x} \in \Gamma_{ij}$. Let C_{ij} be the portion of Γ_{ij} that intersects Ω_{ij}. Integrating (3.5) over Ω_{ij} and using the Gauss-Ostrogradsky theorem yield

$$\int_{\partial\Omega_{ij}} (\bar{K}\nabla_x F_h \cdot \mathbf{n})d\mathbf{x} + \int_{\Omega_{ij}} \frac{\partial(UF_h)}{\partial H} d\mathbf{x} = \int_{\Omega_{ij}} \upsilon F_h d\mathbf{x}.$$

We now assume that the thickness (in the direction orthogonal to C_{ij}) and the volume of Ω_{ij} shrink to zero, while C_{ij} is fixed. Given the finitedness of the integrand functions in both $\int_{\Omega_{ij}} \frac{\partial(UF_h)}{\partial H} d\mathbf{x}$ and $\int_{\Omega_{ij}} \upsilon F_h d\mathbf{x}$ (guaranteed by the continuity of the moments), both integrals approach zero, and

above equation reduces to

$$\int_{C_{ij}} (\bar{K}_i \nabla_x F_{h_i} \cdot \mathbf{n}_i) d\mathbf{x} = \int_{C_{ij}} (\bar{K}_j \nabla_x F_{h_j} \cdot \mathbf{n}_j) d\mathbf{x}.$$

Since above condition is satisfied on any C_{ij}, it leads to following interface condition for CDF equation.

$$F_{h_i|\alpha}(\mathbf{x}) = F_{h_j|\alpha}(\mathbf{x}), \quad \bar{K}_i \nabla F_{h_i|\alpha} \cdot \mathbf{n}_i = \bar{K}_j \nabla F_{h_j|\alpha} \cdot \mathbf{n}_j, \quad \mathbf{x} \in \Gamma_{ij}. \tag{3.9}$$

Finally, the general property of a CDF provides the remaining boundary conditions in H space,

$$F_h(H = H_{\min}; \mathbf{x}) = 0, \qquad F_h(H = H_{\max}; \mathbf{x}) = 1. \tag{3.10}$$

We recall that a combination of the solutions to (3.5)–(3.10) 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 $F_h(H; \mathbf{x})$,

$$F_h(H, \mathbf{x}) = \int F_{h|\alpha}(H; \mathbf{x}|\alpha = a) f_\alpha(a) \mathrm{d}a. \tag{3.11}$$

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.

3.2.3 Numerical Implementation

Numerical solution of the boundary-value problem (3.5)–(3.10) is computed in three steps. The first step involves finite-volume solutions of the moment equations (A.4)–(A.12), 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 [61].

The second step consists of numerical solution of (3.5) and (3.6). Among the plethora of schemes for solving a linear advection-diffusion-reaction equation, such as (3.5), we utilize a finite-volume scheme in which the computation domain $\tilde{\Omega}$ is divided into N_{fv} non-overlapping domains $\tilde{\Omega}_1, \ldots, \tilde{\Omega}_{N_{\text{fv}}}$. A finite-volume solution of (3.5) is obtained by integrating this equation over each element $\tilde{\Omega}_k$ and using the Gauss-Ostrogradsky theorem to replace the volume integrals over $\tilde{\Omega}_k$ with

the surface integrals over their surface $\partial\tilde{\Omega}$,

$$\int_{\partial\tilde{\Omega}_k} (\bar{K}\nabla_{\mathbf{x}} F_{h|\alpha} \cdot \mathbf{n}_{e,x} + U F_{h|\alpha} n_{e,H}) \mathrm{d}\tilde{\mathbf{x}} = \int_{\tilde{\Omega}_k} v F_{h|\alpha} \mathrm{d}\tilde{\mathbf{x}}, \qquad \tilde{\mathbf{x}} = (x_1, \ldots, x_d, H)^\top, \tag{3.12}$$

where $\mathbf{n}_{e,x}$ and $n_{e,H}$ are respectively the spatial and H-direction component component of $\mathbf{n}_e$, the outward unit normal vector of the interface $\partial\tilde{\Omega}_k$. The discrete form of (3.12) is

$$\sum_{l \in \mathrm{adj}(k)} T_{kl}(F^{(l)}_{h|\alpha} - F^{(k)}_{h|\alpha}) + U^+_e F^{(k)}_{h|\alpha} A_{kl} + U^-_e F^{(l)}_{h|\alpha} A_{kl} = v_k F^{(k)}_{h|\alpha} V_k, \qquad k = 1, \ldots, N_{\mathrm{fv}}, \tag{3.13}$$

where $\mathrm{adj}(k)$ is the set of neighbors of k, V_k is the volume of $\tilde{\Omega}_k$, and

$$T_{kl} = \frac{K_{kl} A_{kl}}{\delta x_{kl}}, \qquad U^+_e = \frac{U_e + |U_e|}{2}, \qquad U^-_e = \frac{U_e - |U_e|}{2}, \qquad U_e = U n_{e,H}. \tag{3.14}$$

Here K_{kl}, A_{kl}, and δx_{kl} are harmonically averaged hydraulic conductivity, differential element cross-sectional area, and the distance between the spatially connected computational nodes k and l, respectively.

A main advantage of using the finite volume implementation is that it ensures the continuity conditions (3.9) when the contact interface between the adjacent facies is defined along the computational cell boundaries by its construction. This results in considerable computational speed-up over other RDD-based numerical schemes that enforce the continuity iteratively [40, 110]. Combining (3.13) into a single system of linear algebraic equations, we obtain $\mathbf{A}\mathbf{f} = \mathbf{b}$ where $\mathbf{A}$ is the $N_{\mathrm{fv}} \times N_{\mathrm{fv}}$ coefficient matrix, $\mathbf{b}$ is the $N_{\mathrm{fv}} \times 1$ vector determined by the boundary conditions, and $\mathbf{f}$ is the $N_{\mathrm{fv}} \times 1$ 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 (3.5) 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 (3.5) 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 (3.11). We approximate this integration with the Monte Carlo average of N_α realizations of the geological map with the label α. The numerical integration of $F_{h|\alpha}$, used to compute the marginal CDF, mean, and variance of h, is carried out with the Gaussian quadrature.

3.3 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.

In both flow regimes, the flow domain Ω, a square of unit dimensionless length (normalized with the domain size L), is composed of $N_{\rm gu} = 2$ facies whose uncertain spatial arrangement is represented by four different equiprobable geological models labeled by α (see Figure 3.1). These models are generated using multi-point geostatistics, specifically SNESIM algorithm [87]. 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 $J \equiv (h_{\rm out} - h_{\rm in})/L = 0.1$, with the deterministic dimensionless hydraulic heads $h_{\rm in} = 1.1$ and $h_{\rm out} = 0.1$ (normalized with the reference hydraulic head $h_{\rm ref}$) 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 $h_{\rm well} = 0.1$; 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 $\tilde{\Omega} \equiv \Omega \times (H_{\rm min} = 0, H_{\rm max} = 1.2)$ is discretized using 45, 45, and 120 nodes in the x_1, x_2, and H directions, respectively. Thus, the total number of grids $N_{\rm fv}$ 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 [29]. The convergence study of the two flow scenarios revealed that, for each geological model, it takes $N_{\rm MCS} = 10^4$ 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 $N_{\rm MCS} = 4 \cdot 10^4$, a prohibitively large number in most applications of practical significance.

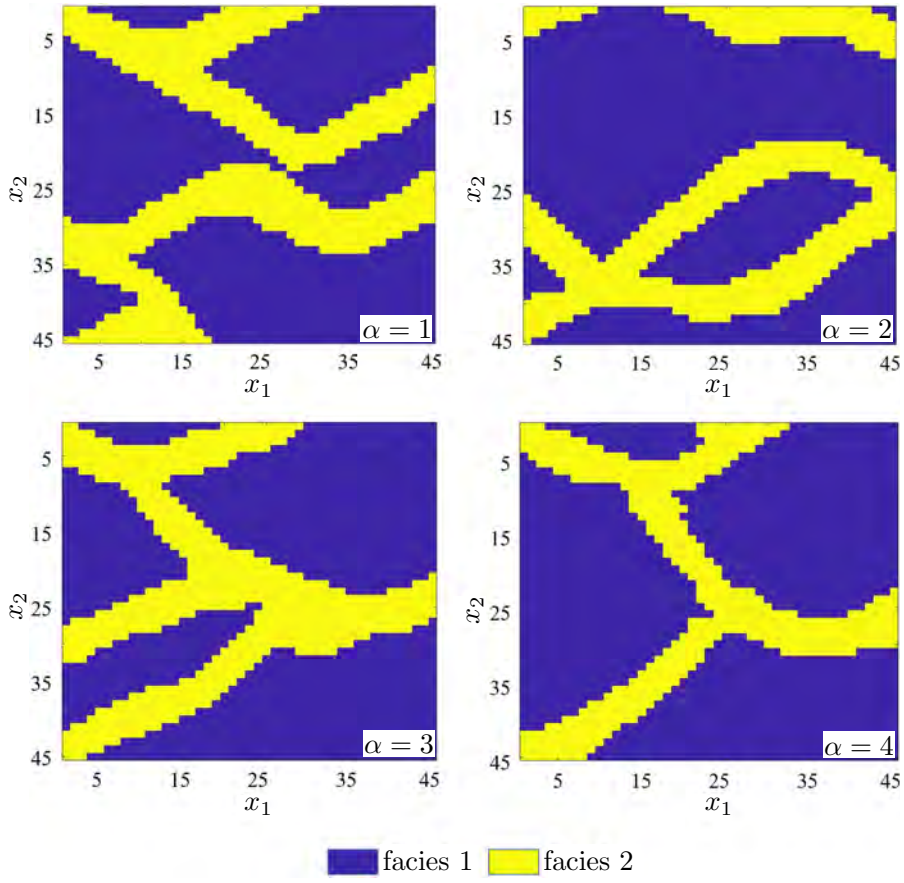

Figure 3.1: Equiprobable geological models used in numerical experiments.

3.3.1 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 Figure 3.1 as a single continuum, it would be characterized by the variance $\sigma_Y^2 \sim (\bar{Y}_1 - \bar{Y}_2)^2$ [110]; 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 3.2 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.

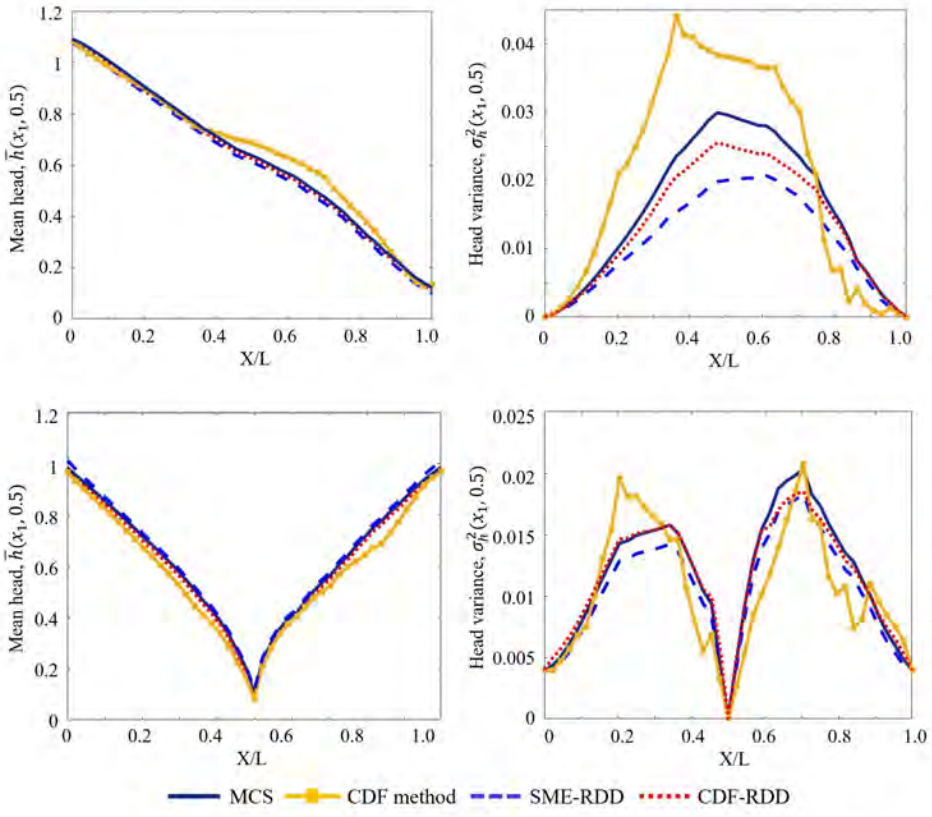

Figure 3.2: Mean (left column) and variance (right column) of hydraulic head in mean uniform flow (top row) and flow to a well (bottom row). These moments are alternatively computed with Monte Carlo simulations (MCS), the RDD-enhanced SME (SME-RDD), and evaluating the moments of F_h in the CDF equation with and without RDD (CDF and CDF-RDD, respectively).

Figure 3.2 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 (Figure 2.3). Figure 3.4 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) defined as (2.15). 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 (Figure 3.6). 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.

3.3.2 Computational Efficiency of CDF-RDD

The CDF method has been shown to be an order of magnitude faster than MCS in statistically homogeneous media [116]. Instead of multiple MC solves of the d-dimensional groundwater flow equation (3.1), it solves a single $(d+1)$-dimensional CDF equation (3.5). 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 3.1, 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

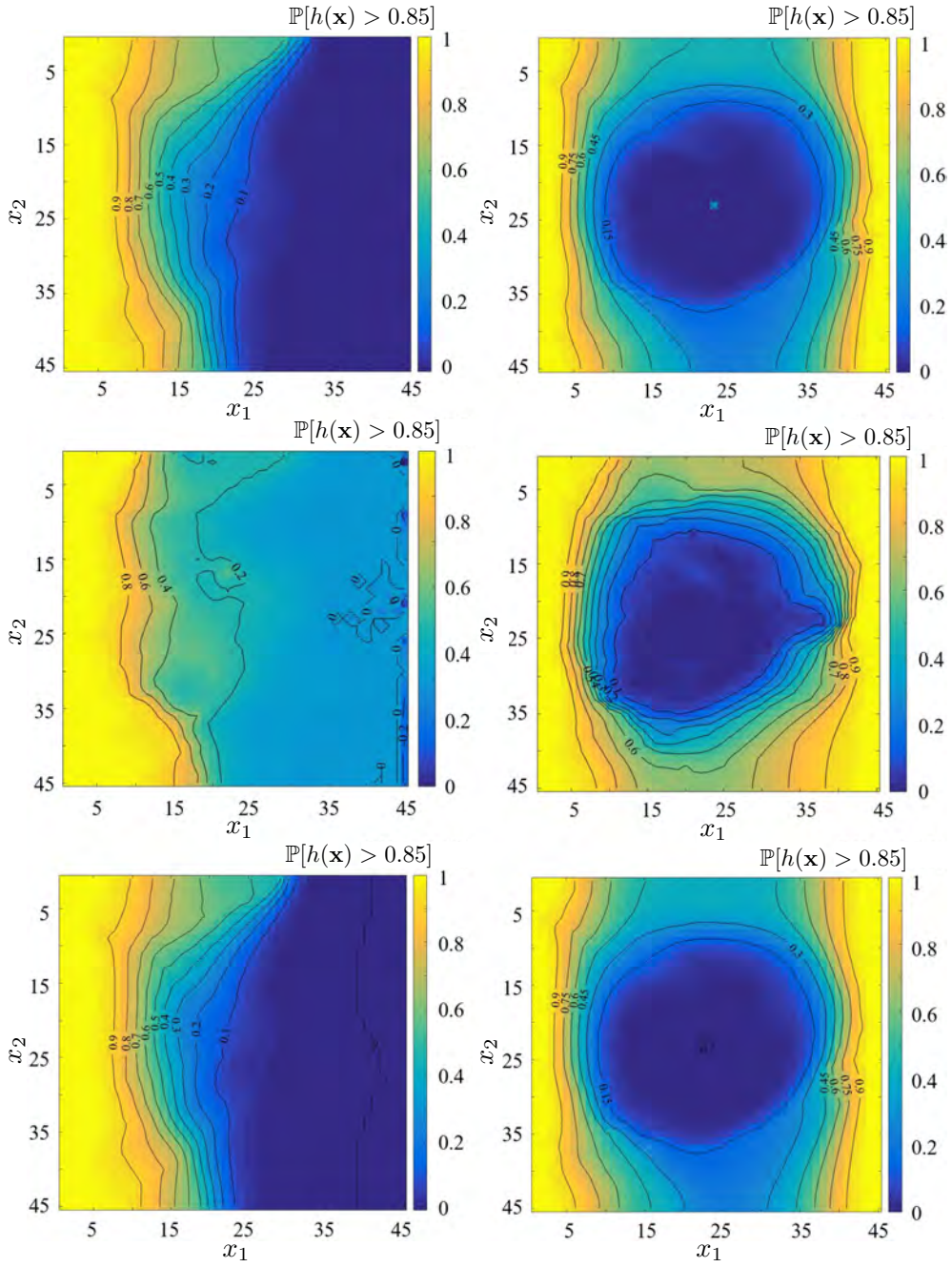

Figure 3.3: Spatial maps of exceedance probability $\mathbb{P}[h(\mathbf{x}) > H = 0.85] = 1 - F_h(H = 0.85; \mathbf{x})$ obtained with MCS (top row), the CDF method (center row), and CDF-RDD (bottom row) for mean uniform flow (left column) and convergent flow (right column).

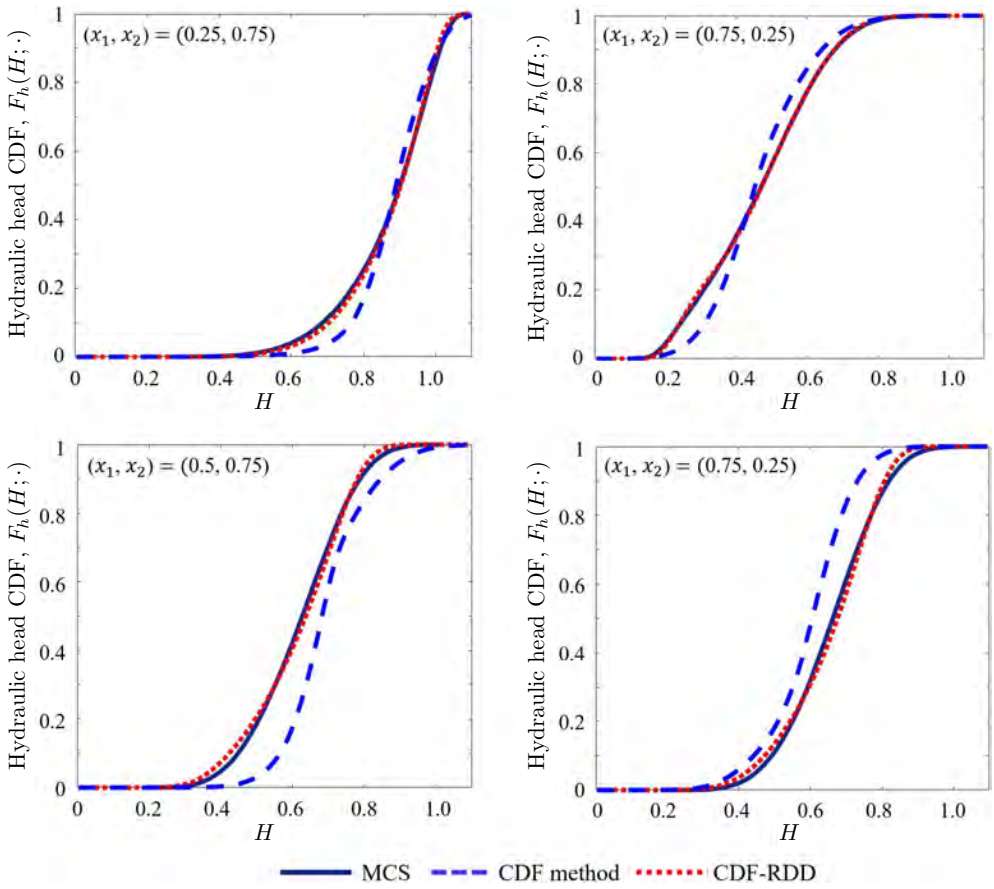

Figure 3.4: Hydraulic head CDFs F_h computed with MCS, CDF method, and CDF-RDD at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the simulation domain for the mean uniform flow (top row) and the convergent flow (bottom row).

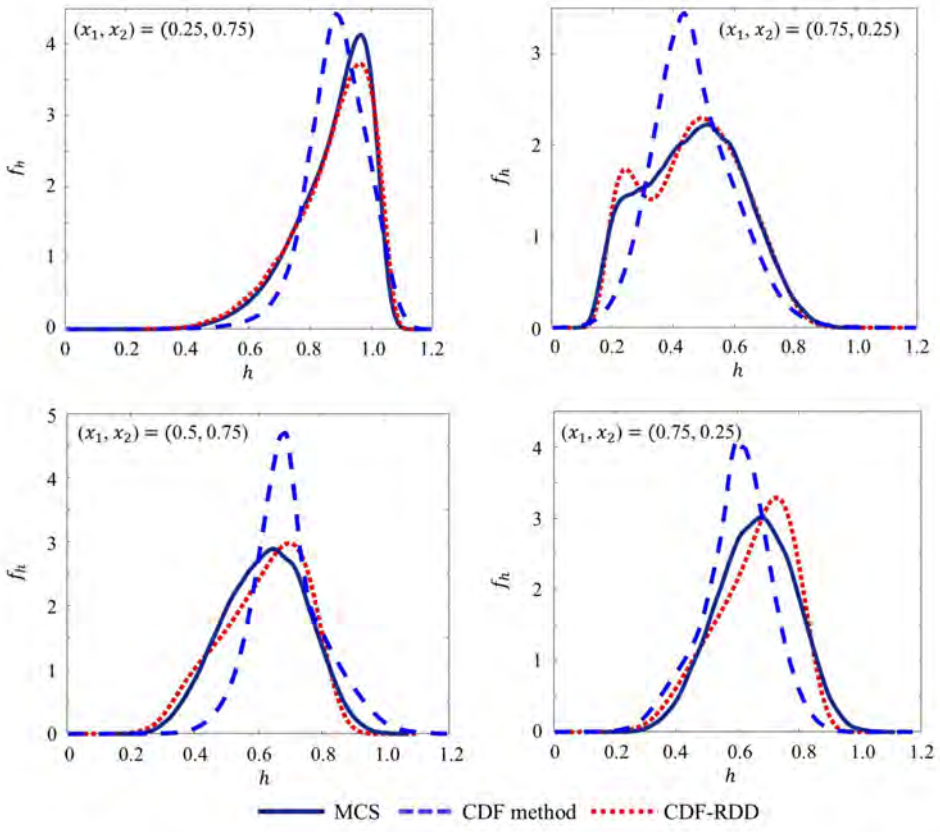

Figure 3.5: Hydraulic head PDFs f_h computed with MCS, CDF method, and CDF-RDD at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the simulation domain for the mean uniform flow (top row) and the convergent flow (bottom row).

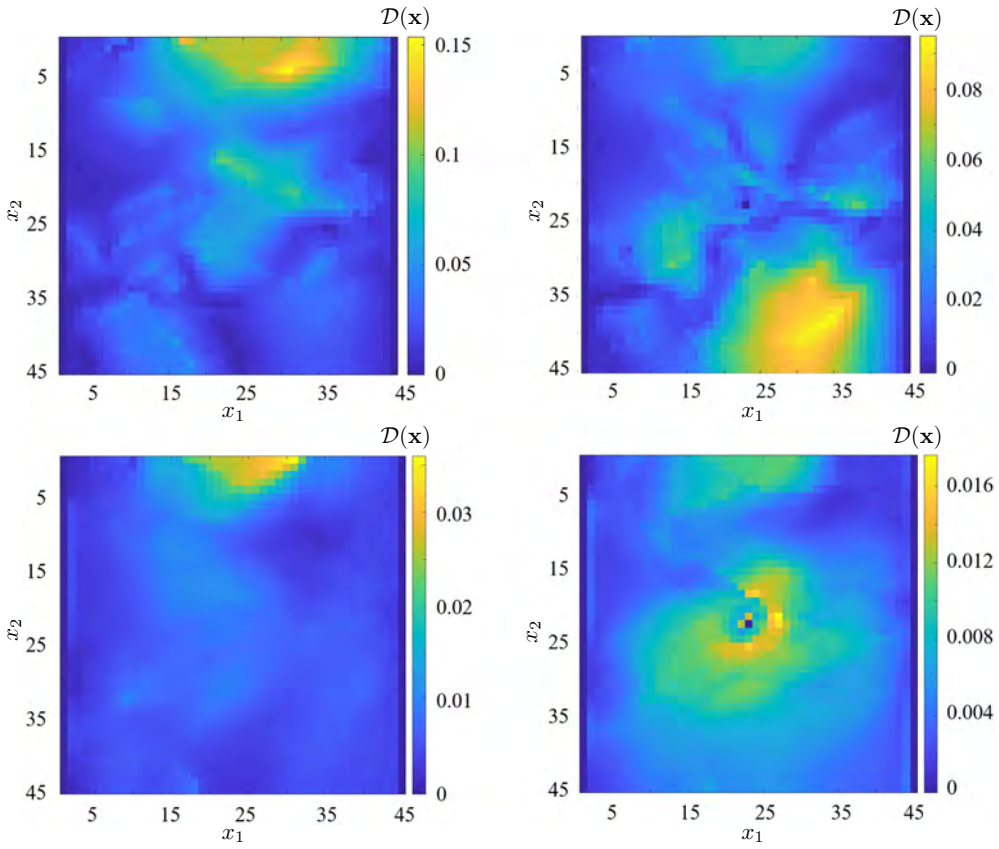

Figure 3.6: Spatial maps of the Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the "exact" MCS estimate of the hydraulic head CDF F_h^{MCS} and its approximations provided by either the CDF method (top row) or CDF-RDD (bottom row), for the mean uniform flow (left column) and the convergent flow (right column).

