Water Resources Research®

RESEARCH ARTICLE

Key Points:

- We present a sampling-free method for probabilistic forecast of immiscible two-phase flow in heterogeneous porous media
- For the same accuracy, cumulative distribution function-frozen streamline method (CDF-FROST) is 10 times faster than Monte Carlo
- CDF-FROST provides maps of exceedance probability for use in risk assessment

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. M. Tartakovsky,
tartakovsky@stanford.edu

Citation:

Author Contributions:

Conceptualization: Hamdi A. Tchelepi, Daniel M. Tartakovsky
Data curation: Hyung Jun Yang
Formal analysis: Hyung Jun Yang
Funding acquisition: Hamdi A. Tchelepi, Daniel M. Tartakovsky
Investigation: Hyung Jun Yang
Methodology: Hyung Jun Yang, Hamdi A. Tchelepi, Daniel M. Tartakovsky
Software: Hyung Jun Yang
Supervision: Hamdi A. Tchelepi, Daniel M. Tartakovsky
Writing – original draft: Hyung Jun Yang, Hamdi A. Tchelepi
Writing – review & editing: Daniel M. Tartakovsky

Method of Distributions for Two-Phase Flow in Heterogeneous Porous Media

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

[1]Department of Energy Science and Engineering, Stanford University, Stanford, CA, USA

Abstract Multiscale heterogeneity and insufficient characterization data for a specific subsurface formation of interest render predictions of multi-phase fluid flow in geologic formations highly uncertain. Quantification of the uncertainty propagation from the geomodel to the fluid-flow response is typically done within a probabilistic framework. This task is computationally demanding due to, for example, the slow convergence of Monte Carlo simulations (MCS), especially when computing the tails of a distribution that are necessary for risk assessment and decision-making under uncertainty. The frozen streamlines method (FROST) accelerates probabilistic predictions of immiscible two-phase fluid flow problems; however, FROST relies on MCS to compute the travel-time distribution, which is then used to perform the transport (phase saturation) computations. To alleviate this computational bottleneck, we replace MCS with a deterministic equation for the cumulative distribution function (CDF) of travel time. The resulting CDF-FROST approach yields the CDF of the saturation field without resorting to sampling-based strategies. Our numerical experiments demonstrate the high accuracy of CDF-FROST in computing the CDFs of both saturation and travel time. For the same accuracy, it is about 5 and 10 times faster than FROST and MCS, respectively.

Plain Language Summary Multiscale heterogeneity and insufficient characterization data for a specific subsurface formation of interest render predictions of multi-phase fluid flow in geologic formations highly uncertain. Quantification of the uncertainty propagation from the geomodel to the fluid-flow response is typically done within a probabilistic framework. This task is computationally demanding, especially when computing the tails of a distribution that are necessary for risk assessment and decision-making under uncertainty. Our method accelerates this computation by several orders of magnitude.

1. Introduction

Quantitative predictions of fluid flow and transport in the subsurface are a key component of proper risk assessment and decision-making in many applications including water resources management, extraction of geothermal energy, geologic carbon sequestration, and remediation of contaminated soils and aquifers (Dagan & Neuman, 1997; Tartakovsky, 2013). Such predictions typically rely on partial differential equations (PDEs) that represent fundamental conservation laws. Parameters in these PDEs reflect relevant properties of the subsurface formation, which is often heterogeneous; moreover, and the available measurements (e.g., permeability from cores) are invariably sparse. In the case of multi-phase flow, reactive transport, and other nonlinear phenomena, these PDEs also involve phenomenological constitutive relations that describe the rock-fluid interactions (e.g., the relative permeability relations). Consequently, model predictions must be accompanied by robust quantification of the predictive uncertainty (Tartakovsky & Winter 2008).

Within the probabilistic framework, uncertain model inputs and solutions of the corresponding PDEs are treated as random fields/processes. In other words, rather than having a single deterministic solution, one ends up with a large number of predictions each of which is consistent with the equations, but some predictions are more likely than others. The probability of a given prediction to be correct is described by the cumulative distribution function (CDF), or its derivative, the probability density function (PDF). Because of their high computational cost, these distributions are often replaced with a few statistical moments, such as the mean and variance of the corresponding PDF/CDF. The mean (first moment) serves as the "average" prediction, while the variance (second moment) provides a measure of the predictive uncertainty. We focus on the computation of PDFs/CDFs, rather than their first two moments. This is necessary to estimate the probability of occurrence of rare events, which, in turn, is used for quantitative risk assessment (Tartakovsky, 2013).