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 3.1: 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	$8.61 \cdot 10^{2}$

3.4 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 steady state flow equation in Chapter 2 provides a computationally efficient alternative to MCS, but its applicability is limited to statistically homogeneous fields. This limitation has been overcome by deploying RDD [111]. 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.

Chapter 4

Method of distributions for two-phase flow in heterogeneous porous media

4.1 Introduction

The uncertainty propagation of stochastic two-phase flow in heterogeneous porous media is often challenging due to the non-linearity of transport equation (i.e., Buckley-Leverett problem) and the necessity of coupling between the saturation and velocity vector. The statistical moment equations (SMEs) based approach relying on low-order approximation has been developed for computing first two moments of the water saturation in stochastic transport problem [120, 57, 61]. This approach formulated on the Lagrangian coordinate (mean streamline coordinate) successfully couples the saturation and velocity vector in terms of travel time. But it still assumes that the distribution of travel time is known. The spectral methods using physically motivated transforms have been developed to build efficient probabilistic collocation methods for nonlinear flow in porous media [58, 59]. For sampling based approach, the multilevel Monte Carlo (MLMC) has been used to speed up the MCS in immiscible oil-water displacement problem [67].

Accelerated versions of MCS, such as multi-level Monte Carlo, provide significant speed-up in computing the statistical moments of system states and associated quantities of interest (QoIs) [67], but might become slower than standard MCS when used to estimate full PDFs/CDFs [98, 97]. Other

direct uncertainty quantification techniques, such as polynomial chaos expansions and stochastic collocation on sparse grids, are also not guaranteed to outperform MCS when the number of input parameters (the so-called stochastic dimension) is large or the governing PDEs are highly nonlinear [4]. Models of multiphase flow in heterogeneous porous media, which are the focus of our study, fall under this category.

We posit that the method of distributions [96] has a potential to provide a computationally efficient alternative to MCS in this setting. By providing a deterministic equation for the spatiotemporal evolution of the PDF or CDF of a state variable, it has proved to be up to an order of magnitude faster than MCS when used to compute the PDF/CDF of hydraulic head in confined heterogeneous aquifers as proved in Chapters 2 and 3, and to yield accurate approximations of the PDF/CDF of the concentration of solutes undergoing geochemical transformations during their migration in the subsurface [95, 7, 8]. While the former class of problems is linear—see, also, [28] and [9] for the PDF solutions of an advection-dispersion equation with uncertain parameters—and the latter nonlinear, the state variable (hydraulic head or concentration) in both cases is smooth.

A complete description of uncertain output via PDF or CDF estimates is important to elaborate the complexity of stochastic saturation field in two-phase flow problem. Among a plethora of methods aiming for uncertainty quantification, the method of distributions [96] provides such information by solving the deterministic equations for PDF or CDF of the QoI. The computational efficiency of this method over other approaches has been proved in Chapters 2 and 3. However, apart from flow problems discussed in previous chapters, complete distribution-based approach for two-phase flow is plagued by non-linearity of problem as well as multi-dimensionality of velocity vector.

Shocks and discontinuous state variables typical of multiphase problems, e.g., fluid saturation whose dynamics is described by the Buckley-Leverett equation, pose a challenge for the method of distributions (and other uncertainty quantification techniques). It requires either the analytical computation of the shock dynamics [105] or the introduction of the so-called kinetic defect that has to be inferred from data [10]. For highly heterogeneous subsurface environments, in which the streamlines are defined largely by geology rather than flow conditions and, hence, remains frozen in time, a version of the method of distributions named FROST [44, 43] exhibits remarkable accuracy and efficiency [35, 36]. Its input is the distribution of travel times, whose Monte Carlo computation requires a large number of flow simulations and streamline tracings. To speed-up FROST further, we eliminate the need for MCS by replacing it with a deterministic CDF equation for travel times.

We start by formulating a two-phase immiscible flow problem with uncertain inputs in Section 4.2. Its treatment with the original FROST method is reviewed in Section 4.3, followed by the derivation of a CDF equation for travel time in Section 4.4. A numerical implementation of the new CDF-FROST method is presented in Section 4.5. Numerical experiments presented in Section 4.6 serve

to demonstrate our method's accuracy and efficiency vis-à-vis MCS. Main findings and conclusions drawn from our study are summarized in Section 4.7.

4.2 Problem formulation

We consider immiscible displacement of a non-wetting fluid (e.g., DNAPL) by a wetting fluid (e.g., water) in a d-dimensional heterogeneous porous medium, $\Omega \subset \mathbb{R}^d$. Both fluids are incompressible, with respective viscosities $\mu_{\rm nw}$ and $\mu_{\rm w}$. Their mobilities in the porous medium of intrinsic permeability $k(\mathbf{x})$ are defined as

$$\lambda_{\rm nw} = \frac{k(\mathbf{x})k^{\rm r}_{\rm nw}(S_{\rm nw})}{\mu_{\rm nw}}, \qquad \lambda_{\rm w} = \frac{k(\mathbf{x})k^{\rm r}_{\rm w}(S_{\rm w})}{\mu_{\rm w}}, \tag{4.1}$$

where $k^{\rm r}_{\rm nw}(S_{\rm nw})$ and $k^{\rm r}_{\rm w}(S_{\rm w})$ are the saturation-dependent relative permeabilities of the porous medium with respect to the non-wetting fluid and the wetting fluid, respectively; and $S_{\rm nw}(\mathbf{x},t)$ and $S_{\rm w}(\mathbf{x},t)$ are the saturations of these two phases, such that $S_{\rm nw} + S_{\rm w} = 1$.

Neglecting the effects of capillary pressure and gravity, the Darcy fluxes $\mathbf{q}_{\rm nw}(\mathbf{x},t)$ and $\mathbf{q}_{\rm w}(\mathbf{x},t)$ of non-wetting and wetting phases are related to the gradient of pressure in both phases, $p(\mathbf{x},t)$, by

$$\mathbf{q}_{\rm nw} = -\lambda_{\rm nw}\nabla p, \qquad \mathbf{q}_{\rm w} = -\lambda_{\rm w}\nabla p, \qquad \mathbf{x}\in\Omega, \quad t>0. \tag{4.2}$$

The total Darcy flux $\mathbf{q}_{\rm tot} = \mathbf{q}_{\rm nw} + \mathbf{q}_{\rm w}$ satisfies the continuity condition, $-\nabla\cdot\mathbf{q}_{\rm tot} + g = 0$, where $g(\mathbf{x},t)$ represents fluid sources and sinks. Neglecting the compressibility of the porous medium, and accounting for (4.2), this yields a pressure equation

$$\nabla\cdot(\lambda_{\rm tot}\nabla p) + g = 0, \qquad \mathbf{x}\in\Omega, \quad t>0, \tag{4.3}$$

where $\lambda_{\rm tot}(S_{\rm w}) = \lambda_{\rm nw}(S_{\rm w}) + \lambda_{\rm w}(S_{\rm w})$ is the total mobility. This equation is subject to boundary conditions either controlled by pressure

$$p = p_{inj} \quad \text{for} \quad \mathbf{x}\in\Gamma_{inj} \qquad \text{and} \qquad p = p_{prod} \quad \text{for} \quad \mathbf{x}\in\Gamma_{prod}, \tag{4.4}$$

or total Darcy flux

$$\mathbf{q}_{\rm tot}\cdot\mathbf{n} = q_{inj} \quad \text{for} \quad \mathbf{x}\in\Gamma_{inj} \qquad \text{and} \qquad \mathbf{q}_{\rm tot}\cdot\mathbf{n} = q_{prod} \quad \text{for} \quad \mathbf{x}\in\Gamma_{prod}, \tag{4.5}$$

which are defined on the injection (Γ_{inj}) and production (Γ_{prod}) segments of the boundary $\partial\Omega = \Gamma_{prod}\cup\Gamma_{inj}$ of the flow domain Ω. Here, p_{inj} and p_{prod} are the pressure imposed along the boundary

segment when boundary is controlled by prescribed pressure. q_{inj} and q_{prod} are the normal component of the Darcy flux through rate-control boundary. $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to $\partial\Omega$.

Conservation of mass of, e.g., the wetting phase gives rise to a transport equation for the saturation $S_{\mathrm{w}}(\mathbf{x}, t)$,

$$\phi(\mathbf{x})\frac{\partial S_{\mathrm{w}}}{\partial t} + \mathbf{q}_{\mathrm{tot}} \cdot \nabla f_{\mathrm{w}}(S_{\mathrm{w}}) = g_w. \tag{4.6}$$

where g_w is the source (sink) term for wetting phase and ϕ is the porosity. The fractional flow function of wetting phase, $f_{\mathrm{w}}(S_{\mathrm{w}})$, is defined as λ_w/λ_{tot}. This equation is subject to following initial and boundary conditions

$$S_{\mathrm{w}}(\mathbf{x}, 0) = S_{\mathrm{w}}^{\mathrm{ir}}, \quad \mathbf{x} \in \Omega; \qquad S_{\mathrm{w}}(\mathbf{x}, t) = 1 - S_{\mathrm{nw}}^{\mathrm{ir}}, \quad \mathbf{x} \in \Gamma_{\mathrm{inj}}, \quad t > 0. \tag{4.7}$$

Here, $S_{\mathrm{w}}^{\mathrm{ir}}$ and $S_{\mathrm{nw}}^{\mathrm{ir}}$ are the irreducible saturations of the wetting and non-wetting fluids, respectively.

Equations (4.3)–(4.7) govern the spatiotemporal evolution of the two state variables, fluid pressure $p(\mathbf{x}, t)$ and saturation of the wetting phase $S_{\mathrm{w}}(\mathbf{x}, t)$. With the sole exception of intrinsic permeability $k(\mathbf{x})$, values of all the parameters in these equations are assumed to be known with certainty. Permeability $k(\mathbf{x})$ is modeled as a second-order stationary multivariate log-normal 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 $\mathbf{x}$ and $\mathbf{y}$ in Ω. Uncertainty (randomness) in the model input, $k(\mathbf{x})$, renders the model prediction, $S_{\mathrm{w}}(\mathbf{x}, t)$, uncertain (random) as well. Our goal is to compute the probability of saturation S_{w}, at any space-time point $(\mathbf{x}, t)$, not exceeding a given value $s \in [S_{\mathrm{w}}^{\mathrm{ir}}, 1 - S_{\mathrm{nw}}^{\mathrm{ir}}]$, $\mathbb{P}\{S_{\mathrm{w}}(\mathbf{x}, t) \leq s\}$. The latter is the definition of the single-point CDF of S_{w}, i.e., $\mathbb{P}\{S_{\mathrm{w}}(\mathbf{x}, t) \leq s\} \equiv F_{S_{\mathrm{w}}}(s; \mathbf{x}, t)$.

4.3 FROST method

Consider a streamline $\mathbf{x}_{\mathrm{sl}}(\tau)$ originating at point $\boldsymbol{\xi}$ at TOF $\tau = 0$. Given the total Darcy flux $\mathbf{q}_{\mathrm{tot}}(\mathbf{x}, t)$, it is defined implicitly by

$$\mathbf{x}_{\mathrm{sl}}(\tau, t) = \boldsymbol{\xi} + \int_0^{\tau} \mathbf{q}_{\mathrm{tot}}(\mathbf{x}_{\mathrm{sl}}(t'), t)\mathrm{d}t'. \tag{4.8}$$

We parameterize this streamline by a natural coordinate r, such that the distance $\mathrm{d}r$ traveled by a particle along this streamline during the TOF interval $\mathrm{d}\tau$ is $\mathrm{d}r = |\mathbf{q}_{\mathrm{tot}}|\mathrm{d}\tau$. A collection

of streamlines forms a streamtube with a variable cross-sectional area $A(r)$; the volumetric flow rate in this streamtube is $Q_{\text{tube}} = |\mathbf{q}_{\text{tot}}|A(r)$. Considering the following fundamental property $\partial Q_{\text{tube}}/\partial r = 0$ derived by the definition of streamtube, Q_{tube} only depends on the time.

With these definitions, we rewrite the saturation equation (4.6) in the streamline coordinate system,

$$\phi(r)A(r)\frac{\partial S_{\text{w}}(r,t)}{\partial t} + Q_{\text{tube}}\frac{\partial f_{\text{w}}(S_{\text{w}}(r,t))}{\partial r} = 0, \qquad r > 0, \quad t > 0. \tag{4.9}$$

Next, we define the cumulative injection volume Q and the cumulative pore volume V as

$$Q(t) = \int_0^t Q_{\text{tube}}(t')\mathrm{d}t', \qquad V(r) = \int_0^r \phi(r')A(r')\mathrm{d}r'. \tag{4.10}$$

Then, with Q playing the role of time and V of spatial coordinate, $S_{\text{w}}(r,t) \mapsto S_{\text{w}}(V,Q)$ and (4.9) takes the form of a one-dimensional Buckley-Leverett equation,

$$\frac{\partial S_w}{\partial Q} + \frac{\partial f_w(S_w)}{\partial V} = 0, \qquad V > 0, \quad Q > 0. \tag{4.11}$$

The initial and boundary conditions in (4.7) map onto

$$S_{\text{w}}(V,0) = S_w^{ir}, \quad S_{\text{w}}(0,Q) = 1 - S_{nw}^{ir}. \tag{4.12}$$

The frozen streamline assumption implies the fixed total flux $\mathbf{q}_{\text{tot}}$ direction. With this assumption, cross sectional area $A(r)$ becomes time-independent and can be expressed as $A(r) = Q_{tube}(t = 0)/|\mathbf{q}_{\text{tot}}(r', t = 0)|$ using total flux $\mathbf{q}_{\text{tot}}$ at $t = 0$. Hence, the ratio $Z = V/Q$ becomes the ratio between the time of flight (TOF) at time $t = 0$, $\tau_0(\mathbf{x})$, and the equivalent injection time (EIT), $T_{\text{inj}}(\mathbf{x}, t)$,

$$Z(\mathbf{x},t) = \frac{\tau_0}{T_{\text{inj}}}, \quad \tau_0(\mathbf{x}) = \int_0^{r(\mathbf{x})} \frac{\phi(r')\,\mathrm{d}r'}{|\mathbf{q}_{\text{tot}}(r',0)|}, \quad T_{\text{inj}}(\mathbf{x},t) = \int_0^t \frac{Q_{tube}(t')}{Q_{tube}(t=0)}\mathrm{d}t'. \tag{4.13}$$

Regardless of spatial heterogeneity, the EIT is approximately uniform in space and can be approximated by its mean value $\bar{T}_{\text{inj}}$ from a relatively few MC realizations, i.e., $T_{\text{inj}}(\mathbf{x},t) \approx T_{\text{inj}}(t) \approx \bar{T}_{\text{inj}}(t)$ [43]. This approximation allows one to express the CDF of the wetting-phase saturation, $F_{S_{\text{w}}}(s;\mathbf{x},t)$, in terms of the CDF $F_Y(y;\mathbf{x})$ of the logarithm of TOF, $Y(\mathbf{x}) = \ln\tau_0(\mathbf{x})$,

$$F_{S_{\text{w}}}(s;\mathbf{x},t) = 1 - F_Y(z;\mathbf{x}), \qquad z = \ln[S_{\text{w}}^{-1}(s)\bar{T}_{\text{inj}}(t)]. \tag{4.14}$$

The deterministic inverse mapping, $S_{\mathrm{w}}^{-1}(s)$, is provided by the analytical solution of (4.11),

$$S_{\mathrm{w}}^{-1}(s) = \begin{cases} \infty, & s < S_{\mathrm{w}}^{\mathrm{ir}} \\ f'_{\mathrm{w}}(s^*), & s \in (S_{\mathrm{w}}^{\mathrm{ir}}, s^*) \\ f'_{\mathrm{w}}(s), & s \in (s^*, 1 - S_{\mathrm{nw}}^{\mathrm{ir}}) \\ 0, & s > 1 - S_{\mathrm{nw}}^{\mathrm{ir}}, \end{cases} \tag{4.15}$$

where $f'_{\mathrm{w}}(\cdot)$ designates the derivative of $f_{\mathrm{w}}(\cdot)$, and s^* satisfies the Rankine-Hugoniot jump condition,

$$f'_w(s^*) = \frac{f_w(s^*) - f_w(S_{\mathrm{w}}^{\mathrm{ir}})}{s^* - S_{\mathrm{w}}^{\mathrm{ir}}}. \tag{4.16}$$

We designate the original version of FROST by MCS-FROST to emphasize its reliance on MCS to estimate both $\bar{T}_{\mathrm{inj}}(t)$ and $F_Y(y; \mathbf{x})$. Estimation of the latter is particularly expensive because it requires a large number of MC realizations, N_{MC}, to converge (in our experiments, $N_{\mathrm{MC}} \geq 5000$). Our version, CDF-FROST, replaces this computational bottleneck with a numerical solution of the deterministic equation for $F_Y(y; \mathbf{x})$.

4.4 CDF Equation for TOF

TOF, also known as travel time, is important in its own right since it provides useful information for flow visualization, models' ranking, and optimization of well placement or operation. As mentioned above, TOF $\tau(r)$ at time t is defined as the time required for a particle to arrive at location r along a given streamline,

$$\tau(r, t) = \int_0^r \frac{\phi(r')\, \mathrm{d}r'}{|\mathbf{q}_{\mathrm{tot}}(r', t)|}. \tag{4.17}$$

For any point $\mathbf{x}$ in the flow domain Ω, TOF from an injection point along the boundary Γ_{inj}, $\tau(\mathbf{x})$, satisfies a differential equation [83]

$$\mathbf{q}_{\mathrm{tot}}(\mathbf{x}, t) \cdot \nabla \tau(\mathbf{x}) = \phi(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{4.18}$$

subject to Dirichlet boundary condition

$$\tau(\mathbf{x}) = 0, \qquad \mathbf{x} \in \Gamma_{\mathrm{inj}}. \tag{4.19}$$

Randomness in the intrinsic permeability $k(\mathbf{x})$ translates into randomness of $\mathbf{q}_{\text{tot}}(\mathbf{x}, t)$, which here serves as an input. We proceed to derive a deterministic equation for the CDF $F_\tau(\mathcal{T}; \mathbf{x})$ of $\tau(\mathbf{x})$ from (4.18) and (4.19).

4.4.1 Derivation of CDF Equation

Consider a function $\Pi(\mathcal{T}, \mathbf{x}) \equiv \mathcal{H}(\mathcal{T} - \tau(\mathbf{x}))$, where $\mathcal{H}(\cdot)$ denotes the Heaviside function and $\mathcal{T}$ is a deterministic value that the random TOF τ at point $\mathbf{x}$ can take. Its ensemble mean over all possible values of the random variables τ is

$$\mathbb{E}\{\Pi(\mathcal{T}, \tau(\mathbf{x}))\} = F_\tau(\mathcal{T}; \mathbf{x}) \equiv \mathbb{P}\{\tau(\mathbf{x}) \leq \mathcal{T}\}. \tag{4.20}$$

Multiplying both sides of (4.18) with $\partial\Pi/\partial\mathcal{T}$, and noting that $\nabla\Pi = -(\partial\Pi/\partial\mathcal{T})\nabla\tau$, yields an equation for $\Pi(\mathcal{T}, \mathbf{x})$,

$$\mathbf{q}_{\text{tot}} \cdot \nabla\Pi + \phi \frac{\partial \Pi}{\partial \mathcal{T}} = 0. \tag{4.21}$$

We use the Reynolds decomposition to represent the random quantities $\mathbf{q}_{\text{tot}}$ and Π as the sum of their ensemble means and zero-mean fluctuation around them, $\mathbf{q}_{\text{tot}} = \bar{\mathbf{q}}_{\text{tot}} + \mathbf{q}'_{\text{tot}}$ and $\Pi = F_\tau + \Pi'$. (Throughout this manuscript, we use $\mathbb{E}\{\cdot\}$, $\bar{\cdot}$, and $\langle\cdot\rangle$ interchangeably to represent the ensemble mean.) Then, the ensemble average of (4.21) yields an unclosed equation for $F_\tau(\mathcal{T}; \mathbf{x})$,

$$\bar{\mathbf{q}}_{\text{tot}} \cdot \nabla F_\tau + \phi \frac{\partial F_h}{\partial T} + \langle \mathbf{q}'_{\text{tot}} \cdot \nabla\Pi' \rangle = 0. \tag{4.22}$$

A closure approximation for the unknown cross-correlation term $\langle \mathbf{q}'_{\text{tot}} \cdot \nabla\Pi' \rangle$ is necessary to render (4.22) computable.

The large-eddy-diffusivity closure exhibits good accuracy and robustness for advection-reaction problems [95, 102]. However, it relies on Green's functions, which are computationally expensive unless given analytically. Instead, we use the moments-preserving closure, a generalization of the interaction by exchange with the mean approximation, [9, 8, 118, 45],

$$\langle \mathbf{q}'_{\text{tot}} \cdot \nabla\Pi' \rangle \approx [\alpha(\mathbf{x})(\mathcal{T} - \bar{\tau}) + \beta(\mathbf{x})] \frac{\partial F_\tau}{\partial T}, \tag{4.23}$$

where $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are closure variables determined below. This yields a closed $(d+1)$-dimensional CDF equation

$$\bar{\mathbf{q}}_{\text{tot}} \cdot \nabla F_\tau + [\alpha(\mathcal{T} - \bar{\tau}) + \beta + \phi] \frac{\partial F_\tau}{\partial T} = 0. \tag{4.24}$$

4.4.2 Coordinate Transformation

To maximize the computational efficiency of the CDF method, we introduce a new transformation of coordinates in the frame of reference defined by the mean streamlines $\bar{\mathbf{x}}_{\rm sl}(\tau)$. Equation (4.17) establishes the duality between TOF along a mean trajectory, τ_{mean}, and the position along the mean streamline traveled during that time, r. Using this duality and taking the ensemble mean of (4.8) yields

$$\bar{\mathbf{x}}_{\rm sl}(\tau_{mean}) = \bar{x}_{sl,0} + \int_0^{\tau_{mean}} \bar{\mathbf{q}}_{\rm tot}(s')\mathrm{d}s'. \tag{4.25}$$

This relation allows one to represent any point $\mathbf{x} \in \Omega$ in the Cartesian grid in terms of the distance r along the closest mean streamline. In the coordinate system spanned by r, $\bar{\mathbf{q}}_{\rm tot} \cdot \nabla = \bar{q}_{\rm tot}\partial/\partial r$, where $\bar{q}_{\rm tot} \equiv |\bar{\mathbf{q}}_{\rm tot}|$ denotes the magnitude of the mean velocity vector $\bar{\mathbf{q}}_{\rm tot}$; $F_\tau(\mathcal{T};\mathbf{x}) \mapsto F_\tau(\mathcal{T};r)$; and (4.24) takes the form,

$$\bar{q}_{\rm tot}\frac{\partial F_\tau}{\partial r} + [\alpha(r)(\mathcal{T}-\bar{\tau}) + \beta(r) + \phi(r)]\frac{\partial F_\tau}{\partial \mathcal{T}} = 0. \tag{4.26}$$

Since $\bar{q}_{\rm tot}$ is constant along a streamline coordinate r by the definition of streamline and incompressibility, we rewrite this CDF equation in a "conservative" form,

$$\frac{\partial(\bar{q}_{\rm tot}F_\tau)}{\partial r} + \frac{\partial(UF_\tau)}{\partial \mathcal{T}} = \alpha F_\tau, \qquad U(r,\mathcal{T}) \equiv \alpha(\mathcal{T}-\bar{\tau}) + \beta + \phi, \tag{4.27}$$

which is more conducive to the subsequent numerical treatment. This CDF equation is subject to the boundary conditions derived from (4.19),

$$F_\tau(T; r=0) = 1, \quad T \geq 0. \tag{4.28}$$

The general properties of a CDF provide the remaining boundary conditions,

$$F_\tau(0;r) = 0, \qquad F_\tau(\tau_{\max};r) = 1. \tag{4.29}$$

Following [9] and [45], we determine expressions for the closure variables $\alpha(r)$ and $\beta(r)$ in (4.27) by enforcing the consistency between the moments $\bar{\tau}$ and σ_τ^2 obtained alternatively by integration of (4.27) and by either using MCS or solving the statistical moment equations (SMEs) derived in A and C. We start by recalling that the mean, $\bar{\tau}(r)$, and variance, $\sigma_\tau^2(r)$, of the TOF $\tau(r)$ with CDF

$F_\tau(\mathcal{T};r)$ are

$$\bar{\tau} = \tau_{\max} - \int_0^{\tau_{\max}} F_\tau(\mathcal{T};r)\mathrm{d}\mathcal{T}, \quad \sigma_\tau^2 = \tau_{\max}^2 - 2\int_0^{\tau_{\max}} \mathcal{T}F_\tau(\mathcal{T};r)\mathrm{d}\mathcal{T} - \bar{\tau}^2, \tag{4.30}$$

where $\tau_{\max}$ is the maximal value the random TOF τ can take; in the absence of additional information, one can set $\tau_{\max} = \infty$.