Monte Carlo simulations (MCS) are often used to propagate parametric uncertainty through the modeling process. They are robust, easy to implement, and perfectly parallelizable. They are also computationally expensive, especially when used to estimate full PDFs/CDFs due to the typically slow convergence rate of the computed PDFs/CDFs (Ibrahima et al., 2018; Yang et al., 2019). For MCS to be accurate, the number of realizations can be extremely large, and each realization requires high-resolution discretization to deal with the wide variations in the coefficients (e.g., permeability). 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) (Müller et al., 2012), but might become slower than standard MCS when used to estimate full PDFs/CDFs (Taverniers et al., 2020; Taverniers & Tartakovsky, 2020, and references therein). 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 (Barajas-Solano & Tartakovsky, 2016). 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 (Tartakovsky & Gremaud, 2016) has the 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 (Yang et al., 2019, 2020), and to yield accurate approximations of the PDF/CDF of the concentration of solutes undergoing geochemical transformations during their migration in the subsurface (Boso et al., 2014, 2018; Tartakovsky & Broyda, 2011). While the former class of problems is linear—see, also, Dentz and Tartakovsky (2010) and Boso and Tartakovsky (2016) 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.

Shocks and discontinuous state variables typical of immiscible multiphase flow problems, for example, fluid saturation whose dynamics are described by the Buckley-Leverett equation, pose a challenge for the method of distributions (and other uncertainty quantification techniques). They require either the analytical computation of the shock dynamics (Wang et al., 2013) or the introduction of the so-called kinetic defect that has to be inferred from data (Boso & Tartakovsky, 2020). For highly heterogeneous subsurface environments, in which the streamlines are defined largely by the geology rather than flow conditions and, hence, remain frozen in time, a version of the method of distributions named FROST (Ibrahima et al., 2015, 2018) exhibits remarkable accuracy and efficiency (Fuks et al., 2019, 2020). The input to FROST is the distribution of travel times, whose Monte Carlo computation requires a large number of flow simulations and streamline tracing. To speed-up FROST further, we eliminate the need for MCS by replacing them with a deterministic CDF equation for travel time.

We start by formulating a two-phase immiscible flow problem with uncertain inputs in Section 2. Its treatment with the original FROST method is reviewed in Section 3, followed by the derivation of a CDF equation for travel time in Section 4. A numerical implementation of the new CDF-FROST method is presented in Section 5. Numerical experiments presented in Section 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 7.

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 μ_{nw} and μ_w. Their mobilities in the porous medium of intrinsic permeability $k(\mathbf{x})$ are defined as

$$\lambda_{nw} = \frac{k(\mathbf{x})k_{nw}^r(S_{nw})}{\mu_{nw}}, \qquad \lambda_w = \frac{k(\mathbf{x})k_w^r(S_w)}{\mu_w}, \tag{1}$$

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

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

$$\mathbf{q}_{\text{nw}} = -\lambda_{\text{nw}} \nabla p, \qquad \mathbf{q}_{\text{w}} = -\lambda_{\text{w}} \nabla p, \qquad \mathbf{x} \in \Omega, \quad t > 0. \tag{2}$$

The total Darcy flux $\mathbf{q}_{\text{tot}} = \mathbf{q}_{\text{nw}} + \mathbf{q}_{\text{w}}$ satisfies the continuity condition, $-\nabla \cdot \mathbf{q}_{\text{tot}} + g = 0$, where $g(\mathbf{x}, t)$ represents fluid sources and sinks. Neglecting the compressibility of the porous medium, and accounting for Equation 2, this yields a pressure equation

$$\nabla \cdot (\lambda_{\text{tot}} \nabla p) + g = 0, \qquad \mathbf{x} \in \Omega, \quad t > 0, \tag{3}$$

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

$$p = p_{\text{inj}} \quad \text{for} \quad \mathbf{x} \in \Gamma_{\text{inj}} \qquad \text{and} \qquad p = p_{\text{prod}} \quad \text{for} \quad \mathbf{x} \in \Gamma_{\text{prod}}, \tag{4}$$

or total Darcy flux

$$\mathbf{q}_{\text{tot}} \cdot \mathbf{n} = q_{\text{inj}} \quad \text{for} \quad \mathbf{x} \in \Gamma_{\text{inj}} \qquad \text{and} \qquad \mathbf{q}_{\text{tot}} \cdot \mathbf{n} = q_{\text{prod}} \quad \text{for} \quad \mathbf{x} \in \Gamma_{\text{prod}}, \tag{5}$$

which are defined on the injection (Γ_{inj}) and production (Γ_{prod}) segments of the boundary $\partial\Omega = \Gamma_{\text{prod}} \cup \Gamma_{\text{inj}}$ 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; and $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to $\partial\Omega$.

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

$$\phi(\mathbf{x}) \frac{\partial S_{\text{w}}}{\partial t} + \mathbf{q}_{\text{tot}} \cdot \nabla f_{\text{w}} (S_{\text{w}}) = g_{\text{w}}. \tag{6}$$

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

$$S_{\text{w}}(\mathbf{x}, 0) = S_{\text{w}}^{\text{ir}}, \quad \mathbf{x} \in \Omega; \qquad S_{\text{w}}(\mathbf{x}, t) = 1 - S_{\text{nw}}^{\text{ir}}, \quad \mathbf{x} \in \Gamma_{\text{inj}}, \quad t > 0. \tag{7}$$

Here, S_{w}^{ir} and $S_{\text{nw}}^{\text{ir}}$ are the irreducible saturations of the wetting and non-wetting fluids, respectively. We assume that the source or sink is located only at the boundaries, Γ_{inj} and Γ_{prod}, which is equivalent to setting $g_{\text{w}} = 0$ in Equation 6.