Integrating (4.26) over $\mathcal{T}$ and employing (4.30) and (4.29), we can derive following equation for $\bar{\tau}$

$$q_{tot}\frac{\partial\bar{\tau}}{\partial r} - \beta(r) - \phi = 0. \tag{4.31}$$

Similarly, multiplying both sides of (4.26) by $\mathcal{T}$ and integrating the resulting equation over $\mathcal{T}$ yields

$$q_{tot}\frac{\partial\sigma_\tau^2}{\partial r} - 2\alpha(\mathbf{x})\sigma_\tau^2 + 2\bar{\tau}[q_{tot}\frac{\partial\bar{\tau}}{\partial r} - \beta(r) - \phi] = 0$$

or, accounting for (4.31),

$$q_{tot}\frac{\partial\sigma_\tau^2}{\partial r} - 2\alpha(\mathbf{x})\sigma_\tau^2 = 0 \tag{4.32}$$

Imposition of the equivalency between the mean ($\bar{\tau}$) and variance (σ_τ^2) equations derived by the CDF method, (4.31) and (4.32), and the moments computed by SMEs or MCS yields expressions for the closure variables $\alpha(r)$ and $\beta(r)$ as

$$\alpha(r) = \frac{q_{\mathrm{tot}}}{2\sigma_\tau^2}\frac{\partial\sigma_\tau^2}{\partial r}, \qquad \beta(r) = q_{\mathrm{tot}}\frac{\partial\bar{\tau}}{\partial r} - \phi. \tag{4.33}$$

We use the SMEs to compute $\bar{\tau}$ and σ_τ^2 becomes this strategy outperforms MCS in terms of computational cost for the same accuracy [61].

The solution to the two-dimensional boundary-value problem (BVP) (4.27)–(4.29) yields the CDF of travel time along each mean streamline. Let N_{sl}, N_r, and $N_{\mathcal{T}}$ denote the total number of streamlines traced in the domain Ω, the number of grid cells along each streamline, and the number of grid cells used to discretize the interval $[0, \tau_{\max}]$, respectively. Then, the computational complexity of solving this BVP is $\mathcal{O}(N_{\mathrm{sl}} \cdot N_r^3 \cdot N_{\mathcal{T}}^3)$. We contrast this with the computational complexity of solving the corresponding CDF equation in the Cartesian coordinate system, i.e., solving (4.24) directly. Even if the flow domain Ω were two-dimensional, its discretization with N_x and N_y cells along each coordinate would yield the computational complexity of solving (4.24) of $\mathcal{O}(N_x^3 \cdot N_y^3 \cdot N_{\mathcal{T}}^3)$. This comparison highlights the main advantage of our coordinate transformation: it enables the computation time that is proportional to the number of streamlines N_{sl} that, unlike the Cartesian-grid calculations, scales linearly with the size of Ω [99].

4.5 Numerical Implementation of CDF-FROST

In our approach, the CDF of saturation is obtained in three steps (Figure 4.1). The first step is to obtain the mean and variance of TOF, $\bar{\tau}(r)$ and $\sigma_\tau^2(r)$, by solving the SMEs (C). The second step involves numerical solution of BVP (4.27)–(4.29) to compute the travel-time CDF $F_\tau(\mathcal{T};r)$. The latter provides an input for the analytical framework of FROST to evaluate the saturation CDF $F_{S_\mathrm{w}}(s;\mathbf{x},t)$.

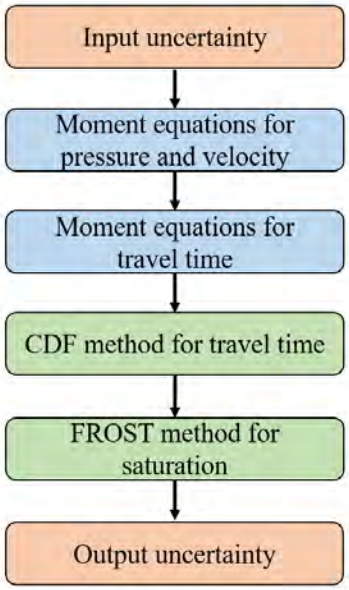

Figure 4.1: Workflow of proposed CDF-FROST method

The first step is carried out with the research code of [61]. The computational cost of the third step is negligible, since it requires no numerical solution of differential equations. The second step, however, poses several challenges for standard numerical methods for hyperbolic partial-differential equations. The discontinuity of the boundary condition in (4.28) precludes the use of a high-order scheme without inducing spurious oscillations. The non-smoothness of the coefficients U and α in (4.27) requires a special treatment to achieve a desired accuracy. The monotonicity of TOF $\tau(r)$ along a streamline necessitates the deployment of a nonuniform grid in the $\mathcal{T}$ coordinate to improve efficiency. To resolve these issues, we deploy three numerical techniques in this step: pseudo-time stepping, a flux-limited method, and an exponential grid spacing.

Pseudo-time stepping

A pseudo-time stepping or pseudo-transient continuation method is particularly appropriate

for the steady-state equation with non-smooth coefficients [47, 34]. It starts with introducing the fictitious or pseudo-time t'. Dropping the subscript τ for simplicity, (4.27) can be transformed into the following time dependent equation

$$\frac{\partial F}{\partial t'} + \frac{\partial (q^{sl} F)}{\partial r} + \frac{\partial (UF)}{\partial T} = vF, \quad F(t'=0) = F_0, \tag{4.34}$$

where F_0 is the initial condition. The main objective of the pseudo-time stepping method is to seek the converged solution $F^* = \lim_{t' \to \infty} F$ which is equivalent to the solution of (4.27) regardless of initial condition.

When (4.34) is discretized within the spatial domain such that $r_i = \sum_{k=1}^{i-1} \Delta r_k + \Delta r_i/2$, $T_j = \sum_{k=1}^{j-1} \Delta T_k + \Delta T_j/2$ where Δr and ΔT are the spatial step, a general explicit conservative scheme with forward time integration can be written as

$$F_{i,j}^{n+1} = (1+v) F_{i,j}^n - \lambda_{r,i} (f_{i+1/2,j}^n - f_{i-1/2,j}^n) - \lambda_{T,j} (g_{i,j+1/2}^n - g_{i,j-1/2}^n). \tag{4.35}$$

The subscripts n and $n+1$ represent the old and new time steps, respectively, and $\lambda_{r,i} = \Delta t'/\Delta r_i$, $\lambda_{r,i} = \Delta t'/\Delta T_i$, where $\Delta t'$ is time-step size. For a finite volume scheme, $F_{i,j}^n$ and $F_{i,j}^{n+1}$ are cell averaged conservative variables on the computational cell defined on the interval $[r_{i-1/2}, r_{i+1/2}] \times [T_{i-1/2}, T_{i+1/2}]$ where $r_{i\pm 1/2} = r_i \pm \Delta r_i$ and $T_{j\pm 1/2} = T_j \pm \Delta T_j$. f and g are respectively the r and T direction components of numerical flux.

Flux limited scheme

Among a plethora of schemes for approximating numerical flux, Roe's first-order upwind method guarantees non-oscillatory solution near discontinuities. Since q_{tot}^{sl} is always greater than 0 along the streamline, flux term f can be expressed as $f_{i+1/2,j}^n = q_{i,j}^{sl} F_{i,j}^n$. For flux term g, the first-order upwind method fails to impose the T space boundary conditions (4.29) properly when the boundaries $T = 0$ or $T = T_{max}$ are downstream boundaries (i.e., outgoing flow direction to the boundary). In order to construct a numerical scheme that maintains high-order accuracy near boundaries, while producing the monotone results at discontinuities, the flux limited scheme is implemented to construct flux term g. The flux limited method represents the flux as a linear combination of low-order (i.e., Roe first-order upwind) and high-order (i.e., Lax-Wendroff) methods as follows

$$g_{i,j+1/2}^n = \begin{cases} U_{i,j} F_{i,j}^n + \frac{1}{2} a_{i,j+1/2} (1 - \lambda_T a_{i,j+1/2}) \phi_{i,j}^+ (F_{i,j+1}^n - F_{i,j}^n), & a_{i,j+1/2} > 0 \\ U_{i,j+1} F_{i,j+1}^n + \frac{1}{2} a_{i,j+1/2} (1 + \lambda_T a_{i,j+1/2}) \phi_{i,j+1}^- (F_{i,j+1}^n - F_{i,j}^n), & a_{i,j+1/2} < 0 \end{cases} \tag{4.36}$$

where wave speed $a_{i,j+1/2}$ is defined as

$$a^n_{i,j+1/2} = \begin{cases} \frac{U_{i,j+1}F^n_{i,j+1} - U_{i,j}F^n_{i,j}}{F^n_{i,j+1} - F^n_{i,j}} & F_{i,j} \neq F_{i,j+1} \\ U_{i,j} & F_{i,j+1} = F_{i,j} \end{cases} \tag{4.37}$$

Here, flux-limiter $\phi^{\pm}$ is the function of $\theta^{\pm}$ which is the ratio of successive gradients indicating the smoothness of the solution. The minmod flux-limiter function $\phi(\theta) = \max[0, \min(1, \theta)]$ is employed in the present study. The ratio $\theta^{\pm}_{i,j}$ can be considered as the measure of smoothness of the data near cell (i, j) and defined as

$$\theta^+_{i,j} = \frac{F_{i,j} - F_{i,j-1}}{F_{i,j+1} - F_{i,j}}, \quad \theta^-_{i,j} = \frac{F_{i,j+1} - F_{i,j}}{F_{i,j} - F_{i,j-1}}. \tag{4.38}$$

The numerical solution $F^n_{i,j}$ is updated iteratively using (4.35) until it converges to the steady state solution. The convergence criteria is given as

$$\max_{i,j} |F^{n+1}_{i,j} - F^n_{i,j}| < 10^{-8} \tag{4.39}$$

Exponential grid spacing

Considering the original definition of travel time (4.17), the travel time increases monotonically along the streamline. Defining the CDF equation (4.27) along the mean streamline coordinate, every points in the streamline should share the same T-direction grid system. As can be seen in Figure 4.2, the fine grid size ΔT_j is required for accurate description of CDF when the distance from the origin r is small. While, the coarser grid could be enough to represent the travel time CDF at large r. To use this fact for further speed up of the computations, we introduce exponential grid spacing instead of uniform grid size.

In the exponential grid spacing system, the grid size $\Delta T_j = \Delta T_0 \xi^{j-1}$ at cells j increases with fixed expansion rate ξ where ΔT_0 is the initial time. The location of each cell T_j is defined as $T_j = T_0 + \Delta T_0 \frac{1-\xi^{j-1}}{1-\xi}$ where T_0 is the location of first cell $j = 0$. In all numerical experiments in this thesis, the expansion rate ξ is fixed as 1.02. Figure 4.3 shows the numerical solutions of (4.27) obtained by two different grid systems with same grid number ($N_T = 150$). It is shown that the exponential grid spacing would be computationally cheaper than uniform grid to get the same level of accuracy.

The third and last step is to obtain the saturation CDF F_{S_w} from the travel time CDF F_τ using FROST method. The logarithm is introduced for better characterization of the distribution. The distribution of initial logarithmic TOF $F_{\log_{\tau_0}}$ can be computed from the results of second

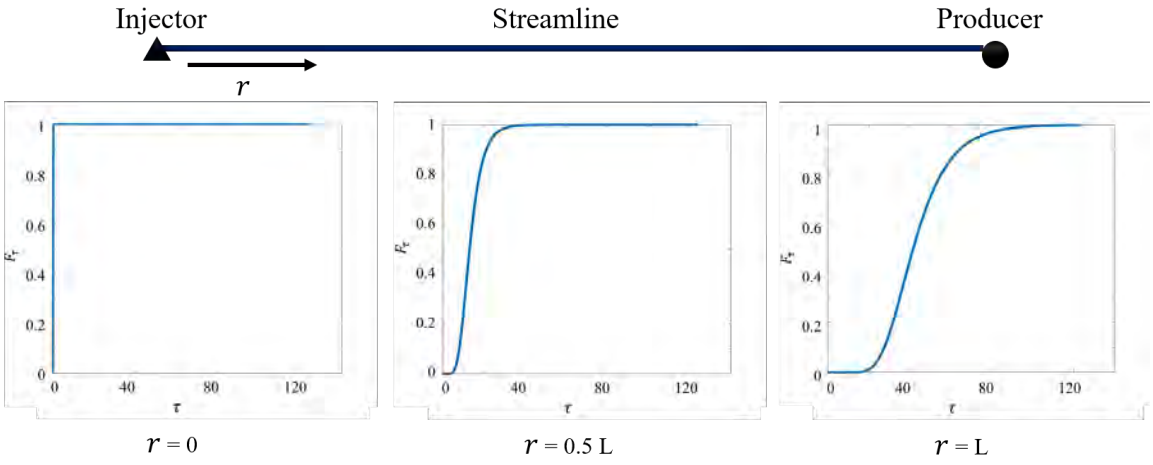

Figure 4.2: Evolution of travel time CDF with distance r from the starting point of mean streamline. L is the total length of mean streamline.

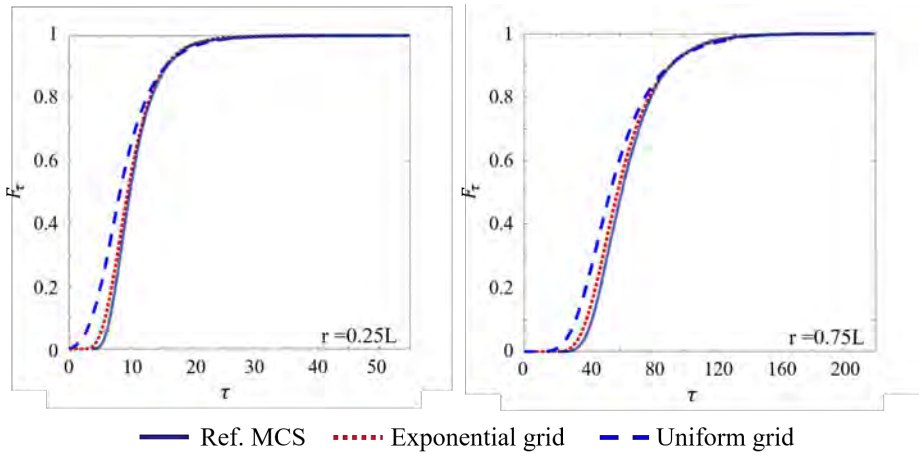

Figure 4.3: Travel time CDFs F_τ computed with reference MCS and the CDF method with uniform and exponential grid system at two selected distance r along the mean streamline for quarter-five spot well configuration. L is the total length of mean streamline.

step. It provides the input of (4.14) which finally produces the distribution of saturation F_{S_w}. The computational cost of last step is relatively cheap compare to the first two steps, since it does not require any numerical solution of PDEs which is most computationally expensive part in CDF-FROST method. The overall procedure of proposed CDF-FROST method is illustrated in Figure 4.1.

4.6 Numerical Experiments

We use two sets of numerical experiments to demonstrate the accuracy, robustness and versatility of CDF-FROST. In both cases, the wetting phase is injected into a two-dimensional flow domain Ω, which is initially filled with non-wetting phase. The viscosities of the non-wetting and wetting phases are $\mu_{\mathrm{nw}} = 2\mathrm{cp}$ and $\mu_{\mathrm{w}} = 1\mathrm{cp}$, respectively, thus the viscosity ratio $M = \mu_{\mathrm{nw}}/\mu_{\mathrm{w}} = 2$. The quadratic model is used for relative permeabilities $k^{\mathrm{r}}_{\mathrm{w}} = (S_{\mathrm{w}} - S^{\mathrm{ir}}_{\mathrm{w}})^2$ and $k^{\mathrm{r}}_{\mathrm{nw}} = (1 - S_{\mathrm{w}})^2$, with $S^{\mathrm{ir}}_{\mathrm{w}} = 0.1$ and $S^{\mathrm{ir}}_{\mathrm{nw}} = 0$. The porosity ϕ is assumed to be uniform $\phi = 0.1$ over the computational domain.

The spatial domain Ω, a square of length $L = 1000$ m, is discretized with a staggered 80×80 grid, and the number of grid points along the $\mathcal{T}$ coordinate is set to $N_{\mathcal{T}} = 150$. The number of mean streamlines launched in the domain is $N_{\mathrm{sl}} = 300$. The log-permeability field $\kappa(\mathbf{x}) \equiv \ln k$ is modeled as a second-order stationary multi-variate Gaussian field with zero mean ($\bar{\kappa} = 0$), variance $\sigma^2_\kappa = 1$, an isotropic exponential covariance function $C_\kappa = \sigma^2_\kappa \exp(-|\mathbf{x} - \mathbf{y}|/\ell_\kappa)$, and dimensionless (normalized with the domain size L) correlation length $\ell_\kappa = 0.1$.

Our numerical experiments mimic two representative flow scenarios. The first is the mean uniform flow driven by a line injection of a wetting fluid at the left boundary, while the non-wetting fluid is extracted at the right boundary. The second is a quarter-five spot problem with one injector at lower left corner and one producer at upper right corner. In both scenarios, the total injection rate is 100 m^3/day and the extraction wells operate at constant pressure 100 Pa.

We compare our estimates of travel-time CDF $F_\tau(\mathcal{T}; \mathbf{x})$ and saturation CDF $F_{S_{\mathrm{w}}}(s; \mathbf{x}, t)$ with their counterparts computed via high-resolution MCS. The latter employ the standard Pollock's method [74] and the finite-volume solver implementing an IMPES scheme [24] to compute, respectively, the travel time and saturation field in each MC realization. The time-step size for the IMPES scheme is set to 20 days, and the total simulation time is 1000 days. MC realizations of the log-permeability $\kappa(\mathbf{x})$ are generated by the sequential Gaussian simulator [29].

4.6.1 Accuracy of the proposed method

We first start by analyzing the ability of the proposed CDF method to accurately approximate first two moments of travel time, mean travel time $\bar{\tau}$ and standard deviation of travel time σ_τ. These two statistics are widely used as uncertainty quantification tools and provide brief confidence intervals of travel time variability. (4.30) enables the computation of any travel time moments from CDF. Numerically, we compute the integrals (4.30) by using 500 quadrature points. Figure 4.4 exhibits these statistical moments along the cross-section $x_1 = x_2$ in two different cases. These profiles $\bar{\tau}(x_1, \cdot)$ and $\sigma^2_\tau(x_1, \cdot)$ are alternatively computed with MCS, the SMEs, and the CDF method.

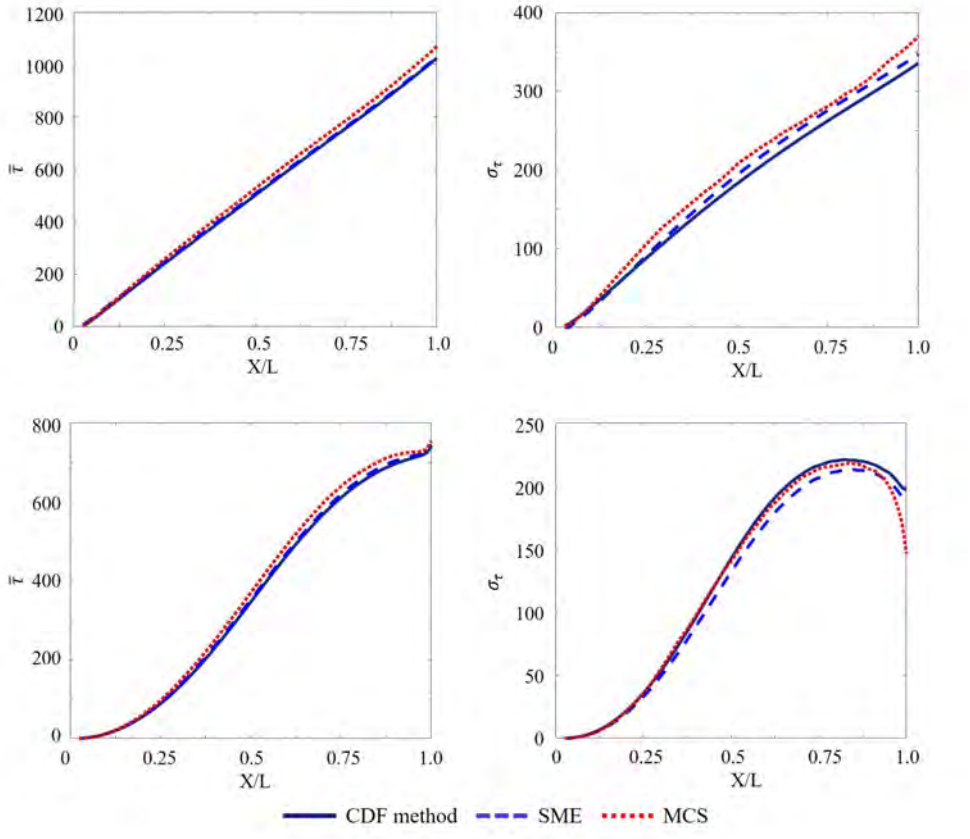

Figure 4.4: Mean (left column) and variance (right column) of hydraulic head, $\bar{h}(x_1, x_2 = 1/2)$ and $\sigma^2_h(x_1, x_2 = 1/2)$, for mean uniform flow (top row) and flow to a well located at the middle of the domain (bottom row). These moments are alternatively computed with Monte Carlo simulations (MCS), the statistical moment equations (SME), and the CDF method. The statistical properties of log-conductivity are $\bar{Y} = 0$; and $\sigma^2_Y = 1$ and $\ell_Y = 0.1$ in both cases.

By construction, the CDF $F_\tau(T;\mathbf{x})$ in (4.27) is to have the same $\bar{\tau}(\mathbf{x})$ and $\sigma_h^2(\mathbf{x})$ as their counterparts computed with the SMEs. The slight discrepancy between the two sets of the statistical moments in Figure 4.4 is due to numerical errors in the solution of the CDF equation and the subsequent evaluation of the quadratures required to compute the first two moments of a CDF. Consistent with the previous SME-focused studies [56, 61, 82], the mean and variance of hydraulic head computed with SMEs are in good agreement with those inferred from MCS, regardless of the flow regime, since our the variance of log-permeability is moderate. The discrepancy between the two approaches is larger for the standard deviation than for the mean.

While useful, these statistics are of limited use in risk assessment, which often requires spatial maps of exceedance/non-exceedance probabilities, $\mathbb{P}\{\tau(\mathbf{x}) > \mathcal{T}\} = 1 - F_\tau(\mathcal{T};\mathbf{x})$, for a selected TOF threshold $\mathcal{T}$. Figure 4.5 exhibits such maps for $\mathcal{T} = 50$ days. Visual inspection of these maps, alternatively computed with the reference MCS and the CDF method, demonstrates the CDF method's accuracy in most of the computational domain, except for the vicinity of its boundaries. In these regions, the total fluid velocity q_{tot} is slow and the MSEs loose their accuracy. The CDFs $F_\tau(\mathcal{T};\mathbf{x})$ presented in Figure 4.6 for several points $\mathbf{x} \in \Omega$ provide another illustration of the accuracy of the CDF method.

The normalized first Wasserstein distance between two distributions,

$$\mathcal{D}(\mathbf{x}) \equiv \frac{\int_0^{\tau_{\max}} \mid F_\tau(\mathcal{T};\mathbf{x}) - F_\tau^{\text{MCS}}(\mathcal{T};\mathbf{x}) \mid \mathrm{d}\mathcal{T}}{\int_0^{\tau_{\max}} F_\tau(\mathcal{T};\mathbf{x})\mathrm{d}\mathcal{T}}, \tag{4.40}$$

provides a more quantitative assessment of the agreement between the CDFs computed with the CDF method (F_h) and the reference MCS (F_h^{MCS}). These integrals are computed with the Gaussian quadrature rule. The resulting contour plots of $\mathcal{D}(\mathbf{x})$ are shown in Figure 4.7. The error metric $\mathcal{D}(\mathbf{x})$ is small throughout much of the domain Ω, but increase around the boundary segments where the streamline density is small. The behavior of $\mathcal{D}(\mathbf{x})$ mirrors that of the TOF variance σ_τ^2 and reflects the error in the perturbation-based estimation of the latter. In both flow scenarios, the average Wasserstein distance $\mathcal{D}_{\text{ave}} = \|\Omega\|^{-1} \int_\Omega \mathcal{D}(\mathbf{x})\mathrm{d}\mathbf{x}$ remains small, 0.032 for the mean uniform flow and 0.046 for the quarter-five well configuration.