Equations 3–7 govern the spatiotemporal evolution of the two state variables, fluid pressure $p(\mathbf{x}, t)$ and saturation of the wetting phase $S_{\text{w}}(\mathbf{x}, t)$. 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. (This treatment is in line with previous studies (Tartakovsky et al., 2003; Wang et al., 2013) and is motivated by the observation that spatial variability of $k(\mathbf{x})$ is largely responsible for flow patterns. Uncertainty in the other parameters can be accounted for as well, at the cost of increased mathematical complexity (Tartakovsky et al., 1999).) 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, $S_{\text{w}}(\mathbf{x}, t)$, 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 $s \in \left[S_{\text{w}}^{\text{ir}}, 1 - S_{\text{nw}}^{\text{ir}}\right]$, $\mathbb{P}\{S_{\text{w}}(\mathbf{x}, t) \leq s\}$. The latter is the definition of the single-point CDF of S_{w}, that is, $\mathbb{P}\{S_{\text{w}}(\mathbf{x}, t) \leq s\} \equiv F_{S_{\text{w}}}(s; \mathbf{x}, t)$.

3. FROST Method

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

$$\mathbf{x}_{\text{sl}}(\tau, t) = \boldsymbol{\xi} + \int_0^\tau \mathbf{q}_{\text{tot}}\big(\mathbf{x}_{\text{sl}} (t'), t\big) \mathrm{d}t'. \tag{8}$$

We parameterize this streamline by a natural coordinate r, such that $\mathrm{d}r$ is the distance traveled by a particle along this streamline during the TOF interval $\mathrm{d}\tau$. Recall that the TOF $\tau(r)$ at time t is defined as the time required

for a particle to arrive at location r along a given streamline. It is computed as (Datta-Gupta & King, 2007; Thiele, 2001)

$$\tau(r,t) = \int_0^r \frac{\phi(r')\,\mathrm{d}r'}{|\mathbf{q}_{\mathrm{tot}}(r',t)|}. \tag{9}$$

The TOF interval $\mathrm{d}\tau$ is represented as $\mathrm{d}r = |\mathbf{q}_{\mathrm{tot}}|\mathrm{d}\tau$.) A collection of streamlines forms a streamtube with a variable cross-sectional area $A(r)$; the volumetric flow rate in this streamtube is $Q_{\mathrm{tube}} = |\mathbf{q}_{\mathrm{tot}}|A(r)$. According to the definition of a streamtube, $\partial Q_{\mathrm{tube}}/\partial r = 0$ so that Q_{tube} is independent of r, that is, $Q_{\mathrm{tube}} = Q_{\mathrm{tube}}(t)$.

With these definitions, we rewrite the saturation Equation 6 in the streamline coordinate system,

$$\phi(r)A(r)\frac{\partial S_{\mathrm{w}}(r,t)}{\partial t} + Q_{\mathrm{tube}}\frac{\partial f_{\mathrm{w}}(S_{\mathrm{w}}(r,t))}{\partial r} = 0, \qquad r > 0, \quad t > 0. \tag{10}$$

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

$$Q(t) = \int_0^t Q_{\mathrm{tube}}(t')\mathrm{d}t', \qquad V(r) = \int_0^r \phi(r')\,A(r')\mathrm{d}r'. \tag{11}$$

Then, with Q playing the role of time and V of spatial coordinate, $S_{\mathrm{w}}(r,t) \mapsto S_{\mathrm{w}}(V,Q)$ and Equation 10 takes the form of a one-dimensional Buckley-Leverett equation,

$$\frac{\partial S_{\mathrm{w}}}{\partial Q} + \frac{\partial f_{\mathrm{w}}(S_{\mathrm{w}})}{\partial V} = 0, \qquad V > 0, \quad Q > 0. \tag{12}$$

The initial and boundary conditions in Equation 7 map onto

$$S_{\mathrm{w}}(V,0) = S_{\mathrm{w}}^{\mathrm{ir}}, \quad S_{\mathrm{w}}(0,Q) = 1 - S_{\mathrm{nw}}^{\mathrm{ir}}. \tag{13}$$

Hewett and Yamada (1997) proved that the analytical solution of Equation 12, S_{w}, can be expressed in terms of $Z = V/Q$. This analytical mapping translates the uncertainty in Z to the stochastic characteristic of S_{w}.