Finally, we verify the accuracy of the CDF-FROST framework. Figure 4.8 and 4.9 exhibits temporal snapshots of the risk maps of saturation, i.e., of the exceedance probability $\mathbb{P}[S_{\text{w}}(\mathbf{x},t) > 0.5] = 1 - F_{S_{\text{w}}}(0.5;\mathbf{x},t)$ computed, alternatively, with CDF-FROST and MCS for the quarter-five spot configuration. In both flow regimes, the two methods yield similar risk estimates, with slight disagreement confined to the areas of small flow velocity. This finding suggests that the discrepancy between the reference MCS and the CDF-FROST method is largely due to the approximation error of the SMEs. Figure 4.10 elaborates this point further by presenting the CDF estimates at several

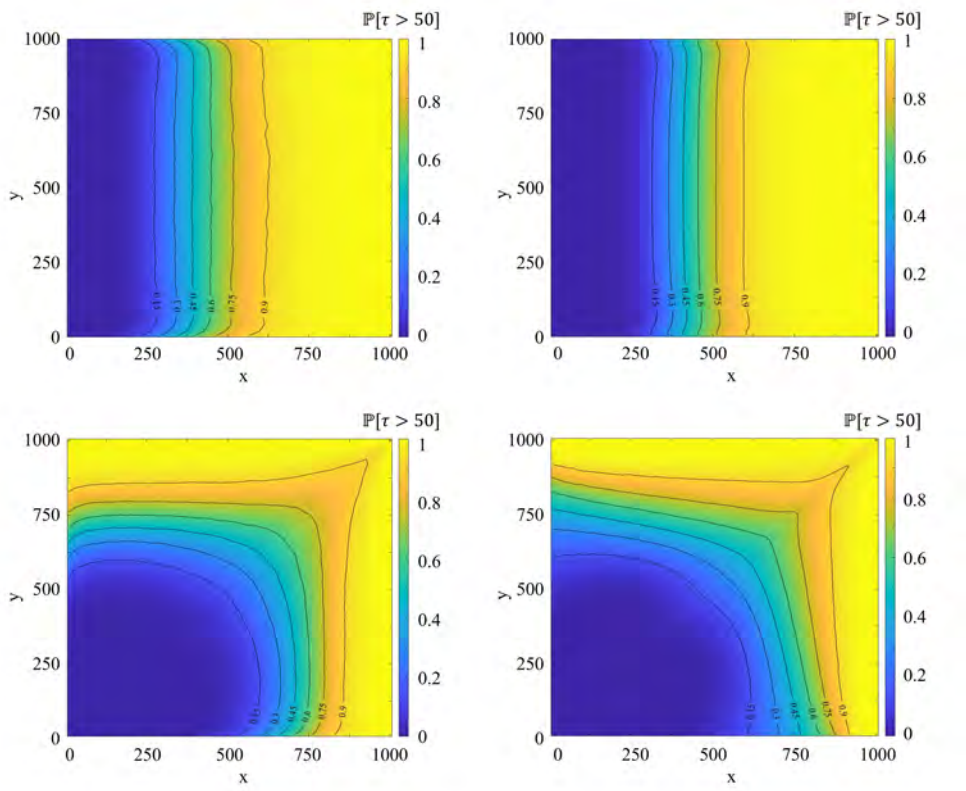

Figure 4.5: Spatial maps of exceedance probability $\mathbb{P}[\tau(\mathbf{x}) > T = 50] = 1 - F_\tau(T = 50; \mathbf{x})$ obtained with MCS (left column) and CDF method (right column) for mean uniform flow (top row) and quarter five spot configuration (bottom row).

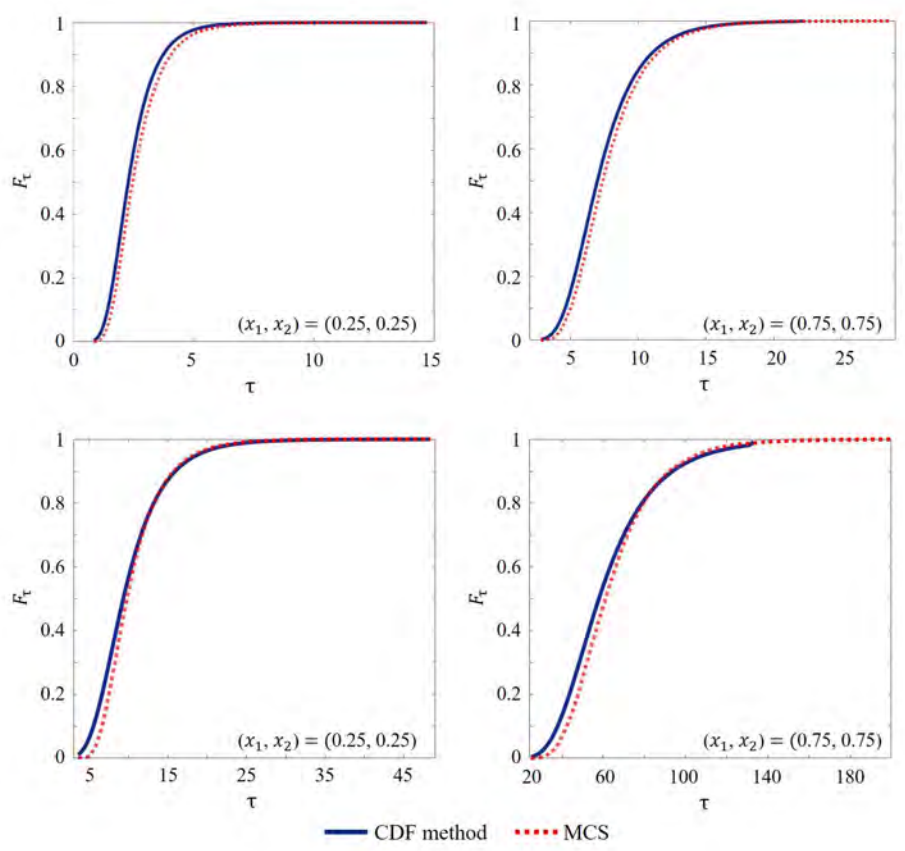

Figure 4.6: Travel time CDFs F_τ computed with MCS and the CDF method at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the flow flow domain for mean uniform flow (top row) and quarter-five spot configuration (bottom row).

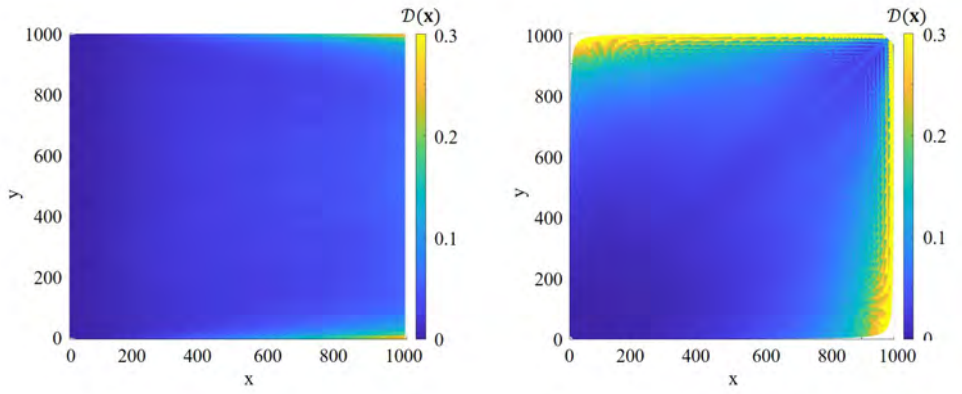

Figure 4.7: Spatial maps of the normalized Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the hydraulic head CDFs computed with the CDF method and Monte Carlo simulations for mean uniform flow (left) and quarter-five spot configuration (right).

points $\mathbf{x} \in \Omega$. The results demonstrate the high accuracy of CDF-FROST. Note that the direct MCS utilizing IMPES scheme exhibits numerical diffusion around the locations of shock fronts, while MCS-FROST and CDF-FROST employing analytical Buckley-Leverett solution does not.

4.6.2 Computational Efficiency of CDF-FROST method

One can expect CDF-FROST to be faster than MCS because the former involves the numerical solution of a fixed number N_{sl} of two-dimensional linear PDEs (4.27), while the latter requires solving a large number of d-dimensional nonlinear coupled PDEs (4.3)–(4.7). CDF-FROST is also expected to be more efficient than MCS-FROST [44, 43], since it obviates the need for any MCS. Table 4.1 confirms these expectations. The CPU times are reported for the same discrepancy level, defined by the average Wasserstein distance $\mathcal{D}_{\text{ave}}$ between the saturation CDFs computed with either MCS, MCS-FROST, or CDF-FROST and the reference MCS. It takes $N_{\text{MCS}} = 2110$ MC realizations to achieve $\mathcal{D}_{\text{ave}} \approx 0.03$ for the mean uniform flow, and $N_{\text{MCS}} = 2630$ MC realizations to achieve $\mathcal{D}_{\text{ave}} \approx 0.05$ for the quarter-five well configuration. To achieve the same accuracy, MCS-FROST requires $N_{\text{MCS}} = 3810$ MC realizations for the mean uniform flow, and $N_{\text{MCS}} = 4160$ MC realizations for the quarter-five well configuration. For the same discrepancy level, CDF-FROST is about five and ten times faster than MCS-FROST and MCS, respectively.

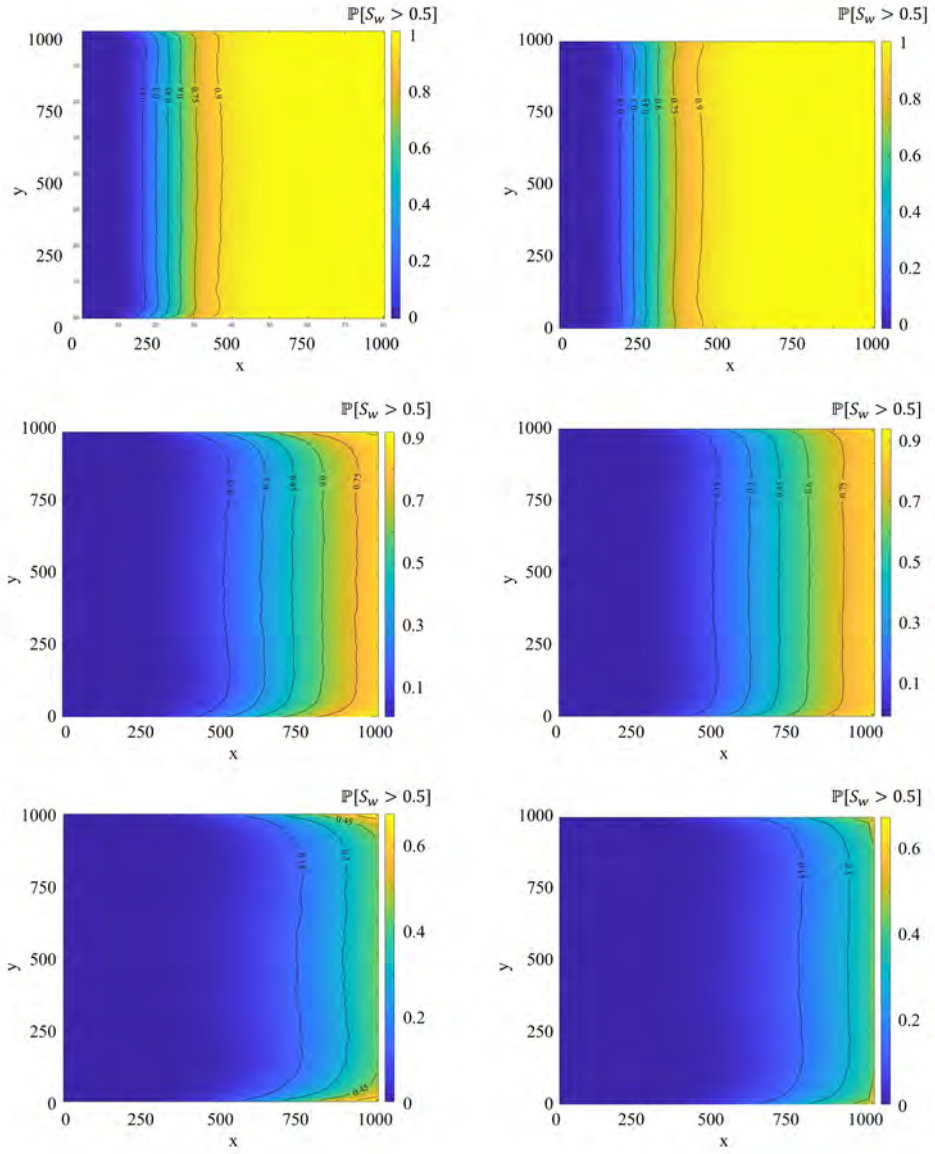

Figure 4.8: Spatial-temporal evolution of exceedance probability $\mathbb{P}[\tau(\mathbf{x}) > T = 50] = 1 - F_\tau(T = 50; \mathbf{x})$ at times $t = 100days$ (top row), $t = 500days$ (center row), $t = 800days$ (top row) obtained with MCS (left column) and CDF method (right column) for mean uniform flow.

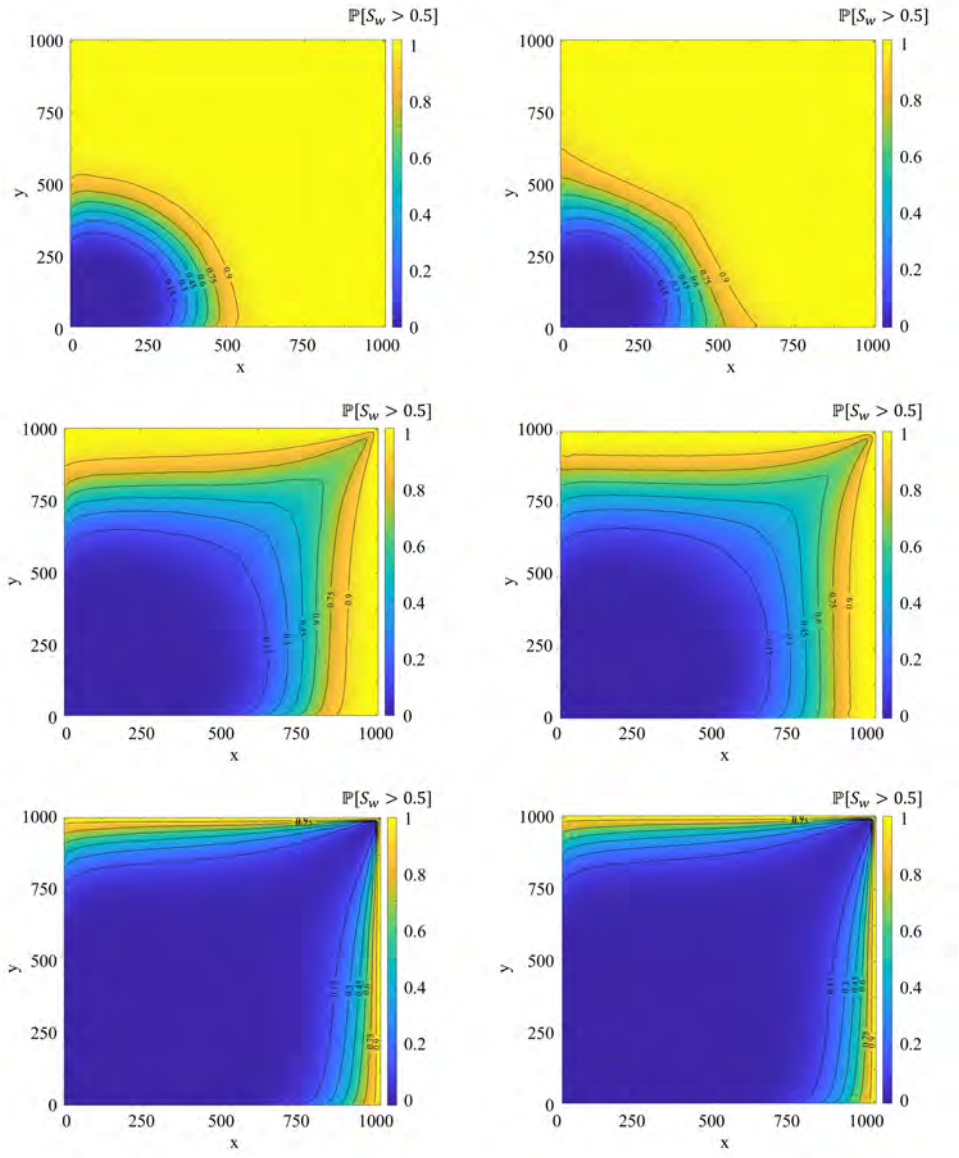

Figure 4.9: Spatial-temporal evolution of exceedance probability $\mathbb{P}[\tau(\mathbf{x}) > T = 50] = 1 - F_\tau(T = 50; \mathbf{x})$ at times $t = 100 days$ (top row), $t = 500 days$ (center row), $t = 800 days$ (top row) obtained with MCS (left column) and CDF method (right column) for quarter-five spot configuration.

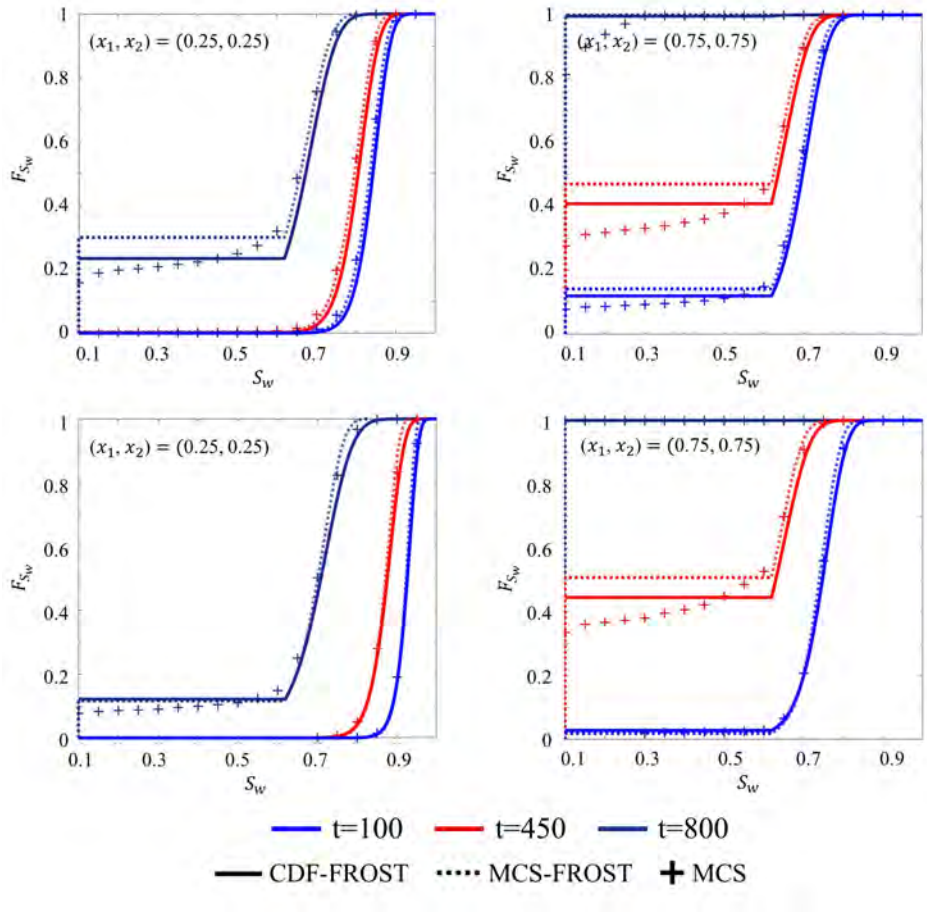

Figure 4.10: Saturation CDFs $F_{S_w}(s;\cdot)$ computed with MCS, MCS-FROST [44, 43], and CDF-FROST for two locations $\mathbf{x} = (x_1, x_2)^\top$ at times $t = 100$ days, $t = 500$ days, and $t = 800$ days. The top and bottom rows correspond to the mean uniform flow and the quarter-five spot configuration, respectively.

Table 4.1: Computational times and accuracy of MCS and the MCS-FROST and CDF-FROST methods.

Flow regime	Method	Error $\mathcal{D}_{\text{ave}}$	CPU time (min)
Mean uniform flow	CDF-FROST	$3.23 \cdot 10^{-2}$	$9.55 \cdot 10^{0}$
	MCS-FROST with $N_{\text{MCS}} = 3810$	$3.23 \cdot 10^{-2}$	$5.01 \cdot 10^{1}$
	MCS with $N_{\text{MCS}} = 2110$	$3.23 \cdot 10^{-2}$	$9.38 \cdot 10^{1}$
	MCS with $N_{\text{MCS}} = 1 \cdot 10^{4}$	0	$4.23 \cdot 10^{2}$
Convergent flow	CDF method	$4.58 \cdot 10^{-2}$	$9.46 \cdot 10^{0}$
	MCS-FROST with $N_{\text{MCS}} = 2630$	$4.58 \cdot 10^{-2}$	$5.34 \cdot 10^{1}$
	MCS with $N_{\text{MCS}} = 4160$	$4.58 \cdot 10^{-2}$	$9.66 \cdot 10^{1}$
	MCS with $N_{\text{MCS}} = 4 \cdot 10^{4}$	0	$4.31 \cdot 10^{2}$

4.7 Conclusion

We proposed a sampling-free CDF-FROST method for probabilistic forecast of two-phase flow in heterogeneous porous media with uncertain permeability. By deploying a fixed-streamline assumption, the original FROST method of [44] outperforms MCS in computing saturation CDFs. However, it has a computational bottleneck, the MC computation of the distribution of travel time and statistics of equivalent injection time. CDF-FROST overcomes this limitation by developing a deterministic equation for CDF of travel time. The equation rests on a moment-preserving closure approximation, whose coefficients are determined by solving moment equations. For computational efficiency, the derived CDF equation is written for a mean streamline grid and solved by a specialized numerical scheme that integrates pseudo-time stepping, flux-limited method, and exponential grid spacing. We performed a series of numerical experiments to demonstrate the accuracy and computational efficiency of CDF-FROST. Our study leads to the following conclusions.

- The method of distributions yields accurate estimates of the CDF of travel time in heterogeneous porous media.
- The saturation CDFs obtained with CDF-FROST are in good agreement with reference MCS.
- For the same accuracy, CDF-FROST method is five and ten times faster than MCS-FROST and MCS, respectively.
- CDF-FROST yields probabilistic information of both travel time and saturation that is necessary for risk assessment and decision-making under uncertainty.

Based on the results of Chapters 2 and 4, we built the distribution-based framework for two-phase flow in heterogeneous porous media. The developed framework utilizes the results of moment equations as the input and computes the distributions of QoIs including pressure, travel time, and saturation. Figure 4.11 summarizes the overall process of proposed distribution-based simulator.

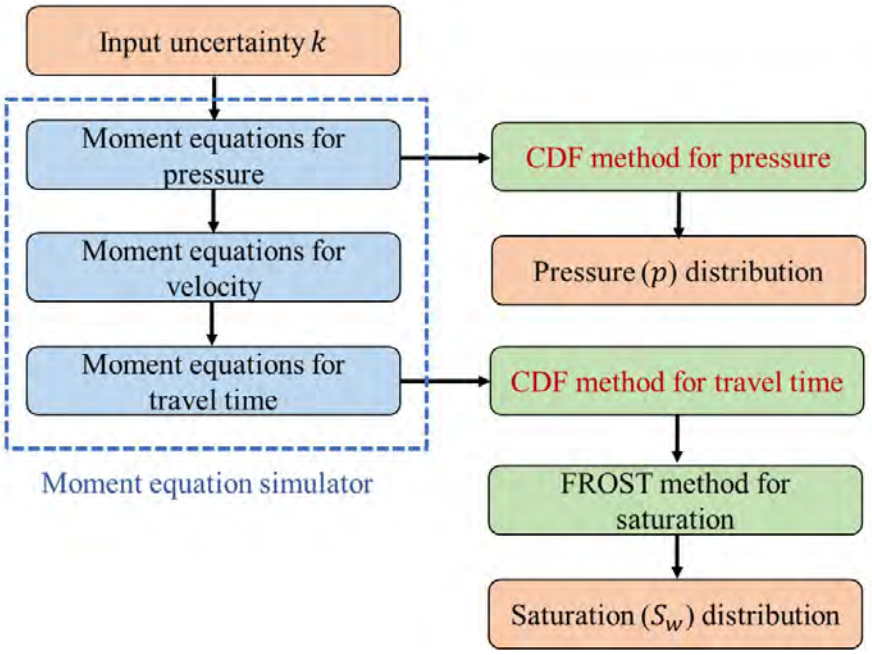

Figure 4.11: Overall flowchart of developed distribution-based simulator

Chapter 5

Consensus equilibrium for subsurface inversion

The concept of this chapter is adopted from a previously published paper:

H. Yang, Y. Lin, B. Wohlberg, & D. M. Tartakovsky, Consensus equilibrium for subsurface delineation. Water Resources Research, 57(10), e2021WR030151.