The frozen streamline assumption implies that the direction of the total flux $\mathbf{q}_{\mathrm{tot}}(r,t)$ is fixed in time. With this assumption, the cross-sectional area $A(r)$ is time-independent and given by $A(r) = Q_{\mathrm{tube}}(0)/|\mathbf{q}_{\mathrm{tot}}(r,0)|$. Hence, the ratio $Z = V/Q$ becomes the ratio between TOF at time $t = 0$, $\tau_0(\mathbf{x})$, and the equivalent injection time (EIT), $T_{\mathrm{inj}}(\mathbf{x},t)$,

$$Z(\mathbf{x},t) = \frac{\tau_0}{T_{\mathrm{inj}}}, \quad \tau_0(\mathbf{x}) = \int_0^{r(\mathbf{x})} \frac{\phi(r')\,\mathrm{d}r'}{|\mathbf{q}_{\mathrm{tot}}(r',0)|}, \quad T_{\mathrm{inj}}(\mathbf{x},t) = \int_0^t \frac{Q_{\mathrm{tube}}(t')}{Q_{\mathrm{tube}}(t=0)}\mathrm{d}t'. \tag{14}$$

Regardless of spatial heterogeneity, the EIT is approximately uniform in space and can be approximated by its mean value $\bar{T}_{\mathrm{inj}}$ from a relatively few MC realizations, that is, $T_{\mathrm{inj}}(\mathbf{x},t) \approx T_{\mathrm{inj}}(t) \approx \bar{T}_{\mathrm{inj}}(t)$ (Ibrahima et al., 2018). This approximation allows one to express the CDF of the wetting-phase saturation, $F_{S_{\mathrm{w}}}(s;\mathbf{x},t)$, in terms of the CDF $F_Y(y;\mathbf{x})$ of the logarithm of TOF, $Y(\mathbf{x}) = \ln \tau_0(\mathbf{x})$,

$$F_{S_{\mathrm{w}}}(s;\mathbf{x},t) = 1 - F_Y(z;\mathbf{x}), \qquad z = \ln\left[S_{\mathrm{w}}^{-1}(s)\bar{T}_{\mathrm{inj}}(t)\right]. \tag{15}$$

Since the fractional flow $f_{\mathrm{w}}(S_{\mathrm{w}})$ is a sigmoidal function, the deterministic inverse mapping, $S_{\mathrm{w}}^{-1}(s)$, is provided by the analytical solution of Equation 12,

$$S_{\mathrm{w}}^{-1}(s) = \begin{cases} \infty, & s < S_{\mathrm{w}}^{\mathrm{ir}} \\ f_{\mathrm{w}}'(s^*), & s \in \left(S_{\mathrm{w}}^{\mathrm{ir}}, s^*\right) \\ f_{\mathrm{w}}'(s), & s \in \left(s^*, 1 - S_{\mathrm{nw}}^{\mathrm{ir}}\right) \\ 0, & s > 1 - S_{\mathrm{nw}}^{\mathrm{ir}}, \end{cases} \tag{16}$$

where $f_{\mathrm{w}}'(\cdot)$ designates the derivative of $f_{\mathrm{w}}(\cdot)$, and s^* satisfies the Rankine-Hugoniot jump condition (Aziz, 1979; Lax, 1973),

$$f'_w(s^*) = \frac{f_w(s^*) - f_w(S_w^{ir})}{s^* - S_w^{ir}}. \tag{17}$$

We designate the original version of FROST by MCS-FROST to emphasize its reliance on MCS to estimate both $\bar{T}_{inj}(t)$ 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, $N_{MC} \geq 5,000$). Our version, CDF-FROST, replaces this computational bottleneck with a numerical solution of the deterministic equation for $F_Y(y; \mathbf{x})$.

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, ranking of geomodels, and optimization of well placement or well operation conditions. 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 (Shahvali et al., 2012)

$$\mathbf{q}_{tot}(\mathbf{x}, t) \cdot \nabla \tau(\mathbf{x}) = \phi(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{18}$$

subject to Dirichlet boundary condition

$$\tau(\mathbf{x}) = 0, \qquad \mathbf{x} \in \Gamma_{inj}. \tag{19}$$

Randomness in the intrinsic permeability $k(\mathbf{x})$ translates into randomness of $\mathbf{q}_{tot}(\mathbf{x}, t)$, 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 Equation 18 and Equation 19.

4.1. Derivation of the 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

$$\mathbb{E}\{\mathcal{H}(\mathcal{T} - \tau(\mathbf{x}))\} = F_\tau(\mathcal{T}; \mathbf{x}) \equiv \mathbb{P}\{\tau(\mathbf{x}) \leq \mathcal{T}\}. \tag{20}$$

Multiplying both sides of Equation 18 with $\partial\Pi/\partial\mathcal{T}$, and noting that $\nabla\Pi = -(\partial\Pi/\partial\mathcal{T})\nabla\tau$, yields an equation for $\Pi(\mathcal{T}, \mathbf{x})$,

$$\mathbf{q}_{tot} \cdot \nabla\Pi + \phi\frac{\partial\Pi}{\partial\mathcal{T}} = 0. \tag{21}$$

We use the Reynolds decomposition to represent the random quantities $\mathbf{q}_{tot}$ and Π as the sum of their ensemble means and zero-mean fluctuation around them, $\mathbf{q}_{tot} = \bar{\mathbf{q}}_{tot} + \mathbf{q}'_{tot}$ and $\Pi = F_\tau + \Pi'$. (Throughout this paper, we use $\mathbb{E}\{\cdot\}$, $\bar{\cdot}$, and $\langle\cdot\rangle$ interchangeably to represent the ensemble mean.) Then, the ensemble average of Equation 21 yields an unclosed equation for $F_\tau(\mathcal{T}; \mathbf{x})$,

$$\bar{\mathbf{q}}_{tot} \cdot \nabla F_\tau + \phi\frac{\partial F_\tau}{\partial\mathcal{T}} + \langle\mathbf{q}'_{tot} \cdot \nabla\Pi'\rangle = 0. \tag{22}$$

A closure approximation for the unknown cross-correlation term $\langle\mathbf{q}'_{tot} \cdot \nabla\Pi'\rangle$ is necessary to render Equation 22 computable.

The large-eddy-diffusivity closure exhibits good accuracy and robustness for advection-reaction problems (Tartakovsky & Broyda, 2011; Venturi et al., 2013). 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 (Boso et al., 2018; Boso & Tartakovsky, 2016; Yang et al., 2019, 2020),

$$\langle\mathbf{q}'_{tot} \cdot \nabla\Pi'\rangle \approx \left[\alpha(\mathbf{x})\left(\mathcal{T} - \bar{\tau}\right) + \beta(\mathbf{x})\right]\frac{\partial F_\tau}{\partial\mathcal{T}}, \tag{23}$$

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

$$\bar{\mathbf{q}}_{\text{tot}} \cdot \nabla F_\tau + \left[\alpha \left(\mathcal{T} - \bar{\tau} \right) + \beta + \phi \right] \frac{\partial F_\tau}{\partial \mathcal{T}} = 0. \tag{24}$$

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 $\bar{\mathbf{x}}_{\text{sl}}(\tau_{\text{mean}})$. Equation 9 establishes the duality between TOF along a mean trajectory, τ_{mean}, and the position along the mean streamline traveled during that time, r. Similarly, when the mean velocity $\bar{\mathbf{q}}_{\text{tot}}$ is given, $\bar{\mathbf{x}}_{\text{sl}}$ and τ_{mean} are defined as

$$\bar{\mathbf{x}}_{\text{sl}}(\tau_{\text{mean}}) = \bar{x}_{\text{sl},0} + \int_0^{\tau_{\text{mean}}} \bar{\mathbf{q}}_{\text{tot}}\left(r'\right) dr', \quad \tau_{\text{mean}}(r) = \int_0^r \frac{\phi(s') ds'}{\bar{q}_{\text{tot}}(s')}, \tag{25}$$

where $\bar{x}_{\text{sl},0}$ is the mean origin point. 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, $\bar{\mathbf{q}}_{\text{tot}} \cdot \nabla = \bar{q}_{\text{tot}} \partial / \partial r$, where $\bar{q}_{\text{tot}} \equiv |\bar{\mathbf{q}}_{\text{tot}}|$ denotes the magnitude of the mean velocity vector $\bar{\mathbf{q}}_{\text{tot}}$; $F_\tau(\mathcal{T}; \mathbf{x}) \mapsto F_\tau(\mathcal{T}; r)$; and Equation 24 takes the form,

$$\bar{q}_{\text{tot}} \frac{\partial F_\tau}{\partial r} + \left[\alpha(r) \left(\mathcal{T} - \bar{\tau} \right) + \beta(r) + \phi(r) \right] \frac{\partial F_\tau}{\partial \mathcal{T}} = 0. \tag{26}$$

We rewrite this CDF equation in a "conservative" form,

$$\frac{\partial \left(\bar{q}_{\text{tot}} F_\tau \right)}{\partial r} + \frac{\partial \left(U F_\tau \right)}{\partial \mathcal{T}} = \left[\alpha + \frac{\partial \bar{q}_{\text{tot}}}{\partial r} \right] F_\tau, \qquad U(r, \mathcal{T}) \equiv \alpha(\mathcal{T} - \bar{\tau}) + \beta + \phi, \tag{27}$$

which is more conducive to the subsequent numerical treatment. This CDF equation is subject to the boundary conditions derived from Equation 19,

$$F_\tau(\mathcal{T}; r = 0) = 1, \quad \mathcal{T} \geq 0. \tag{28}$$

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

$$F_\tau(0; r) = 0, \qquad F_\tau(\tau_{\max}; r) = 1. \tag{29}$$

Following Boso and Tartakovsky (2016) and Yang et al. (2019), we determine expressions for the closure variables $\alpha(r)$ and $\beta(r)$ in Equation 27 by enforcing the consistency between the moments $\bar{\tau}$ and σ_τ^2 obtained alternatively by integration of Equation 27 and by either using MCS or solving the statistical moment equations (SMEs) derived in Section S1 in Supporting Information S1. This procedure yields (Appendix A)

$$\alpha(r) = \frac{q_{\text{tot}}}{2\sigma_\tau^2} \frac{\partial \sigma_\tau^2}{\partial r}, \qquad \beta(r) = q_{\text{tot}} \frac{\partial \bar{\tau}}{\partial r} - \phi. \tag{30}$$

These expressions relate the closure variables $\alpha(r)$ and $\beta(r)$ to the spatial variability (gradients) of the mean and variance of the TOF, σ_τ^2 and $\bar{\tau}$. We use the SMEs to compute $\bar{\tau}$ and σ_τ^2, because this strategy outperforms MCS in terms of computational cost for the same accuracy (Likanapaisal et al., 2012).