Most of figures and equations are identical to corresponding sections of the paper.

5.1 Introduction

Reliable characterization of the subsurface is a key component of quantitative predictions of flow and transport in geologic formations. Subsurface characterization generally entails inversion (aka history matching or data assimilation), a computational procedure that converts observations (e.g., of hydraulic head or solute concentration) into multi-dimensional images of model parameters (e.g., hydraulic conductivity or dispersivity). While inverse strategies vary widely, most of them involve minimization of a discrepancy between observations and model predictions.

Gradient-based methods [81, and the references therein] and evolutionary algorithms [19, and the references therein] are some of the most successful approaches to optimization-based inversion. Among these, adjoint gradient-based methods boast high efficiency, because they require only one forward and one backward simulation to compute model sensitivity [81, and the references therein].

These and similar optimization-based procedures generate one optimal solution at a time; they can be combined with the randomized maximum-likelihood method [49] to obtain multiple posterior models for uncertainty quantification. Sampling-based approaches for data assimilation, e.g., (dual) ensemble Kalman filter [106], ensemble smoother [22], Markov chain Monte Carlo [122], and DREAM algorithm [52] combined with Markov chain Monte Carlo, allow one both to estimate model parameters and to assess predictive uncertainty. Even though iterative ensemble-based schemes [22, 107] ameliorate the convergence issue that plagues most inversion algorithms, sampling-based approaches are generally more computationally expensive than adjoint gradient method. Yet, they are often more efficient than gradient-based approaches without the adjoints.

Regardless of the strategy used to achieve it, subsurface inversion is generally an ill-posed problem that has to be regularized. One way to do so is to incorporate prior information in the form of geological constraints, which would guarantee a geologically realistic solution. A majority of the current approaches rely on explicitly definable priors such as two-point statistics (mean and covariance), which renders them inappropriate for complex geology. One alternative is to employ a parameterization procedure aiming to represent geological maps in terms of a small number of parameters. Methods of this class include principal component analysis [80, 104] and deep learning-based techniques [20, 53, 52, 64]. Though these approaches have shown good performance in many subsurface applications, their results vary with the subjectively defined number of parameters and parameterization methods, i.e., require significant fine-tuning.

We posit that high (parametric) dimensionality of subsurface images and non-differentiability of the image reconstruction problem argue for the deployment of proximal algorithms such as alternating direction method of multipliers (ADMM) [33, 1, 12]. More specifically, we adopt plug-and-play (PnP) priors [101], which extend the previous proximal algorithms by relying on recent progress in machine learning in general and deep neural networks (DNN) in particular. Inspired by the mathematical equivalence between the proximal operator and an image denoiser, PnP methodologies provide a flexibility to integrate various heterogeneous priors that may not be explicitly defined [85, 21, 71, 46, 88]. Among them, consensus equilibrium (CE) [16], a generalization of the ADMM-based PnP scheme, integrates multiple advanced prior models (e.g., denoisers, data fidelity agents, deblurring maps, etc.) within an optimization-free framework; examples of its application to image reconstruction problems can be found in [37, 86].

We introduce CE as a means to achieve geologically realistic results for subsurface inversion problems, which are formulated in Section 5.2. The CE approach, presented in Section 5.3, utilizes a denoising prior and deep learning-based prior to maintain geological realism in the inversion of hydraulic head data. The "agents" that seek equilibrium in our setting, i.e., operators responsible for data fidelity, for data denoising, and for geological priors, are introduced in Section 5.4. In

Section 5.5, we demonstrate the performance of the proposed CE framework via a series of numerical experiments dealing with two-dimensional transient flow in an aquifer having channelized spatial arrangement. Main findings and conclusions drawn from our study are summarized in Section 5.6.

5.2 Problem Formulation

The goal of an inverse problem is to recover an unknown subsurface model parameter set $\mathbf{m} \in \mathbb{R}^n$ (e.g., values of hydraulic conductivity K in n elements of a numerical grid) from a set of measurements $\mathbf{d} \in \mathbb{R}^{m \cdot k}$ (e.g., of hydraulic head observed in m wells at k time intervals); in a typical application, $m \ll n$ and $m \cdot k \ll n$. In terms of the maximum-a-posteriori (MAP) estimate, this problem takes the form of an optimization problem,

$$\mathbf{m}^* = \underset{\mathbf{m} \in \mathbb{R}^n}{\operatorname{argmin}} \, f(\mathbf{m}). \tag{5.1}$$

The MAP cost function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^+$ is defined by

$$f(\mathbf{m}) = -\ln p(\mathbf{d}|\mathbf{m}) - \ln p(\mathbf{m}) + \text{const}, \tag{5.2}$$

where $p(\mathbf{d}|\mathbf{m})$ and $p(\mathbf{m})$ represent a data-fidelity function and a prior distribution, respectively. It is common to assume that the distribution of random observation errors is a multivariate normal, in which case the data-fidelity function, $-\ln p(\mathbf{d}|\mathbf{m})$, is expressed as

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{d})^\top \mathbf{C}_D^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}), \tag{5.3}$$

where $\mathbf{C}_D \in \mathbb{R}^{(m \cdot k) \times (m \cdot k)}$ is the covariance matrix of the measurement errors, and $\mathbf{g}(\cdot) : \mathbb{R}^n \to \mathbb{R}^{m \cdot k}$ represents the model predictions of an observable at space-time points $(\mathbf{x}, t)$ at which the observable's measurements $\mathbf{d}$ are available.

The prior distribution $p(\mathbf{m})$ is often selected to be standard, e.g., multivariate Gaussian. However, complex priors representative of realistic geological environments are often poorly described by such explicit distributions, and alternatives are needed. We propose the joint use of multiple advanced priors (e.g., machine learning-based priors) and conventional optimization-based approaches within the CE framework to generate geologically realistic models.

5.3 CE Framework

The novelty of CE framework is to fuse sets of heterogeneous models that may or may not arise from the regularized optimization. The plug-and-play (PnP) reconstruction [101] emerging from ADMM is the first method to incorporate the denoising operators that have no underlying optimization problem. In addition to providing an optimization-free interpretation of PnP, CE extends PnP to handle problems involving more than two "agents", allowing the use of more than one implicit regularizer for same problem.

To formulate the CE equations, we split the MAP cost function $f(\mathbf{m})$ in (5.2) into N auxiliary functions $f_i(\mathbf{m}) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, N)$, such that (5.1) becomes

$$\min \sum_{i=1}^{N} f_i(\mathbf{m}_i); \qquad \text{subject to } \mathbf{m}_i = \mathbf{m}, \quad i = 1, \ldots, N, \tag{5.4}$$

with variable $\mathbf{m}_i \in \mathbb{R}^n$. A proximal mapping operator $F_i : \mathbb{R}^n \to \mathbb{R}^n$, corresponding to the cost function f_i, is defined by

$$F_i(\mathbf{m}) = \underset{\mathbf{v}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \frac{\|\mathbf{v} - \mathbf{m}\|^2}{2\sigma_i^2} + f_i(\mathbf{v}) \right\}. \tag{5.5}$$

The regularization parameters σ_i controls the convergence speed of the PnP algorithm. If f_i is a lower-semicontinuous and convex function on $\mathbb{R}^n$, then a solution of (5.4) and, hence, of (5.1) is given by a solution, $\mathbf{m}^*$, of the CE equations [16],

$$F_i(\mathbf{m}^* + \mathbf{u}_i^*) = \mathbf{m}^*, \quad i = 1, \ldots, N; \quad \text{and } \sum_{i=1}^{N} \mathbf{u}_i^* = \mathbf{0}. \tag{5.6}$$

The main advantage of the CE framework over optimization-based methods is that other non-expansive operators, called agents, can be employed in lieu of the proximal mapping operator F_i [16, 11]. An operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be non-expansive if there exists a real number $0 \leq k \leq 1$ such that

$$\|T(\mathbf{x}) - T(\mathbf{y})\| \leq k\|\mathbf{x} - \mathbf{y}\|, \tag{5.7}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The non-expansiveness, a weaker condition than being a proximal operator, allows us to use much richer class of actions including machine learning models.

The CE equations (5.6) can be solved with several proximal point algorithms, such as ADMM

[33] and Douglas-Rachford [12]. We introduce stacked operators $\mathbf{F}$ and $\mathbf{G}$ such that

$$\mathbf{F}(\mathbf{m}) = \begin{bmatrix} F_1(\mathbf{m}_1) \\ \vdots \\ F_N(\mathbf{m}_N) \end{bmatrix}, \qquad \mathbf{G}(\mathbf{m}) = \begin{bmatrix} \bar{\mathbf{m}} \\ \vdots \\ \bar{\mathbf{m}} \end{bmatrix}, \qquad \bar{\mathbf{m}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i. \tag{5.8}$$

Then (5.6) takes the form

$$\mathbf{F}(\mathbf{v}^*) = \mathbf{G}(\mathbf{v}^*), \qquad \mathbf{v}^* = \mathbf{m}^* + \mathbf{u}^*. \tag{5.9}$$

The averaging function $\mathbf{G}$ has a property $2\mathbf{G} - \mathbf{I} = \mathbf{I}$, where $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix. This property gives rise to a fixed-point problem

$$(\mathbf{2G} - \mathbf{I})(\mathbf{2F} - \mathbf{I})\mathbf{v}^* = \mathbf{v}^*. \tag{5.10}$$

We solve it using the Mann iteration [79], which is equivalent to the Douglas-Rachford algorithm [39],

$$\mathbf{v^{k+1}} = (1-\rho)\mathbf{v^k} + \rho(\mathbf{2G} - \mathbf{I})(\mathbf{2F} - \mathbf{I})\mathbf{v^k}, \tag{5.11}$$

where $\rho \in (0, 1)$ is the fixed parameter. The convergence of the Mann iteration is guaranteed when the operator $\mathbf{T} \equiv (2\mathbf{G} - \mathbf{I})(2\mathbf{F} - \mathbf{I})$ is non-expansive [11]. If $\rho = 0.5$ and $N = 2$, then the fixed point approach (5.11) is identical to the PnP with ADMM algorithm [85].

5.4 Selection of Agents for Subsurface Delineation

Our inverse modeling strategy is to deploy the CE framework (5.9) with $N = 3$ agents F_i: a data fidelity agent F_{dat}, a denoising prior agent F_{den}, and a geology prior agent F_{geo}. The first of these, F_{dat}, is introduced to reduce the mismatch between observations, $\mathbf{d}$, and predictions of the reconstructed model, $\mathbf{g}(\mathbf{m})$. It is defined as a proximal mapping (5.5) of the data fidelity function (5.3),

$$F_{\text{dat}}(\mathbf{m}) = \underset{\mathbf{v}\in\mathbb{R}^n}{\text{argmin}} \left\{ \frac{\|\mathbf{v} - \mathbf{m}\|^2}{2\sigma_{\text{dat}}^2} + \frac{1}{2}(\mathbf{g}(\mathbf{v}) - \mathbf{d})^\top \mathbf{C}_D^{-1}(\mathbf{g}(\mathbf{v}) - \mathbf{d}) \right\}, \tag{5.12}$$

where σ_{dat} is an internal parameter controlling the strength of the regularization term $\|\mathbf{v} - \mathbf{m}\|^2$. We solve the minimization problem in (5.12) using the gradient-based method L-BFGS-B, in which adjoints are used to compute the gradient [13, 81, 70]. The adjoint method is employed here because

of its computational efficiency and lack of the Gaussianity assumption, but other advanced inversion techniques [22, 107, 52, 122] can be plugged into CE as F_{dat}.

Agent F_{den} represents an image denoising prior, which plays the role of an implicit regularizer; its role is to provide reconstructions with sharper, less smeared hydrofacies boundaries. The underlying assumption behind the use of the denoising prior is that the subsurface is composed of distinct hydrofacies with relatively low heterogeneity within each facies [109, 110, 108, 117]. Examples of image denoisers include total variation (TV) [72], BM3D [26], and DnCNN [121]. BM3D typically provides substantially better denoising performance than TV, and the more recent DnCNN typically outperforms BM3D by a much smaller margin. TV is posed as an optimization problem and is easier to implement as a regularizer for more complex problems than denoising via standard methods [5]. Neither BM3D nor DnCNN, however, has an explicit form as an optimization problem, BM3D being based on a complex algorithm involving block matching and coefficient shrinkage in the transform domain, and DnCNN being a convolutional neural network (CNN).

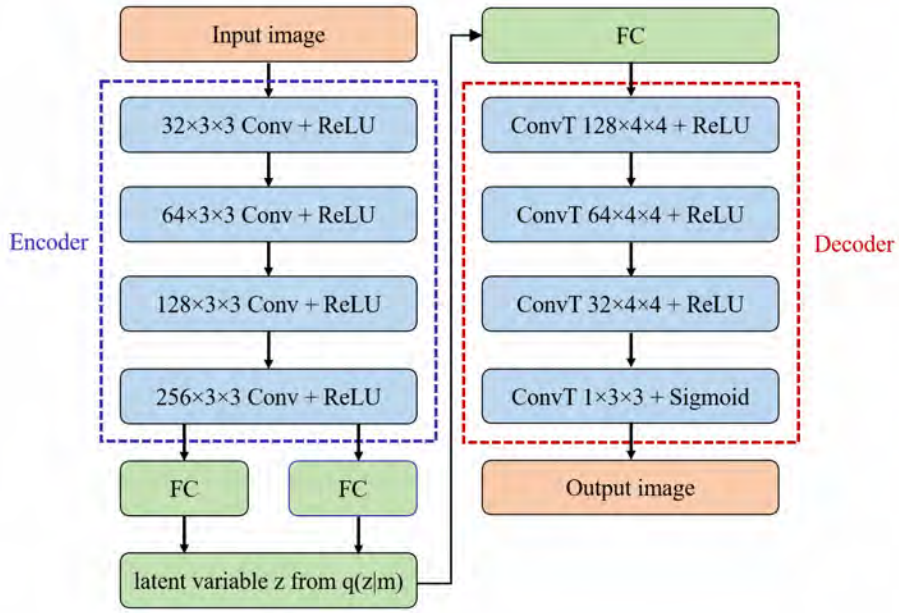

Figure 5.1: Overall architecture of our VAE for incorporating prior geological information. The abbreviations in Figure are defined as; Conv: convolution layer, FC: fully connected layer, and ConvT: transposed convolution layer.

Agent F_{geo} enables our model to preserve the prior geological information such as shapes, sizes, positions and orientations of geological objects. We use variational autoencoders (VAEs), one of the popular generative models, to build such geological information into our model. In a typical implementation, VAEs use deep neural networks to learn latent representations from complex input

data [48]. In so doing, an encoder is used to estimate latent variables $\mathbf{z}$ from input data $\mathbf{m}$, then multiple realizations of $\mathbf{z}$ are generated and used by a decoder to generate reconstructed data $\tilde{\mathbf{m}}$ (Figure 5.1). Each layer of encoder/decoder contains a convolution or transposed-convolution unit to effectively extract and integrate the spatial features of input realizations. The convolution layer performs linear filtering on the output from the previous layer. When the input X is a 2D image, the output feature map, h, is obtained by $N \times N$ filter $\mathbf{w}$. One output pixel, $h_{i,j}$ is calculated as

$$h_{i,j}(Xi,j) = f\left(\sum_{m=1}^{N}\sum_{n=1}^{N} w_{i,j}X_{i+m,j+n}\right), \tag{5.13}$$

where the activation function f is a generally rectified linear unit (ReLU) $f(x) = \max(0, x)$. The transposed convolutional layer reverses the operation of a standard convolutional layer. Details of the convolution and transposed convolution layers can be found in [32, 51]. Training data for VAE can be generated with geostatistical algorithms capable of producing geologically plausible realizations, including object-based [30], process-based methods [73], and multipoint geostatistics [87].

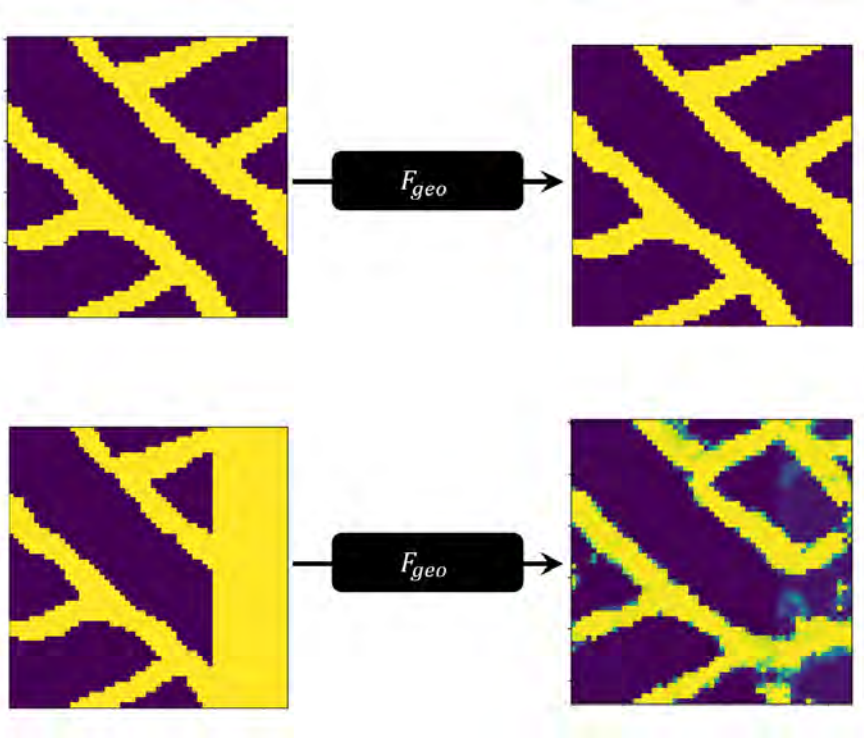

Figure 5.2: Examples of F_{geo} operations applied to geologically realistic input (top row) and unrealistic input (bottom row).

The encoder and decoder are trained simultaneously by minimizing VAE loss L_{VAE}. Let $\boldsymbol{\phi}$ denote a parameterization of the encoder that infers latent variables $\mathbf{z}$ from $\mathbf{m}$; the inferred distribution is

$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{m})$. When the decoder is parameterized with another parameter set $\boldsymbol{\theta}$, yielding a distribution $p_{\boldsymbol{\theta}}(\mathbf{z})$, VAE loss L_{VAE} is formulated as

$$L_{\text{VAE}}(\mathbf{m};\boldsymbol{\phi},\boldsymbol{\theta}) = D_{\text{KL}}[q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{m}), p_{\boldsymbol{\theta}}(\mathbf{z})] - \mathbb{E}_{\mathbf{z}\sim q_{\boldsymbol{\phi}}}[\ln(p_{\boldsymbol{\theta}}(\mathbf{m}|\mathbf{z}))]. \tag{5.14}$$

Here $p_{\boldsymbol{\theta}}(\mathbf{m}|\mathbf{z})$ is the likelihood of $\mathbf{m}$ given $\mathbf{z}$ under the decoder model. The term $D_{\text{KL}}[\cdot]$ is the Kullback-Leibler (KL) divergence, which provides a measure of discrepancy between the distribution of the latent variables, $p_{\boldsymbol{\theta}}(\mathbf{z})$, and the inferred distribution, $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{m})$. The term $\mathbb{E}_{\mathbf{z}\sim q_{\boldsymbol{\phi}}}[\cdot]$, the expectation of the likelihood that the input image $\mathbf{m}$ can be generated from latent variable $\mathbf{z}$, represents the reconstruction error between the actual $\mathbf{x}$ and the reconstructed image $\tilde{\mathbf{m}}$ from the decoder. To compute L_{VAE} efficiently, one represents $q_{\boldsymbol{\phi}}(\cdot)$ as a known readily parametrizable distribution. When $\mathbf{z}$ is a continuous latent variable, $q_{\boldsymbol{\phi}}(\cdot)$ is generally assumed to have a multivariate Gaussian distribution, whose mean μ and variance σ^2 are determined by the encoder model.

If the VAE model is well trained on a sufficient number of realizations, its loss effectively measures the similarity between the input data and the training dataset. To ensure the consistency between an updated image and the prior geological information, we define agent F_{geo} as the proximal map of the VAE loss L_{VAE},

$$F_{\text{geo}}(\mathbf{m}) = \underset{\mathbf{v}\in\mathbb{R}^n}{\text{argmin}} \left\{ \frac{\|\mathbf{v}-\mathbf{m}\|^2}{2\sigma_{\text{geo}}^2} + L_{\text{VAE}}(\mathbf{v};\cdot) \right\}, \tag{5.15}$$

where σ_{geo} is the regularization coefficient. Figure 5.2 shows examples of F_{geo}-agent operations. When the input is geologically realistic, F_{geo} preserves the original input. In contrast, F_{geo} improves the input significantly when the input is inconsistent with the prior geology.

5.5 Numerical Experiments

We consider two-dimensional (vertically averaged) transient flow in an aquifer $\Omega \subset \mathbb{R}^2$ bounded by a surface $\partial\Omega$. Spatiotemporal distribution of hydraulic head $h(\mathbf{x},t)$ is described by the groundwater flow equation,

$$S_s \frac{\partial h}{\partial t} = \nabla\cdot(K\nabla h) - r(\mathbf{x}), \qquad \mathbf{x}\in\Omega, \quad t>0, \tag{5.16}$$

where $K(\mathbf{x})$ and $S_s(\mathbf{x})$ are the aquifer's hydraulic conductivity and specific storage, respectively; and $r(\mathbf{x},t)$ represents sources and sinks (e.g., wells and recharge). This equation is subject to initial

and boundary conditions

$$h(\mathbf{x},0)=H,\quad \mathbf{x}\in\Omega;\qquad h=\phi,\quad \mathbf{x}\in\Gamma_D;\qquad -K\nabla h\cdot\mathbf{n}=\psi,\quad \mathbf{x}\in\Gamma_N. \tag{5.17}$$

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

The groundwater flow model (5.16) and (5.17) is supplemented with (noisy) measurements, $\mathbf{d}$, of hydraulic head h collected at a few locations (e.g., wells) throughout the aquifer during a certain time horizon. In a typical application, an aquifer's properties (K and S_s), auxiliary functions (H, ϕ, and ψ) and sources (r) are all uncertain and have to be inferred from geologic considerations and measurements $\mathbf{d}$. This inverse problem is variously referred to as model calibration or history matching. In the examples reported below, we treat hydraulic conductivity $K(\mathbf{x})$ as the only unknown parameter, and relate its natural logarithm, $Y(\mathbf{x})=\ln K(\mathbf{x})$, to the corresponding values of specific storage $S_s(\mathbf{x})$ via a linear regression,

$$S_s=aY+b. \tag{5.18}$$

Following [55], we assume the regression coefficient a to be positive. With this simplification, and under a suitable discretization of the flow domain Ω into n elements (or nodes), the flow problem (5.16) and (5.17) is uniquely characterized by a set of parameters $\mathbf{m}=\{K_1,\ldots,K_n\}$. A numerical solution of this problem is denoted by $\mathbf{h}=\mathbf{g}(\mathbf{m})$.

As the previous studies in geo-inversion [81, 78, 53, 25, 89], we assume each hydrofacies to be homogeneous, i.e., characterized by a constant value of hydraulic conductivity K and, hence, specific storage S_s. This assumption is introduced to verify our method's ability to reconstruct large-scale geological structures; it is not necessary for our methodology to work. In fact, since the discontinuity of K_i is not desirable for efficient algorithms such as gradient-based methods, we relax the inversion problem formulation by allowing K to be continuous and by constraining its range [80, 104, 64].

We use numerical experimentation to illustrate the performance of our CE framework. The simulation parameters and other settings for these experiments are borrowed from [52, 50]. The numerical simulations are performed in Python using the `FEniCS` software library [2].[1] The first setting (Section 5.5.1) deals with deterministic inversion, in which CE has only two agents, F_{dat} and F_{den}. The second (Section 5.5.2) setting provides probabilistic treatment of a more complex geology and CE has three agents (F_{dat}, F_{den}, and F_{geo}).

[1] The data and codes are available at `https://github.com/DDMS-ERE-Stanford/CE.git`.