The solution to the two-dimensional boundary-value problem (BVP) in Equations 27–29 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 $\mathcal{O}\left(N_{\text{sl}} \cdot N_r^3 \cdot N_{\mathcal{T}}^3 \right)$. We contrast this with the computational complexity of solving the corresponding CDF equation in the Cartesian coordinate system, that is, solving Equation 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 Equation 24 of $\mathcal{O}\left(N_x^3 \cdot N_y^3 \cdot N_{\mathcal{T}}^3 \right)$. 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
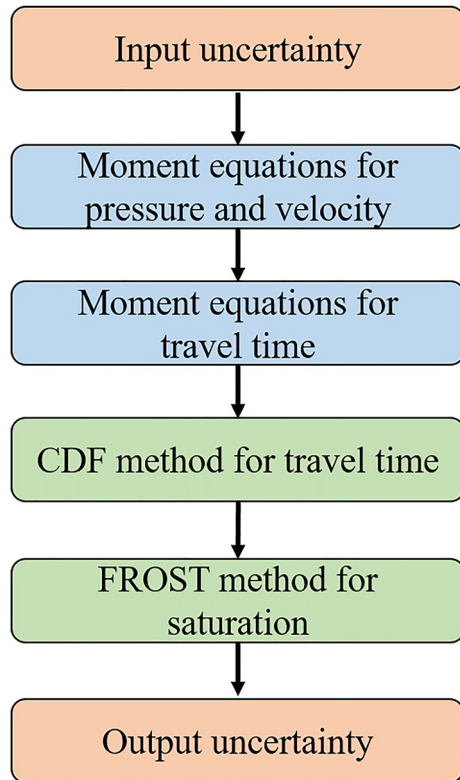
Figure 1. Workflow of the proposed cumulative distribution function-frozen streamline method (CDF-FROST) method.

calculations, scales linearly with the size of Ω (Thiele, 2001). As a result, our method can be used for high-resolution and/or three-dimensional simulations without any significant increase in computational cost.

5. Numerical Implementation of CDF-FROST

In our approach, the CDF of saturation is evaluated in three steps (Figure 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 (Section S1 in Supporting Information S1). The second step involves numerical solution of BVP Equations 27–29 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 $F_{S_w}(s; \mathbf{x}, t)$.

The first step is carried out with the research code of Likanapaisal et al. (2012). 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 Equation 28 precludes the use of a high-order scheme without inducing spurious oscillations. The non-smoothness of the coefficients U and α in Equation 27 requires 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 exponential grid spacing. These are detailed in Section S2 of Supporting Information S1.