The meta-parameters introduced in Sections 5.3 and 5.4 are optimized to achieve the best image reconstruction performance using the exhaustive grid search [54]. Specifically, the optimal parameter for the BM3D denoiser is chosen experimentally by visual inspection of the reconstructed images produced by the CE framework. Tuning of the DnCNN denoiser is more difficult because it does not have an explicit noise parameter [115], the noise regime being implicitly selected during the training stage [121]. For the experiments reported here, we compare performance for all pre-trained models provided by [121], selecting the one (i.e., $\sigma = 50$ model for greyscale image) that gives the best CE reconstruction according to visual inspection. The Mann iteration parameter ρ and the number of iterations are set to 0.5 and 30, respectively for the fast convergence of algorithm. The regularization parameters are set to $\sigma_{\text{dat}} = 20$ and $\sigma_{\text{geo}} = 0.5$ to achieve the best-quality reconstructed image.

5.5.1 Deterministic inversion without geological prior

We consider a channelized aquifer represented by a 45×45 grid consisting of 10 m $\times$ 10 m grid cells (Figure 5.3). Values of the hydraulic conductivity of the channels and the ambient matrix are set to 10^{-2} m/s and 10^{-4} m/s, respectively; the true (unknown) conductivity field is shown in Figure 5.3a. Radial flow is induced by a pumping well (the red circle in Figure 5.3a) operating with a fixed hydraulic head of $h_{\text{well}} = 1$ m; constant head $h = 21$ m is prescribed along the left ($x_1 = 0$) and right ($x_1 = 450$ m) sides of the square aquifer; the remaining two boundaries ($x_2 = 0$ and 450 m) are impermeable; the initial hydraulic head over the entire domain is 21 m. The true hydraulic head field $h(\mathbf{x}, t)$ is computed as a numerical simulation of (5.16) and (3.2) with the true hydraulic conductivity field.

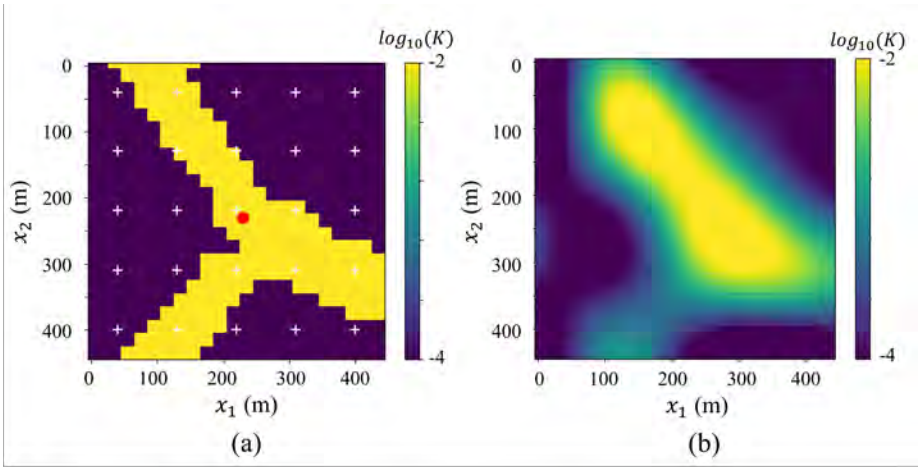

Figure 5.3: True log hydraulic field (left) and its initial guess (right) estimated from conductivity measurements via support vector regression.

Hydraulic head measurements are collected at 25 observation wells (the white circles in Figure 5.3a) every 5 hours for the first 30 hours of the simulation. Gaussian noise with zero mean and variance of 0.2 is added to these values to account for observation errors. The resulting dataset $\mathbf{d}$, as well as conductivity values at these 16 locations, are used in inverse modeling to reconstruct the hydraulic conductivity field. Given both the relative simplicity of the geological structure in Figure 5.3a and the relatively high sampling density, we use this setting to perform deterministic inversion by reconstructing only one geological map. The prior geological constraint is not enforced for this case, thus our CE framework contains only two agents, F_{dat} and F_{den}.

An initial guess for the inversion can be either provided by an expert (e.g., geologist) or constructed from the 16 conductivity measurements. Figure 5.3b presents the initial guess estimated by support vector regression, which showed good performance for facies delineation [112].

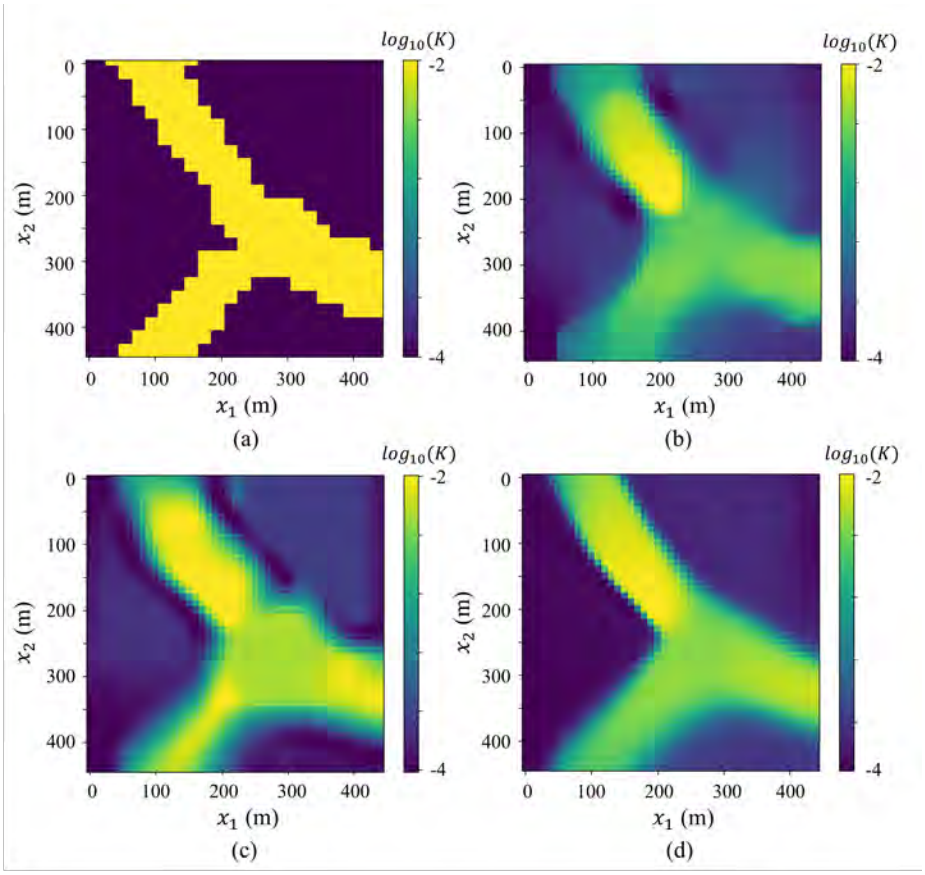

Figure 5.4: True hydraulic conductivity map (a) and its reconstructions via the CE-based inversion with the TV (b), BM3D (c), and DnCNN (d) denoisers.

Figure 5.4 shows the hydraulic conductivity maps obtained via the CE-based inversion with three

alternative types of agent F_{den}: TV, BM3D, and DnCNN. Regardless of the denoiser type, our CE framework captures the channel connectivity and generates realistic images. Visual inspection of these images reveals that CE with the DnCNN image denoiser delineates the facies most accurately.

Predictions of hydraulic head $h(\mathbf{x}, t)$, corresponding to the reconstructed conductivity $K(\mathbf{x})$, at four locations are presented in Figure 5.5. Our CE-based inversion considerably reduces the discrepancy between the true and predicted hydraulic head values, regardless of the denoiser type.

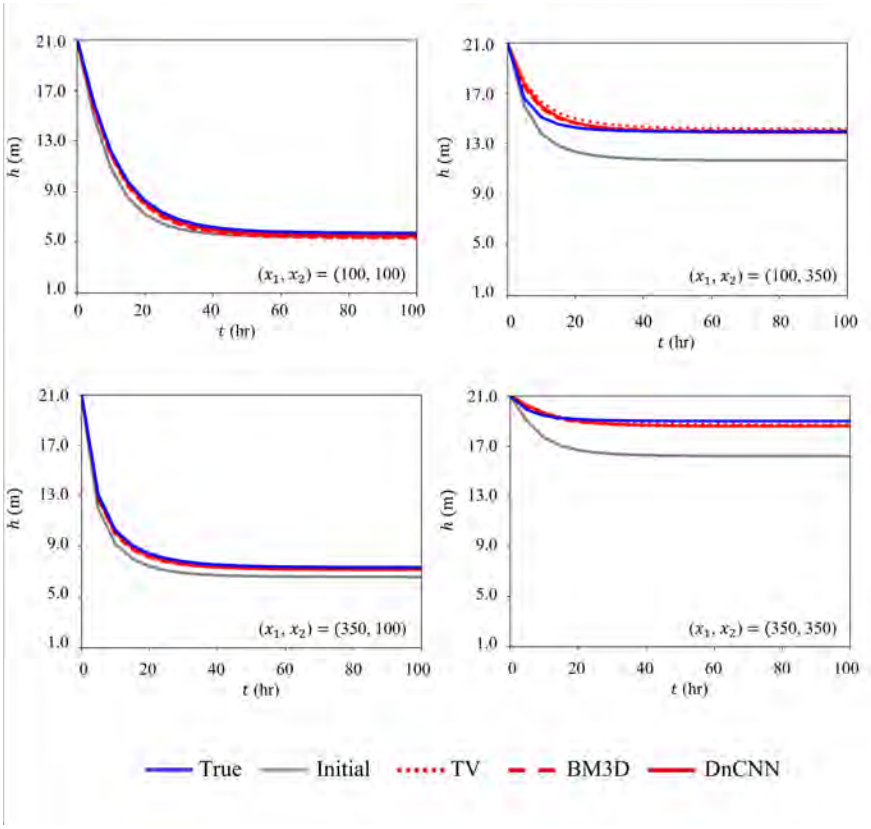

Figure 5.5: Temporal evolution of hydraulic head $h(\mathbf{x}, t)$ at four selected locations $\mathbf{x} = (x_1, x_2)^T$ predicted with the initial guess of $K(\mathbf{x})$ based only on conductivity measurements and with the CE-based inversion with the TV, BM3D, and DnCNN denoisers.

To quantify the relative performance of the CE framework, we introduce a classification error computed as the fraction of misclassified grid points in the total number of grid points. The facies classification is done by setting a threshold value for $\ln K$ as -3, which is the average of the log hydraulic conductivities in the two distinct facies. The best method, CE with the DnCNN denoiser, has the classification error of 5.65%; a remarkably good performance given small number of observation points (0.79% of the total number of grid points). Classification errors of CE with the TV and BM3D denoisers are 6.19% and 6.47%, respectively.

5.5.2 Probabilistic inversion with geologic prior

Prior geological information can be provided by an expert/geologist in the form of a training image (TI) that describes the morphology and key characteristics of hydrofacies. CE incorporates this conceptual geological knowledge through the geology prior agent, F_{geo}, which is used in addition to the other two agents, F_{dat} and F_{den}. We use the DnCNN denoiser, since it performed best on the previous problem. Multi-point geostatistics, specifically the SNESIM algorithm [87], is used to generate 2000 realizations which form the training dataset for the VAE model F_{geo}. These realizations also provide the initial guess for our inversion. The probabilistic inversion is conducted by running multiple inversions with an arbitrarily sampled initial guess; this approach is known as randomized maximum-likelihood or RML [49]. We use 50 RML runs for the stochastic data assimilation.

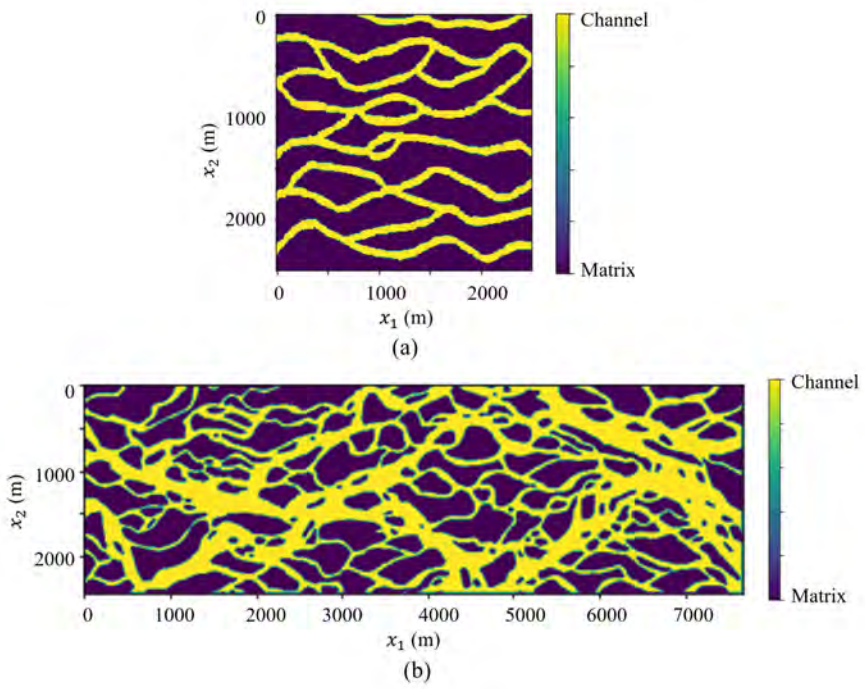

Figure 5.6: Training images used for multi-point geostatistical simulations: (a) a 250×250 hand-made drawing for case 1 and (b) a 768×243 image for case 2 generated from the satellite image [66].

We consider two channelized aquifers with hydraulic conductivities 10^{-2} m/s and 10^{-4} m/s for channels and matrix, respectively. Figure 5.6 shows the TIs for the two different cases[2]. The TI for

[2]The images are available at https://wp.unil.ch/gaia/downloads/

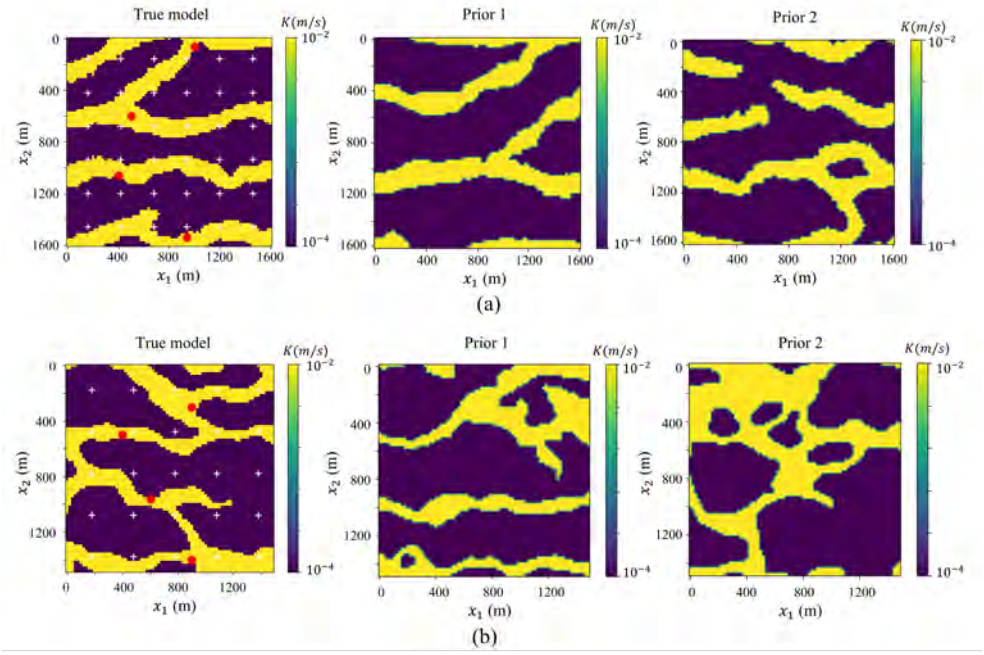

Figure 5.7: True geological maps (left column) and representative prior realizations generated by the SNESIM algorithm (the remaining two columns) for Cases 1 (top row) and 2 (bottom row). The white circles mark locations of observation wells.

Case 1 represents a fluvial channelized aquifer used in [87]. The TI for Case 2 comes from a satellite image of the Ganges delta in Bangladesh [66, 65]. The aquifers are discretized with 81×81 and 75×75 grids for Cases 1 and 2, respectively. In both cases, the size of each grid cell is 20 m $\times$ 20 m with thickness of 1 m. Figure 5.7 shows the true hydraulic conductivity fields and realizations generated from the given TIs.

Initially, groundwater flow is driven by constant heads of 9.0 m and 10.0 m imposed on the left and right boundaries of the domain, respectively; the remaining two boundaries are impermeable. The initial hydraulic head distribution is computed by running the flow simulator until steady state is achieved. At that time, four pumping wells operating with the fixed hydraulic head 8.0 m are installed. Observation wells record the hydraulic head response to groundwater withdrawal. Locations of pumping wells and observation wells are indicated by the red circles and white crosses, respectively. Hydraulic head $h(\mathbf{x}, t)$ at these observation wells is simulated with (5.16) and (3.2) for the true conductivity fields, and recorded every 20 days for the first 80 days. Dataset $\mathbf{d}$ is constructed by corrupting these simulated values with zero-mean Gaussian noise of variance 0.01.

Figures 5.8 and 5.9 exhibit the true and estimated conductivity fields. The latter represent sample averages of the realizations obtained with the initial guess, CE without F_{geo}, and CE with the DnCNN denoiser. Visual inspection of these figures demonstrates a close agreement between the reconstructed geological maps and their true counterparts. On the other hand, the image reconstruction without F_{geo} fails to preserve the geological realism and has a large discrepancy with the true image. The sample-averaged hydraulic conductivity fields for Cases 1 and 2 obtained by CE without F_{geo} (with the threshold conductivity value of 10^{-3} m/s) have the classification errors of 14.6% and 12.8%, respectively. The corresponding classification errors of CE with F_{geo} are 7.6% and 8.1% for Cases 1 and 2, respectively.

Figures 5.10 and 5.11 present the temporal evolution of hydraulic head h at two different locations for Cases 1 and 2, respectively. The RML approach allows us to quantify the prior (before inversion) and posterior uncertainty ranges. The blue areas in these figures represent the 90% confidence intervals; and the dashed lines, $t = 60$ days, indicate the end of the assimilation period. Visually, the recorded drawdowns at two different locations are dissimilar in Cases 1 and 2 due to the different degrees of connectivity to pumping wells. This distinct relationship between hydraulic head response and channel connectivity enables the close agreement between the reconstructed image and true image. The CE-enabled assimilation of hydraulic head data leads to significant uncertainty reduction in the posterior realizations. The hydraulic head profiles of true model fall within the 90% confidence interval in both Cases 1 and 2. These results demonstrate that our methodology has a robust prediction performance.

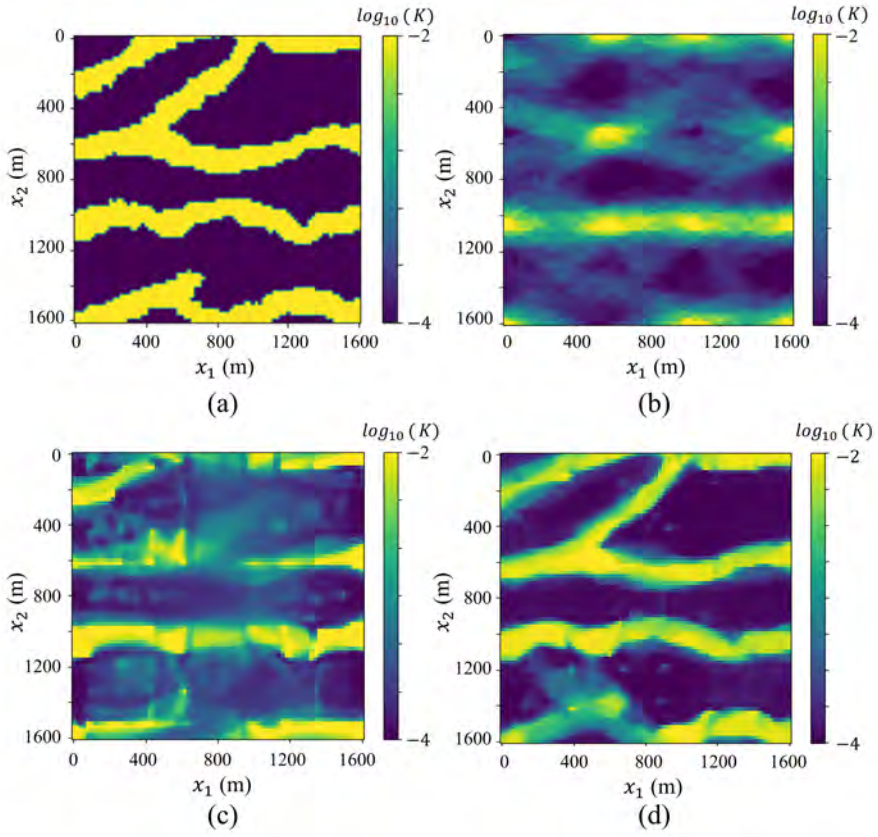

Figure 5.8: Case 1: (a) the true conductivity map and its reconstructions obtained by averaging the realizations of conductivity maps of (b) the initial guess, (c) CE without prior geology agent, and (d) CE with prior geology agent.

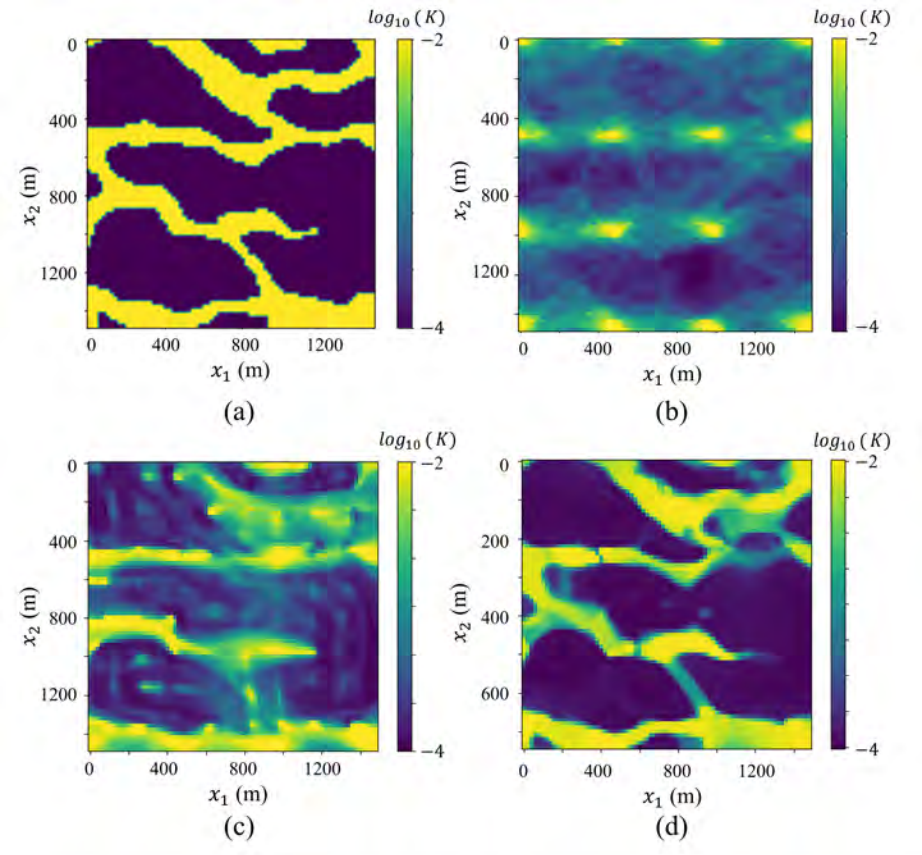

Figure 5.9: Case 2: (a) the true conductivity map and its reconstructions obtained by averaging the realizations of conductivity maps of (b) the initial guess, (c) CE without prior geology agent, and (d) CE with prior geology agent.