6. Numerical Experiments

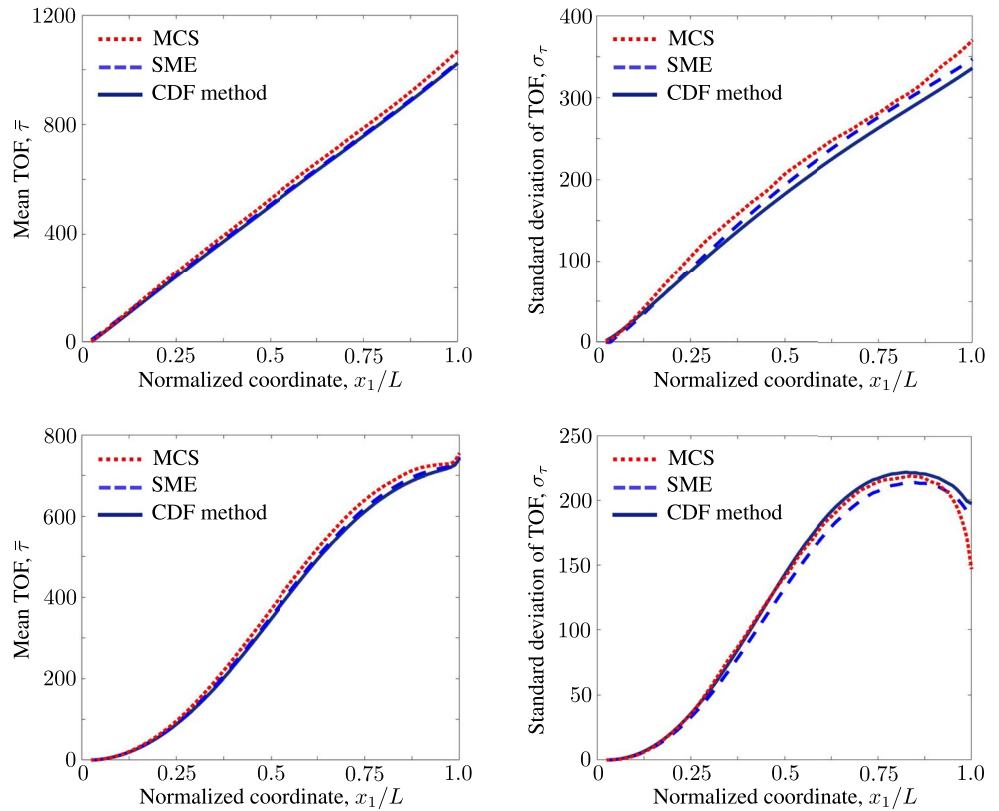We use two sets of numerical experiments to demonstrate the accuracy, robustness, and versatility of CDF-FROST. In both cases, the wetting phase is injected into a two-dimensional flow domain Ω, which is initially filled with non-wetting phase. The viscosities of the non-wetting and wetting phases are $\mu_{\text{nw}} = 2$ cP and $\mu_{\text{w}} = 1$ cP, respectively, thus the viscosity ratio $M = \mu_{\text{nw}}/\mu_{\text{w}} = 2$. The quadratic model is used for relative permeabilities $k_{\text{w}}^{\text{r}} = \left(S_{\text{w}} - S_{\text{w}}^{\text{ir}}\right)^2$ and $k_{\text{nw}}^{\text{r}} = (1 - S_{\text{w}})^2$, with $S_{\text{w}}^{\text{ir}} = 0.1$ and $S_{\text{nw}}^{\text{ir}} = 0$. The porosity ϕ is assumed to be uniform $\phi = 0.1$ over the computational domain. The spatial domain Ω, a square of length $L = 1{,}000$ 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 $N_{\text{sl}} = 300$. 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 $\sigma_\kappa^2 = 1$, an isotropic exponential covariance function $C_\kappa = \sigma_\kappa^2 \exp\left(-|\mathbf{x} - \mathbf{y}|/\ell_\kappa\right)$, 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 of 100 Pa. These examples are motivated by such activities as CO_2 sequestration and remediation of groundwater contaminated with non-aqueous phase liquids. Predictive uncertainty inherent in these and other subsurface models requires one to estimate the probability of failure of the operations guided by the model predictions; the latter is encoded in travel-time CDF $F_\tau(\mathcal{T}; \mathbf{x})$ and saturation CDF $F_{S_w}(s; \mathbf{x}, t)$.

We compare our estimates of $F_\tau(\mathcal{T}; \mathbf{x})$ and $F_{S_w}(s; \mathbf{x}, t)$ with their counterparts computed via high-resolution MCS. The latter employ the standard Pollock's method (Pollock, 1988) and the finite-volume solver implementing an implicit pressure, explicit saturation (IMPES) scheme (Coats, 2000) 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

Figure 2. Mean (left column) and standard deviation (right column) of time of flight (TOF), $\bar{\tau}(x_1, x_2 = 1/2)$ and $\sigma_\tau(x_1, x_2 = 1/2)$, 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 cumulative distribution function (CDF) method.

total simulation time is 1,000 days. MC realizations of the log-permeability $\kappa(\mathbf{x})$ are generated by the sequential Gaussian simulator (Deutsch & Journel, 1998).

6.1. Accuracy of the Proposed Method

We start by analyzing the ability of our CDF equation to accurately approximate the first two moments of TOF $\tau(\mathbf{x})$, its mean $\bar{\tau}(\mathbf{x})$ and standard deviation $\sigma_\tau(\mathbf{x})$. These statistics are widely used to equip the TOF predictor, $\bar{\tau}$, with a confidence interval, for example, $\bar{\tau} \pm \sigma_\tau$. The quadratures in Equation A1 are evaluated via Gauss-Legendre quadrature rule using 500 quadrature points. Figure 2 exhibits these statistical moments, for the two flow regimes, along the cross-section $x_1 = x_2$. The profiles $\bar{\tau}(x_1, x_1)$ and $\sigma_\tau(x_1, x_1)$ are alternatively computed with MCS, the SMEs, and the CDF method.