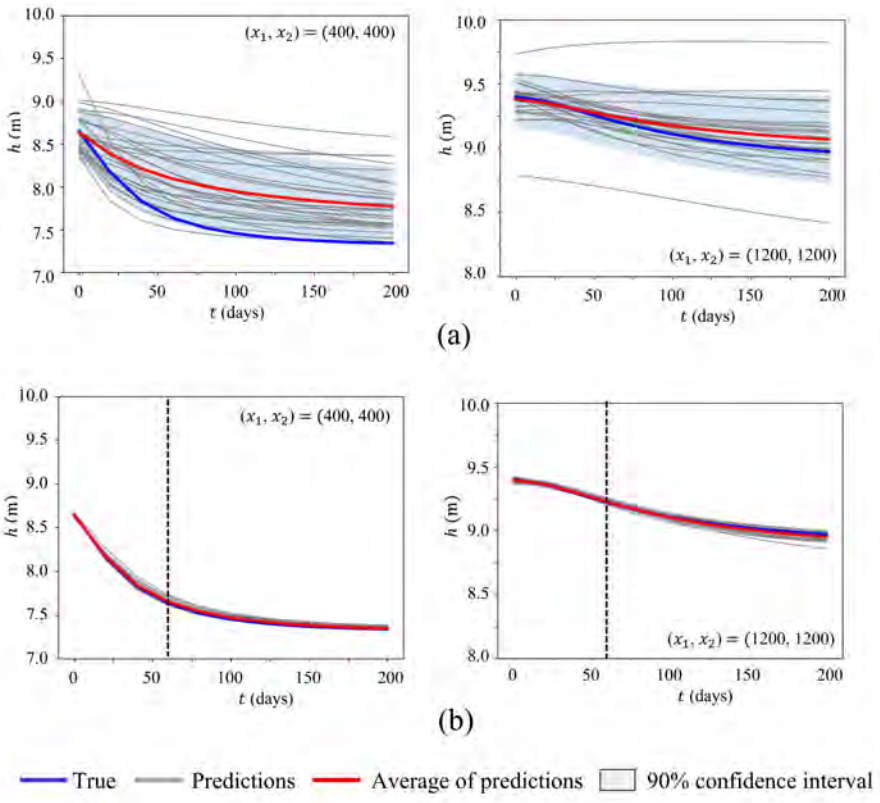

Figure 5.10: Hydraulic head h evolution with time at selected locations $\mathbf{x} = (x_1, x_2)^T$ of aquifers for case 1: (a) initial and (b) posterior.

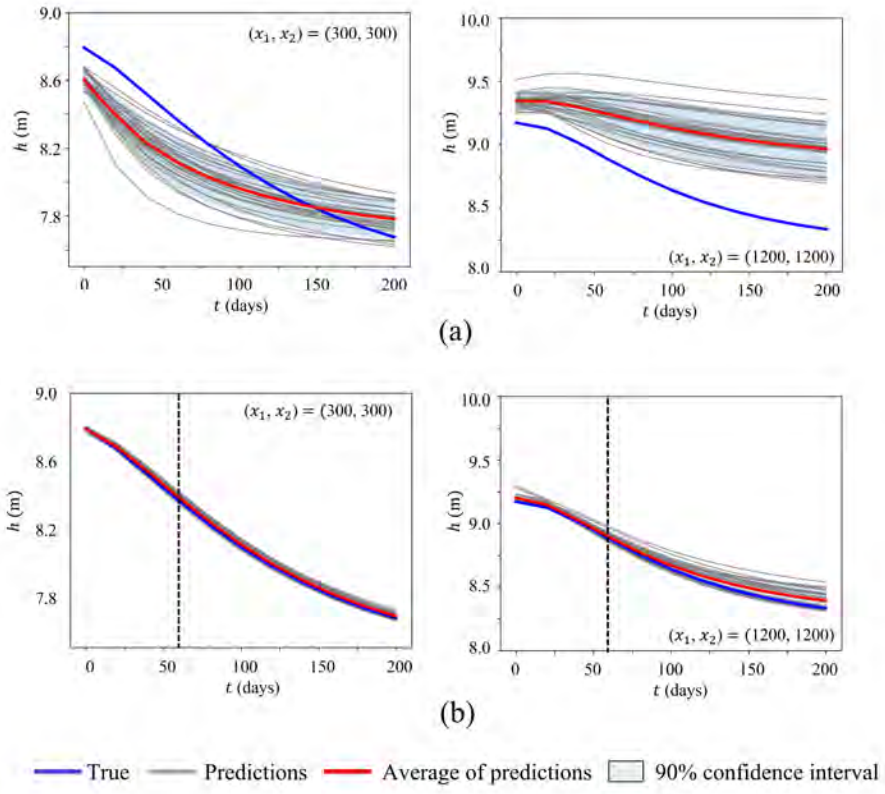

Figure 5.11: Hydraulic head h evolution with time at selected locations $\mathbf{x} = (x_1, x_2)^T$ of aquifers for case 2: (a) initial and (b) posterior.

5.5.3 Computational efficiency of the proposed method

In our CE approach, each of the agents is applied sequentially at each iteration. Since the averaging and updating operations in (5.11) carry negligible computational costs, we approximate the overall computational cost by the sum of the computational times for each agents. Table 5.1 collates the computational burden for the three experiments considered in this study; the computation costs are reported for an Intel Xeon e5-2670 machine running at 2.3 GHz, and all the runs of the probabilistic inversion are parallelized on 50 computer nodes. Agent F_{dat}, an equivalent of the regularized adjoint-based inversion, consumes most of the overall computational cost. This finding implies that our method adds only a small amount of computational burden to the existing optimization-based approaches, while significantly improving the reconstruction performance.

Table 5.1: Computational times of the CE algorithm for the three tests considered.

Tests	Agents	CPU time (min)
Deterministic	F_{dat}	$4.39 \cdot 10^1$
	F_{den} with DnCNN	$5.3 \cdot 10^0$
Probabilistic, Case 1	F_{dat}	$7.22 \cdot 10^1$
	F_{den}	$9.81 \cdot 10^0$
	F_{geo}	$5.78 \cdot 10^0$
Probabilistic, Case 2	F_{dat}	$3.49 \cdot 10^2$
	F_{den}	$2.19 \cdot 10^1$
	F_{geo}	$1.65 \cdot 10^0$

5.6 Summary and Conclusions

We developed and applied a new plug-and-play approach to solve subsurface inversion problems with complex geology. Conventional optimization-based approaches are widely used for this purpose, but their applicability is often limited to geological formations characterized by multi-Gaussian fields. We overcome this limitation within the CE approach by fusing multiple heterogeneous priors with conventional physics-based inversion. Our CE strategy involves three different agents. The data fidelity agent F_{dat} uses an adjoint method to force the consistency between a solution of an inverse problem and observed data. This choice of F_{dat} is due to its computational efficiency and lack of the Gaussianity assumption, but other advanced inversion techniques can be plugged into CE as F_{dat}. The denoiser agent F_{den} minimizes the noise within each hydrofacies and generates realistic geological maps. The geology prior agent F_{geo} incorporates the prior geological knowledge using the VAE model.

We performed three numerical experiments to check the robustness of the proposed method.

First, we solved a deterministic inversion problem on a relatively simple synthetic model. Next, we used our method for probabilistic reconstruction of geologically realistic models. Our numerical experiments lead to the following conclusions.

- The CE framework without agent F_{geo} performs well for the relatively simple geology. However, it cannot reflect the prior geological knowledge and fails to get a geologically realistic maps for more complex formation.
- Among several alternative denoisers, DnCNN (a CNN-based denoiser) shows the best performance as a CE component.
- When the prior geological information, such as shapes or orientations of geo-bodies, is available, the VAE agent trained on the prior realizations significantly improves the CE performance.
- When combined with the RML approach, our method allows one to quantify posterior uncertainties in estimates of both hydraulic parameters and flow response. Our method effectively estimates the posterior uncertainty range.

Chapter 6

Conclusions

In this thesis, we studied the probabilistic framework for the uncertainty quantification of flow and transport in heterogeneous porous media. Specifically, we investigated two different stochastic approaches for efficient characterization of uncertainty - method of distributions for forward uncertainty propagation and consensus equilibrium for realistic inverse modeling.

First, we developed the method of distributions to predict saturated flow in porous media with uncertain hydraulic conductivity or permeability and uncertain boundary functions. The method of distributions yields probability information for QoI which is necessary for probabilistic risk assessment. The proposed method derives and solves the deterministic partial differential equation (PDE) for the cumulative distribution function (CDF) of hydraulic head (or pressure). The moment-preserving closure approximation is deployed to express the coefficients of the CDF equation in terms of mean and variance. The mean and variance can be computed either statistical moment equations (SMEs) or MCS, but the best computational efficiency is achieved when CDF method is combined with SMEs. The series of numerical experiments have shown that 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 method is two-orders of magnitude faster than MCS, and this computational speed up stems from the the smoothness of coefficients in CDF equations.

Secondly, we extend the method of distributions developed in previous chapter to quantify the geologic uncertainty as well as parametric uncertainty of saturated flow in porous media. Using random domain decomposition (RDD), we derive a deterministic equation satisfied by conditional CDF conditioned on a realization of the site geology. The sample average, over alternative geological maps, of the solutions yields unconditional CDFs of pressure/hydraulic head. The proposed method,

which we refer to CDF-RDD, yields accurate estimates of the hydraulic head CDF (exceedance probability) for statistically inhomogeneous porous media in both linear and radial flow regimes. For the same accuracy, the CDF-RDD method is an order of magnitude faster MCS in both radial and linear flow regimes.

Next, We proposed the integrated CDF-FROST framework for complete distribution-based approach for the two-phase flow in heterogeneous porous media. We first develop a CDF method for computing the travel-time distribution. The derivation of this equation relies on a moment-preserving closure approximation whose coefficients are determined by mean and variance of travel-time. We transform the derived CDF equation into mean streamline grid for computational efficiency and solve it using a new numerical scheme designed to solve the derived equation. The specialized numerical approach includes pseudo-time stepping, flux-limited method, and exponential grid spacing. Then CDF method for travel-time is combined with FROST method [44] within CDF-FROST framework to compute the probability distribution without the simulations of MC realizations. The numerical expereiments show that CDFs obtained with CDF-FROST method are in good agreement n both linear and quarter-five spot flow regimes. For the same accuracy, the proposed CDF-FROST method is ten times faster than naive MCS.

Lastly, we developed and applied a new method called consensus equilibrium (CE) to solve subsurface inversion problems with complex geology. The proposed framework is plug-and-play approach designed to characterize non-Gaussian formations which cannot be characterized by conventional optimization-based approaches. The proposed CE approach integrates multiple heterogeneous priors modeled by advanced approaches like machine learning with conventional physics-based inversion. Our CE framework involves three different agents including data fidelity, denoiser, and geology prior agent. The data fidelity agent uses an adjoint method to force the consistency between a solution of an inverse problem and observed data. The denoiser agent minimizes the noise within each hydrofacies and generates realistic geological maps. The geology prior agent incorporates the prior geological knowledge using the variational auto-encoder. The numerical experiments verified that the CE framework with data fidelity and denoiser agents performs well for relatively simple geology. In the presence of prior geological information, such as shapes or orientations of geo-bodies, the VAE agent trained on the prior realizations significantly improves the CE performance. When proposed framework is combined with randomized likelihood approach, the conditioned uncertainties in both geology and dynamic response are effectively quantified.

Appendix A

Moment Equations for Flow Problem

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 [68, 92, 93, 56, 82]. A brief derivation of the MDEs implemented numerically by [61] is presented below for completeness.

The steady-state groundwater flow equation (2.1) is rewritten in terms of log hydraulic conductivity $Y(\mathbf{x}) = \ln K(\mathbf{x})$ as

$$\nabla^2 h + \nabla Y \cdot \nabla h = g(\mathbf{x}) \mathrm{e}^{-Y}. \tag{A.1}$$

Using the Reynolds decomposition $Y(\mathbf{x}) = \bar{Y} + Y'(\mathbf{x})$, recalling that $Y(\mathbf{x})$ is second-order stationary multivariate Gaussian, i.e., that its mean $\bar{Y}$ and variance $\sigma_Y^2 = \langle Y'^2 \rangle$ are constant, defining by $K_{\mathrm{G}} = \exp(\bar{Y})$ 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

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

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 zero. 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,

$$\bar{h} = \bar{h}^{(0)} + \bar{h}^{(1)} + \cdots, \qquad \langle Y' \nabla h' \rangle = \langle Y' \nabla h' \rangle^{(1)} + \langle Y' \nabla h' \rangle^{(2)} + \cdots, \tag{A.3}$$

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 (A.2) yields a recursive set of partial differential equations

$$\nabla^2 \bar{h}^{(0)} = \frac{g}{K_\mathrm{G}}, \qquad \nabla^2 \bar{h}^{(n)} + \nabla \cdot \langle Y' \nabla h \rangle^{(n)} = \frac{g}{2K_\mathrm{G}} \sigma_Y^{2n}, \quad n \geq 1. \tag{A.4}$$

The boundary conditions for these equations are obtained from (2.2) by following a similar procedure,

$$\bar{h}^{(0)} = \bar{\phi}(\mathbf{x}), \quad \bar{h}^{(n)} = 0, \quad n \geq 1, \qquad \mathbf{x} \in \Gamma_D; \tag{A.5a}$$

$$-K_\mathrm{G} \nabla \bar{h}^{(n)} \cdot \mathbf{n}(\mathbf{x}) = \frac{1}{(2n)!} \bar{\psi}(\mathbf{x}) \sigma_Y^{2n}, \quad n \geq 0, \quad \mathbf{x} \in \Gamma_N; \tag{A.5b}$$

$$-K_G \nabla \bar{h}^{(n)} \cdot \mathbf{n}(\mathbf{x}) + \frac{1}{(2n)!} a h^{(n)} \sigma_Y^{2n} = \frac{1}{(2n)!} \bar{\varphi}(\mathbf{x}) \sigma_Y^{2n}, \quad n \geq 0, \quad \mathbf{x} \in \Gamma_R. \tag{A.5c}$$

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 (A.4) 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 (A.1) with $Y'(\boldsymbol{\chi})$, taking the ensemble mean, and retaining the terms of order σ_Y^2,

$$\nabla_\mathbf{x}^2 C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) + \nabla_\mathbf{x} C_Y(\mathbf{x}, \boldsymbol{\chi}) \cdot \nabla_\mathbf{x} \bar{h}^{(0)} = -\frac{g}{K_\mathrm{G}} C_Y(\mathbf{x}, \boldsymbol{\chi}), \tag{A.6}$$

where $C_Y(\mathbf{x}, \boldsymbol{\chi}) = \langle Y'(\mathbf{x}) Y'(\boldsymbol{\chi}) \rangle$ is the auto-correlation of $Y(\mathbf{x})$. Accounting for the lack of correlation between Y and the boundary functions ϕ and ψ, it follows from (2.2) that the moment equation (A.6) is subject to boundary conditions

$$C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) = 0, \quad \mathbf{x} \in \Gamma_D; \qquad \nabla_\mathbf{x} C_{Yh}(\boldsymbol{\chi}, \mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \frac{\bar{\psi}}{K_\mathrm{G}} C_Y(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_N. \tag{A.7}$$

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, i.e., approximate the mean head $\bar{h}$ with $\tilde{h} = \bar{h}^{(0)} + \bar{h}^{(1)}$. Multiplying (A.4) with $\bar{K}$ and summing the equations for $\bar{h}^{(0)}$ and $\bar{h}^{(1)}$ yields the equation (2.11) for the first order

approximation of mean head subject to boundary conditions

$$\tilde{h} = \bar{\phi}(\mathbf{x}), \quad \mathbf{x} \in \Gamma_D; \qquad -K_{\mathrm{G}} \nabla \tilde{h} \cdot \mathbf{n}(\mathbf{x}) = \bar{\psi}(\mathbf{x}) \left(1 + \frac{\sigma_Y^2}{2}\right), \quad \mathbf{x} \in \Gamma_N;$$
$$-K_G \nabla \tilde{h} \cdot \mathbf{n}(\mathbf{x}) + a\tilde{h}(\mathbf{x}) \left(1 + \frac{\sigma_Y^2}{2}\right) = \bar{\varphi}(\mathbf{x}) \left(1 + \frac{\sigma_Y^2}{2}\right), \quad \mathbf{x} \in \Gamma_R. \tag{A.8}$$

An equation for the first-order approximation of the head variance, $\tilde{\sigma}_h^2$, is derived by subtracting (2.11) from (A.4), multiplying the resulting equation with $h'(\mathbf{x})$, and taking the ensemble mean,

$$\nabla_{\mathbf{x}}^2 \sigma_h^2(\mathbf{x}) - 2\langle \nabla h \cdot \nabla h' \rangle^{(1)} + 2\nabla_{\mathbf{x}} \bar{h}^{(0)} \cdot \langle \nabla Y' h \rangle^{(1)} = -\frac{2g(\mathbf{x})}{K_{\mathrm{G}}} C_{Yh}(\mathbf{x}, \mathbf{x}). \tag{A.9}$$

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 auto-covariance 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 (A.1) with $h'(\boldsymbol{\chi})$, taking the ensemble mean, and retaining the terms of order σ_Y^2,

$$\nabla_{\mathbf{x}}^2 C_h(\mathbf{x}, \boldsymbol{\chi}) + \nabla_{\mathbf{x}} C_{Yh}(\mathbf{x}, \boldsymbol{\chi}) \cdot \nabla_{\mathbf{x}} \bar{h}^{(0)} = -\frac{g(\mathbf{x})}{K_{\mathrm{G}}} C_{Yh}(\mathbf{x}, \boldsymbol{\chi}). \tag{A.10}$$

The boundary conditions for this equations are obtained by multiplying (2.2) with $h'(\boldsymbol{\chi})$ taking the ensemble average, and retaining the terms of order σ_Y^2,

$$C_h(\mathbf{x}, \boldsymbol{\chi}) = C_{\phi h}(\mathbf{x}, \boldsymbol{\chi}), \quad \mathbf{x} \in \Gamma_D; \qquad \nabla_{\mathbf{x}} C_h(\mathbf{x}, \boldsymbol{\chi}) \cdot \mathbf{n}(\mathbf{x}) = \frac{C_{\psi h}(\mathbf{x}, \boldsymbol{\chi}) - \bar{\psi}(\mathbf{x}) C_{Yh}(\mathbf{x}, \boldsymbol{\chi})}{K_{\mathrm{G}}}, \quad \mathbf{x} \in \Gamma_N,$$
$$-\nabla_{\mathbf{x}} C_h(\mathbf{x}, \boldsymbol{\chi}) \cdot \mathbf{n}(\mathbf{x}) + \frac{a}{K_G} C_h(\mathbf{x}, \boldsymbol{\chi}) = \frac{-C_{\varphi h}(\mathbf{x}, \boldsymbol{\chi}) + \left(\bar{\varphi}(\mathbf{x}) - a\tilde{h}(\mathbf{x})\right) C_{Yh}(\mathbf{x}, \boldsymbol{\chi})}{K_{\mathrm{G}}}, \quad \mathbf{x} \in \Gamma_R. \tag{A.11}$$

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 (2.2) 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, i.e., $C_h(\mathbf{x}, \boldsymbol{\chi})$ is computed, we evaluate $\langle \nabla h \cdot \nabla h' \rangle^{(1)} = \lim_{\boldsymbol{\chi} \to \mathbf{x}} [\nabla_{\mathbf{x}} \cdot \nabla_{\chi} C_h(\boldsymbol{\chi}, \mathbf{x})]$. Multiplying (A.9) with $\bar{K}$ and evaluating $\langle \nabla h \cdot \nabla h' \rangle^{(1)}$ and $\langle \nabla h \cdot \nabla h' \rangle^{(1)}$ lead to the closed equations (2.12) for the first-order approximation of the head variance

subject to boundary conditions

$$\tilde{\sigma}_h^2(\mathbf{x}) = C_{\phi h}(\mathbf{x},\mathbf{x}), \quad \mathbf{x}\in\Gamma_D; \qquad \nabla_{\mathbf{x}}\tilde{\sigma}_h^2(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) = \frac{2C_{\psi h}(\mathbf{x},\mathbf{x}) - 2\bar{\psi}(\mathbf{x})C_{Yh}(\mathbf{x},\mathbf{x})}{K_{\mathrm{G}}}, \quad \mathbf{x}\in\Gamma_N;$$
$$-\nabla_{\mathbf{x}}\tilde{\sigma}_h^2(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) + \frac{a}{K_G}\tilde{\sigma}_h^2(\mathbf{x}) = \frac{-2C_{\varphi h}(\mathbf{x},\boldsymbol{\chi}) + 2\left(\bar{\varphi}(\mathbf{x}) - a\tilde{h}(\mathbf{x})\right)C_{Yh}(\mathbf{x},\boldsymbol{\chi})}{K_{\mathrm{G}}} \quad \mathbf{x}\in\Gamma_R. \tag{A.12}$$

Alternatively, $\tilde{\sigma}_h^2$ can be obtained by taking the limit of the head's auto-covariance function $C_h(\boldsymbol{\chi},\mathbf{x})$, i.e., $\tilde{\sigma}_h^2 = \lim_{\boldsymbol{\chi}\to\mathbf{x}} C_h(\boldsymbol{\chi},\mathbf{x})$. The limit can be computed from the numerical solution of the (A.10), the C_h between the grid point x and all grid points in the domain.

Next, we derive the first and second moment equations for the stochastic velocity field. For flow in porous media, Darcy's law with unit viscosity can be written as

$$\mathbf{v}(\mathbf{x}) = -\frac{e^{Y(\mathbf{x})}}{\phi}\nabla h(\mathbf{x}) \tag{A.13}$$

As in the derivation of (A.4), Reynolds decomposition and taking the ensemble mean lead to the following equation,

$$\bar{\mathbf{v}}(\mathbf{x}) = -\frac{K_G}{\phi}\sum_{n=0}^{\infty}\frac{1}{2n!}(\nabla\bar{h}^{(0)}\sigma_Y^{2n} + \langle Y'^n\nabla h'\rangle). \tag{A.14}$$

The first order approximation of mean velocity, $\tilde{\mathbf{v}}$ can be written as

$$\tilde{\mathbf{v}}(\mathbf{x}) = -\frac{K_G}{\phi}[\nabla\bar{h}^{(0)}(1+\frac{\sigma_Y^2}{2}) + \nabla C_{Yh}(\mathbf{x},\mathbf{x}))]. \tag{A.15}$$

An equation for the covariance between the i-th and j-th components of the velocity vector, $C_{v_iv_j}(\mathbf{x},\chi)$, is obtained by subtracting (A.14) with (A.15), multiplying $v_j'(\mathbf{x})$, and taking ensemble mean,

$$C_{v_iv_j}(\mathbf{x},\chi) = \frac{K_G(\mathbf{x})K_G(\chi)}{\phi^2}[\frac{\partial\bar{h}(\mathbf{x})}{\partial x_i}\frac{\partial\bar{h}(\chi)}{\partial\chi_j}C_{YY}(\mathbf{x},\chi) + \frac{\partial^2 C_{hh}(\mathbf{x},\chi)}{\partial x_i\partial\chi_j} + \frac{\partial\bar{h}(\mathbf{x})}{\partial x_i}\frac{\partial C_{Yh}(\mathbf{x},\chi)}{\partial\chi_j} + \frac{\partial\bar{h}(\chi)}{\partial\chi_j}\frac{\partial C_{Yh}(\chi,\mathbf{x})}{\partial x_i}]. \tag{A.16}$$

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 (2.2) 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 (2.2) along Γ_R by $\partial\Pi/\partial H$ to obtain

$$-K(\mathbf{x})\nabla\Pi\cdot\mathbf{n}(\mathbf{x}) - aH\frac{\partial\Pi}{\partial H} = -\varphi\frac{\partial\Pi}{\partial H}. \tag{B.1}$$

Ensemble averaging of (B.1) yields

$$-\bar{K}(\mathbf{x})\nabla F\cdot\mathbf{n}(\mathbf{x}) = aH\frac{\partial F}{\partial H} - \bar{\varphi}\frac{\partial F}{\partial H} + \langle K'\frac{\partial\Pi'}{\partial H}\rangle - \langle\varphi'\frac{\partial\Pi'}{\partial H}\rangle \tag{B.2}$$

which requires closure. Consistently with the IEM closure developed for (2.1), we impose

$$-\bar{K}(\mathbf{x})\nabla F\cdot\mathbf{n}(\mathbf{x}) = aH\frac{\partial F}{\partial H} - \bar{\varphi}\frac{\partial F}{\partial H} + \big(\gamma(\mathbf{x})(H-\bar{h}(\mathbf{x})) + \eta(\mathbf{x})\big)\frac{\partial F}{\partial H}, \quad \mathbf{x}\in\Gamma_R, \tag{B.3}$$

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

$$\gamma(\mathbf{x}) = \frac{\frac{\bar{K}}{2}\nabla\sigma_h^2\cdot\mathbf{n}(\mathbf{x}) - a\sigma_h^2(\mathbf{x}) + 2a\bar{h}^2(\mathbf{x})}{\big(\sigma_h^2(\mathbf{x}) - 2\bar{h}^2(\mathbf{x})\big)}, \quad \eta(\mathbf{x}) = \bar{K}\nabla\bar{h}\cdot\mathbf{n}(\mathbf{x}) - a\bar{h}(\mathbf{x}) + \bar{\varphi}(\mathbf{x}). \tag{B.4}$$

Appendix C

Moment Equations for Transport Problem

The first two moments of the velocity field obtained by (A.15) and (A.16) can be used to generate mean streamline. Then, the travel time moments are computed using the perturbation expansion around the mean streamline. The random travel-time τ can be expressed along the mean streamline coordinate as

$$\tau = \int_0^r \frac{1}{v_r(\chi, \eta(\chi))} d\chi \tag{C.1}$$

where r is the distance along the mean streamline, v_r is the velocity component whose direction is along the mean streamline, and χ is a dummy variable to parameterize mean streamline. η is transverse displacement, i.e., the displacement perpendicular to the mean streamline, is defined as

$$\eta(r) = \int_0^r \frac{v_\eta(\chi)}{v_r(\chi)} d\chi. \tag{C.2}$$

Here, v_η is the velocity component perpendicular to the mean streamline. Considering the definition of mean streamline, the mean traverse displacement $\langle \eta \rangle = 0$.

The underlying assumption for perturbation expansion is that the variance of traverse displacement $\eta(\chi)$ and velocity along the streamline v_r is relatively small. Taylor expansion in terms of these two terms is employed to perturb all streamline-related random quantities around the mean streamline. Applying Taylor expansion to (C.1) and neglecting the terms higher than first order, we

can derive the following truncated equation for travel-time τ,

$$\tau(r) = \int_0^r \left[\frac{1}{\langle v_r(\chi, \langle\eta\rangle)\rangle} - \frac{v_r'(\chi, \langle\eta\rangle)}{\langle v_r(\chi, \langle\eta\rangle)\rangle^2} + \frac{\eta'(\chi)}{\langle v_r(\chi, \langle\eta\rangle)\rangle^2} \frac{\partial \langle v_r(\chi, \eta)\rangle}{\partial \eta(\chi)}\bigg|_{\eta=\langle\eta\rangle} \right] d\chi \tag{C.3}$$

Taking ensemble average to the (C.3) yields the equation of the zeroth-order mean travel-time $\bar{\tau}(r)^{(0)}$,

$$\bar{\tau}(r)^{(0)} = \int_0^r \frac{1}{\langle v_r(\chi, \langle\eta\rangle)\rangle} d\chi. \tag{C.4}$$

We subtract the mean equation (C.4) from (C.3) to obtain the first-order fluctuation term,

$$\tau'(r) = \int_0^r \frac{1}{\langle v_r(\chi, \langle\eta\rangle)\rangle^2} \left[v_r'(\chi, \langle\eta\rangle) + \eta'(\chi) \frac{\partial \langle v_r(\chi, \eta)\rangle}{\partial \eta(\chi)}\bigg|_{\eta=\langle\eta\rangle} \right] d\chi \tag{C.5}$$

Multiplying two travel-time fluctuations at different locations and taking ensemble average lead to the following expression for travel-time covariance

$$\begin{aligned} C_\tau(r_1, r_2) = \int_0^{r_1} \int_0^{r_2} \frac{1}{\langle v_{r1}\rangle^2 \langle v_{r2}\rangle^2} \Bigg[& \langle v_{r1}' v_{r2}' \rangle + \frac{\partial \langle v_{r1}\rangle}{\partial \eta}\bigg|_{\eta_1=\langle\eta_1\rangle} \langle v_{r2}' \eta_1' \rangle \\ & + \frac{\partial \langle v_{r2}\rangle}{\partial \eta}\bigg|_{\eta_2=\langle\eta_2\rangle} \langle v_{r1}' \eta_2' \rangle \\ & + \frac{\partial \langle v_{r1}\rangle}{\partial \eta}\bigg|_{\eta_1=\langle\eta_1\rangle} \frac{\partial \langle v_{r2}\rangle}{\partial \eta}\bigg|_{\eta_2=\langle\eta_2\rangle} \langle \eta_1' \eta_2' \rangle \Bigg] d\chi_1 d\chi_2 \end{aligned} \tag{C.6}$$

where $\eta_1 = \eta(\chi_1)$, $\eta_2 = \eta(\chi_2)$, $v_{r1} = v_r(\chi_1, \langle\eta_1\rangle)$, and $v_{r2} = v_r(\chi_2, \langle\eta_2\rangle)$. The closure terms related to transverse displacement, $\langle v_r' \eta' \rangle$ and $\langle \eta_1' \eta_2' \rangle$ can be obtained from (C.2). By expanding v_η and $1/v_r$ with a Taylor series expansion around $\langle v_r \rangle$, we can write the transverse displacement fluctuation as

$$\eta'(r) = \int_0^r \frac{v_\eta'(\chi)}{\langle v_r(\chi)\rangle} \left[1 - \frac{v_r'(\chi)}{\langle v_r(\chi)\rangle} + \frac{{v_r'}^2(\chi)}{\langle v_r(\chi)\rangle^2} - \cdots \right] d\chi. \tag{C.7}$$

With this fluctuation term, we can express the transverse displacement covariance as the expected value of two traverse displacements with first order accuracy,

$$\langle \eta_1'(r_1) \eta_2'(r_2) \rangle = \int_0^{r_2} \int_0^{r_1} \frac{\langle v_{\eta,1}'(\chi_1) v_{\eta,2}'(\chi_2)\rangle}{\langle v_{r,1}\rangle(\chi_1) \langle v_{r,2}\rangle(\chi_2)} d\chi_1 d\chi_2. \tag{C.8}$$

Similarly, the following first-order equation for covariance between transverse displacement and velocity along a streamline can be derived by multiplying (C.7) with v_r' and taking ensemble average:

$$\langle v_r'(r_1, \langle\eta_1\rangle) \eta'(r_2) \rangle = \int_0^{r_2} \frac{\langle v_\eta'(r_1, \langle\eta(r_1)\rangle) v_\eta'(\chi)\rangle}{\langle v_r(\chi)\rangle} d\chi. \tag{C.9}$$

In summary, the travel time moments can be computed by the series of integration and velocity moments derived in Appendix A. The numerical integration generally requires less computational cost than linear solver. Hence, the travel-time moments can be obtained rapidly from the results of Appendix A.

Bibliography

[1] Manya V Afonso, José M Bioucas-Dias, and Mário A T Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010.

[2] Martin Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E Rognes, and Garth N Wells. The FEniCS project version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[3] D. A. Barajas-Solano and D. M. Tartakovsky. Stochastic collocation methods for nonlinear parabolic equations with random coefficients. *SIAM/ASA J. Uncert. Quant.*, 4(1):475–494, 2016.

[4] D. A. Barajas-Solano and D. M. Tartakovsky. Stochastic collocation methods for nonlinear parabolic equations with random coefficients. *SIAM/ASA J. Uncert. Quant.*, 4(1):475–494, 2016.

[5] David A. Barajas-Solano, Brendt Wohlberg, Velimir Vesselinov, and Daniel M. Tartakovsky. Linear functional minimization for inverse modeling. *Water Resources Research*, 51(6):4516–4531, 2015.

[6] Jacob Bear. *Dynamics of fluids in porous media*. Courier Corporation, 1988.

[7] F. Boso, S. V. Broyda, and D. M. Tartakovsky. Cumulative distribution function solutions of advection-reaction equations with uncertain parameters. *Proc. R. Soc. A*, 470(2166):20140189, 2014.

[8] F. Boso, A. Marzadri, and D. M. Tartakovsky. Probabilistic forecasting of nitrogen dynamics in hyporheic zone. *Water Resour. Res.*, 54(7):4417–4431, 2018.

[9] F. Boso and D. M. Tartakovsky. The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resour. Res.*, 52(6):4700–4712, 2016.

[10] F. Boso and D. M. Tartakovsky. Data-informed method of distributions for hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 42(1):A559–A583, 2020.

[11] Charles A Bouman. Model based image processing. *Purdue University*, 2013.

[12] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[13] D. R. Brouwer and J. D. Jansen. Dynamic optimization of water flooding with smart wells using optimal control theory. *SPE Journal*, 9(4):391–402, 2004.

[14] Eric C Bryant and WaiChing Sun. A micromorphically regularized cam-clay model for capturing size-dependent anisotropy of geomaterials. *Computer Methods in Applied Mechanics and Engineering*, 354:56–95, 2019.

[15] SJ Buckley, JA Howell, HD Enge, BLS Leren, and TH Kurz. Integration of terrestrial laser scanning, digital photogrammetry and geostatistical methods for high-resolution modelling of geological outcrops. *Remote Sensing and Spatial Information Science*, 36, 2006.

[16] Gregery T Buzzard, Stanley H Chan, Suhas Sreehari, and Charles A Bouman. Plug-and-play unplugged: Optimization-free reconstruction using consensus equilibrium. *SIAM Journal on Imaging Sciences*, 11(3):2001–2020, 2018.

[17] Jef Caers. History matching under training-image-based geological model constraints. *SPE journal*, 8(03):218–226, 2003.

[18] R. E. Caflisch. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7:1–49, 1998.

[19] David A Cameron, Louis J Durlofsky, and Sally M Benson. Use of above-zone pressure data to locate and quantify leaks during carbon storage operations. *International Journal of Greenhouse Gas Control*, 52:32–43, 2016.

[20] Shing Chan and Ahmed H Elsheikh. Parametric generation of conditional geological realizations using generative neural networks. *Comput. Geosci.*, 23:925–952, 2019.

[21] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2016.

[22] Yan Chen and Dean S Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.

[23] V. Ciriello, I. Lauriola, S. Bonvicini, V. Cozzani, V. Di Federico, and D. M. Tartakovsky. Impact of hydrogeological uncertainty on estimation of environmental risks posed by hydrocarbon transportation networks. *Water Resour. Res.*, 53(11):8686–8697, 2017.

[24] Keith H Coats. A note on impes and some impes-based simulation models. *SPE Journal*, 5(03):245–251, 2000.

[25] Alessandro Comunian and Mauro Giudici. Hybrid inversion method to estimate hydraulic transmissivity by combining multiple-point statistics and a direct inversion method. *Mathematical Geosciences*, 50(2):147–167, 2018.

[26] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[27] G. Dagan and S. P. Neuman, editors. *Subsurface flow and transport: A stochastic approach*. Cambridge, New York, 1997.

[28] M. Dentz and D. M. Tartakovsky. Probability density functions for passive scalars dispersed in random velocity fields. *Geophys. Res. Lett.*, 37:L24406, 2010.

[29] C. V. Deutsch and A. Journel. *GSLIB: Geostatistics Software Library and User's Guide*. Oxford Univ. Press, 1998.

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

[31] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA J. Uncert. Quant.*, 3:1075–1108, 2015.

[32] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, 2016.

[33] Jonathan Eckstein and Dimitri P Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[34] Kathleen R Fowler and Carl Tim Kelley. Pseudo-transient continuation for nonsmooth nonlinear equations. *SIAM journal on numerical analysis*, 43(4):1385–1406, 2005.

[35] Olga Fuks, Fayadhoi Ibrahima, Pavel Tomin, and Hamdi A Tchelepi. Analysis of travel time distributions for uncertainty propagation in channelized porous systems. *Transport in Porous Media*, 126(1):115–137, 2019.

[36] Olga Fuks, Fayadhoi Ibrahima, Pavel Tomin, and Hamdi A Tchelepi. Uncertainty propagation for compositional flow using a probability distribution method. *Transport in Porous Media*, 132(3):113–133, 2020.

[37] Muhammad Usman Ghani and W Clem Karl. Integrating data and image domain deep learning for limited angle tomography using consensus equilibrium. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3922–3932. IEEE, 2019.

[38] M. B. Giles, T. Nagapetyan, and K. Ritter. Multilevel Monte Carlo approximation of distribution functions and densities. *SIAM/ASA J. Uncert. Quant.*, 3:267–295, 2015.

[39] Pontus Giselsson. Tight global linear convergence rate bounds for Douglas–Rachford splitting. *Journal of Fixed Point Theory and Applications*, 19(4):2241–2270, 2017.

[40] Alberto Guadagnini, Laura Guadagnini, Daniel M Tartakovsky, and C Larrabee Winter. Random domain decomposition for flow in heterogeneous stratified aquifers. *Stochastic Environmental Research and Risk Assessment*, 17(6):394–407, 2003.

[41] L. Guadagnini, A. Guadagnini, and D. M. Tartakovsky. Probabilistic reconstruction of geologic facies. *J. Hydrol.*, 294:57–67, 2004.

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

[43] F. Ibrahima, H. A. Tchelepi, and D. W. Meyer. An efficient distribution method for nonlinear two-phase flow in highly heterogeneous multidimensional stochastic porous media. *Comput. Geosci.*, 22:389–412, 2018.

[44] Fayadhoi Ibrahima, Daniel W Meyer, and Hamdi A Tchelepi. Distribution functions of saturation for stochastic nonlinear two-phase flow. *Transp. Porous Media*, 109(1):81–107, 2015.

[45] Hyung Jun Yang, Francesca Boso, Hamdi A Tchelepi, and Daniel M Tartakovsky. Probabilistic forecast of single-phase flow in porous media with uncertain properties. *Water Resources Research*, 55(11):8631–8645, 2019.

[46] Ulugbek S Kamilov, Hassan Mansour, and Brendt Wohlberg. A plug-and-play priors approach for solving nonlinear imaging inverse problems. *IEEE Signal Processing Letters*, 24(12):1872–1876, 2017.

[47] Carl Timothy Kelley and David E Keyes. Convergence analysis of pseudo-transient continuation. *SIAM Journal on Numerical Analysis*, 35(2):508–523, 1998.

[48] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[49] Peter K Kitanidis. Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10):2411–2419, 1995.

[50] Ole Klein, Olaf A Cirpka, Peter Bastian, and Olaf Ippisch. Efficient geostatistical inversion of transient groundwater flow using preconditioned nonlinear conjugate gradients. *Advances in Water Resources*, 102:161–177, 2017.

[51] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105, 2012.

[52] Eric Laloy, Romain Hérault, Diederik Jacques, and Niklas Linde. Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, 54(1):381–406, 2018.

[53] Eric Laloy, Romain Hérault, John Lee, Diederik Jacques, and Niklas Linde. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network. *Advances in Water Resources*, 110:387–405, 2017.

[54] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In *Proceedings of the 24th International Conference on Machine Learning*, pages 473–480, 2007.

[55] Chuanliang Li, Xiaofan Chen, Zhimin Du, et al. A new relationship of rock compressibility with porosity. In *SPE Asia Pacific oil and gas conference and exhibition*. Society of Petroleum Engineers, 2004.

[56] Liyong Li, Hamdi A Tchelepi, and Dongxiao Zhang. Perturbation-based moment equation approach for flow in heterogeneous porous media: Applicability range and analysis of high-order terms. *J. Comput. Phys.*, 108:296–317, 2003.

[57] Liyong Li, Hamdi A Tchelepi, and Dongxiao Zhang. Perturbation-based moment equation approach for flow in heterogeneous porous media: applicability range and analysis of high-order terms. *Journal of Computational Physics*, 188(1):296–317, 2003.

[58] Qinzhuo Liao and Dongxiao Zhang. Probabilistic collocation method for strongly nonlinear problems: 1. transform by location. *Water Resources Research*, 49(12):7911–7928, 2013.

[59] Qinzhuo Liao and Dongxiao Zhang. Probabilistic collocation method for strongly nonlinear problems: 2. transform by displacement. *Water Resources Research*, 50(11):8736–8759, 2014.

[60] Peter C Lichtner and D M Tartakovsky. Stochastic analysis of effective rate constant for heterogeneous reactions. *Stoch. Environ. Res. Risk Assess.*, 17(6):419–429, 2003.

[61] Pipat Likanapaisal, Liyong Li, and Hamdi A Tchelepi. Dynamic data integration and quantification of prediction uncertainty using statistical-moment equations. *SPE J.*, 17(1):98–111, 2012.

[62] Guang Lin, Alexandre M Tartakovsky, and Daniel M Tartakovsky. Uncertainty quantification via random domain decomposition and probabilistic collocation on sparse grids. *Journal of Computational Physics*, 229(19):6995–7012, 2010.

[63] S. Liodakis, P. Kyriakidis, and P. Gaganis. Conditional Latin hypercube simulation of (log)Gaussian random fields. *Math. Geosci.*, 50:127–146, 2018.

[64] Yimin Liu, Wenyue Sun, and Louis J Durlofsky. A deep-learning-based geological parameterization for history matching complex models. *Mathematical Geosciences*, 51(6):725–766, 2019.

[65] Gregoire Mariethoz and Jef Caers. *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons, 2014.

[66] Gregoire Mariethoz and Sylvain Lefebvre. Bridges between multiple-point geostatistics and texture synthesis: Review and guidelines for future research. *Computers & Geosciences*, 66:66–80, 2014.

[67] Florian Müller, Patrick Jenny, and Daniel W Meyer. Multilevel monte carlo for two phase flow and transport in random heterogeneous porous media. In *Zurich, Switzerland: Seminar für Angewandte Mathematik, Eidgen össische Technische Hochschule*, 2012.

[68] S. P. Neuman, D. Tartakovsky, T. C. Wallstrom, and C. Winter. 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, 1996.

[69] offshore technology. Seismic data enhancement. `https://www.offshore-technology.com/products/seismic-data-enhancement/`. Accessed: 2021-11-7.

[70] Dean S Oliver, Albert C Reynolds, and Ning Liu. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, 2008.

[71] Shunsuke Ono. Primal-dual plug-and-play image restoration. *IEEE Signal Processing Letters*, 24(8):1108–1112, 2017.

[72] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[73] Chris Paola. Quantitative models of sedimentary basin filling. *Sedimentology*, 47:121–178, 2000.

[74] David W Pollock. Semianalytical computation of path lines for finite-difference models. *Groundwater*, 26(6):743–750, 1988.

[75] Stephen B Pope. Turbulent flows, 2001.

[76] V. Raman, H. Pitsch, and R. O. Fox. Hybrid large-eddy simulation/Lagrangian filtered-density-function approach for simulating turbulent combustion. *Combust. Flame*, 143:56–78, 2005.

[77] researchgate. Heterogeneity in rock masses, howpublished = `https://www.researchgate.net/post/how-can-i-represent-heterogeneity-in-rock-masses`, note = Accessed: 2021-11-7.

[78] Michael J Ronayne, Steven M Gorelick, and Jef Caers. Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach. *Water Resources Research*, 44(8), 2008.

[79] Ernest K Ryu and Stephen Boyd. Primer on monotone operator methods. *Appl. Comput. Math.*, 15(1):3–43, 2016.

[80] Pallav Sarma, Louis J Durlofsky, and Khalid Aziz. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical Geosciences*, 40(1):3–32, 2008.

[81] Pallav Sarma, Louis J Durlofsky, Khalid Aziz, and Wen H Chen. Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Computational Geosciences*, 10(1):3–36, 2006.

[82] G. Severino and S. De Bartolo. Stochastic analysis of steady seepage underneath a water-retaining wall through highly anisotropic porous media. *J. Fluid Mech.*, 778:253–272, 2015.

[83] Mohammad Shahvali, Bradley Mallison, Kaihong Wei, and Herve Gross. An alternative to streamlines for flow diagnostics on structured and unstructured grids. *SPE Journal*, 17(03):768–778, 2012.

[84] Mark Shvidler and Kenzi Karasaki. Probability density functions for solute transport in random field. *Transport in porous media*, 50(3):243–266, 2003.

[85] Suhas Sreehari, S Venkat Venkatakrishnan, Brendt Wohlberg, Gregery T Buzzard, Lawrence F Drummy, Jeffrey P Simmons, and Charles A Bouman. Plug-and-play priors for bright field electron tomography and sparse interpolation. *IEEE Transactions on Computational Imaging*, 2(4):408–423, 2016.

[86] Venkatesh Sridhar, Xiao Wang, Gregery Buzzard, and Charles Bouman. Distributed iterative CT reconstruction using multi-agent consensus equilibrium. *IEEE Transactions on Computational Imaging*, 6:1153–1166, 2020.

[87] Sebastien Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002.

[88] Yu Sun, Brendt Wohlberg, and Ulugbek S Kamilov. An online plug-and-play algorithm for regularized image reconstruction. *IEEE Transactions on Computational Imaging*, 5(3):395–408, 2019.

[89] Meng Tang, Yimin Liu, and Louis J Durlofsky. Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, 376:113636, 2021.

[90] D. M. Tartakovsky. Probabilistic risk analysis in subsurface hydrology. *Geophys. Res. Lett.*, 34:L05404, 2007.

[91] D. M. Tartakovsky. Assessment and management of risk in subsurface hydrology: A review and perspective. *Adv. Water Resour.*, 51:247–260, 2013.

[92] D. M. Tartakovsky and S. P. Neuman. Transient flow in bounded randomly heterogeneous domains 1. Exact conditional moment equations and recursive approximations. *Water Resour. Res.*, 34(1):1–12, 1998.

[93] D. M. Tartakovsky and S. P. Neuman. Transient flow in bounded randomly heterogeneous domains 2. Localization of conditional mean equations and temporal nonlocality effects. *Water Resour. Res.*, 34(1):13–20, 1998.

[94] D. M. Tartakovsky, B. E. Wohlberg, and A. Guadagnini. Nearest neighbor classification for facies delineation. *Water Resour. Res.*, 34:L05404, 2007.

[95] Daniel M Tartakovsky and Svetlana Broyda. Pdf equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of contaminant hydrology*, 120:129–140, 2011.

[96] Daniel M Tartakovsky and Pierre A Gremaud. Method of distributions for uncertainty quantification. *Handbook of Uncertainty Quantification*, pages 1–22, 2016.

[97] S. Taverniers, S. B. M. Bosma, and D. M. Tartakovsky. Accelerated multilevel Monte Carlo with kernel-based smoothing and Latinized stratification. *Water Resour. Res.*, 56(9):e2019WR026984, 2020.

[98] S. Taverniers and D. M. Tartakovsky. Estimation of distributions via multilevel monte carlo with stratified sampling. *J. Comput. Phys.*, 419:109572, 2020.

[99] Marco R Thiele. Streamline simulation. In *6th International Forum on Reservoir Simulation*, pages 3–7. Schloss Fuschl, Austria Society of Petroleum Engineers, 2001.

[100] USGS. Slabbed drill core display. https://www.usgs.gov/media/images/slabbed-drill-core-display. Accessed: 2021-11-7.

[101] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 945–948, Austin, TX, USA, 2013.

[102] Daniele Venturi, Daniel M Tartakovsky, Alexandre M Tartakovsky, and George E Karniadakis. Exact pdf equations and closure approximations for advective-reactive transport. *Journal of Computational Physics*, 243:323–343, 2013.

[103] Jacques Villermaux and Laurent Falk. A generalized mixing model for initial contacting of reactive fluids. *Chem. Engrg. Sci.*, 49(24):5127–5140, 1994.

[104] Hai X Vo and Louis J Durlofsky. A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models. *Mathematical Geosciences*, 46(7):775–813, 2014.

[105] P. Wang, D. M. Tartakovsky, Jr. K. D. Jarman, and A. M. Tartakovsky. CDF solutions of Buckley-Leverett equation with uncertain parameters. *Multiscale Model. Simul.*, 11(1):118–133, 2013.

[106] P. Wang, L. Yang, H. Wang, D. M. Tartakovsky, and S. Onori. Temperature estimation from current and voltage measurements in lithium-ion battery systems. *J. Power Storage*, 34:102133, 2021.

[107] Jeremy T White. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environmental Modelling & Software*, 109:191–201, 2018.

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

[109] C. L. Winter and D. M. Tartakovsky. Groundwater flow in heterogeneous composite aquifers. *Water Resour. Res.*, 38(8), 2002.

[110] C. L. Winter, D. M. Tartakovsky, and A. Guadagnini. Moment equations for flow in highly heterogeneous porous media. *Surv. Geophys.*, 24(1):81–106, 2003.

[111] C Larrabee Winter and Daniel M Tartakovsky. Mean flow in composite porous media. *Geophysical Research Letters*, 27(12):1759–1762, 2000.

[112] Brendt Wohlberg, Daniel M Tartakovsky, and Alberto Guadagnini. Subsurface characterization with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 44(1):47–57, 2005.

[113] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, 2010.

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

[115] Xiaojian Xu, Jiaming Liu, Yu Sun, Brendt Wohlberg, and Ulugbek S. Kamilov. Boosting the performance of plug-and-play priors via denoiser scaling. Technical Report 2002.11546, arXiv, February 2020.

[116] H.-J. Yang, F. Boso, H. A. Tchelepi, and D. M. Tartakovsky. Probabilistic forecast of single-phase flow in porous media with uncertain properties. *Water Resour. Res.*, 55(11):8631–8645, 2019.

[117] H.-J. Yang, F. Boso, H. A. Tchelepi, and D. M. Tartakovsky. Method of distributions for quantification of geologic uncertainty in flow simulations. *Water Resour. Res.*, 56(7):e2020WR027643, 2020.

[118] Hyung Jun Yang, Francesca Boso, Hamdi A Tchelepi, and Daniel M Tartakovsky. Method of distributions for quantification of geologic uncertainty in flow simulations. *Water Resources Research*, 56(7):e2020WR027643, 2020.

[119] Hyung Jun Yang, Youzuo Lin, Brendt Wohlberg, and Daniel M Tartakovsky. Consensus equilibrium for subsurface delineation. *Water Resources Research*, 57(10):e2021WR030151, 2021.

[120] Dongxiao Zhang, Liyong Li, and HA Tchelepi. Stochastic formulation for uncertainty analysis of two-phase flow in heterogeneous reservoirs. *Spe Journal*, 5(01):60–70, 2000.

[121] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[122] Z. Zhou and D. M. Tartakovsky. Markov chain Monte Carlo with neural network surrogates: Application to contaminant source identification. *Stoch. Environ. Res. Risk Assess.*, 35(3):639–651, 2021.