The CDF $F_\tau(\mathcal{T}; \mathbf{x})$ in Equation 27 is constructed to have the same $\bar{\tau}(\mathbf{x})$ and $\sigma_\tau(\mathbf{x})$ as their counterparts computed via the SMEs in Section S1 of Supporting Information S1. The slight discrepancy between the two sets of the statistical moments in Figure 2 is due to numerical errors in the solution of the CDF equation and the subsequent numerical evaluation of the quadratures in Equation A1. Consistent with the previous SME-focused studies (Likanapaisal et al., 2012; Severino & De Bartolo, 2015, among many others), the mean and variance of hydraulic head computed with SMEs are in good agreement with those inferred from MCS as long as the variance of log-conductivity is below 7, which is the case in our simulations. As usual, 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 3 exhibits such maps for $\mathcal{T} = 50$ days. Visual inspection of these maps, alternatively computed with the reference
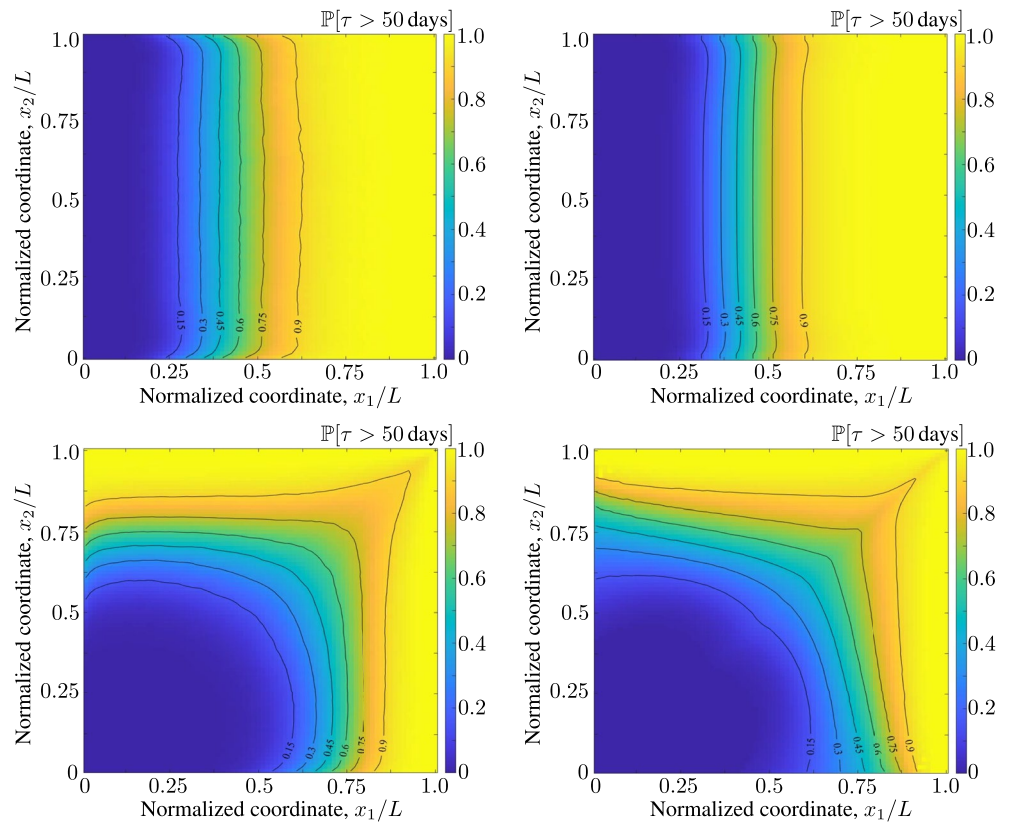
Figure 3. Spatial maps of exceedance probability, $\mathbb{P}[\tau(\mathbf{x}) > 50 \text{ days}]$, obtained with Monte Carlo simulations (MCS) (left column) and the cumulative distribution function (CDF) method (right column) for the mean uniform flow (top row) and the quarter five spot configuration (bottom row).

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 SMEs loose their accuracy. The CDFs $F_\tau(\mathcal{T}; \mathbf{x})$ presented in Figure 4 for several points $\mathbf{x} \in \Omega$ provide another illustration of the accuracy of the CDF method.

The normalized first Wasserstein distance (a.k.a. Earth Mover's metric) between two distributions (Rubner et al., 2000),

$$\mathcal{D}(\mathbf{x}) \equiv \frac{\int_0^{\tau_{\max}} |F_\tau(\mathcal{T}; \mathbf{x}) - F_\tau^{\text{MCS}}(\mathcal{T}; \mathbf{x})| \mathrm{d}\mathcal{T}}{\int_0^{\tau_{\max}} F_\tau(\mathcal{T}; \mathbf{x}) \mathrm{d}\mathcal{T}}, \tag{31}$$

provides a more quantitative assessment of the agreement between the CDFs computed with the CDF method (F_τ) and the reference MCS (F_τ^{MCS}). These integrals are computed with the Gaussian quadrature rule. The resulting contour plots of $\mathcal{D}(\mathbf{x})$ are shown in Figure 5. The error metric $\mathcal{D}(\mathbf{x})$ is small throughout much of the domain Ω, but increases 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 6 exhibits temporal snapshots of the risk maps of saturation, that is, of the exceedance probability $\mathbb{P}[S_w(\mathbf{x}, t) > 0.5] = 1 - F_{S_w}(0.5; \mathbf{x}, t)$ computed, alternatively, with CDF-FROST and MCS for the quarter-five spot configuration. We also computed similar maps for the mean uniform flow, but do not show them here. 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
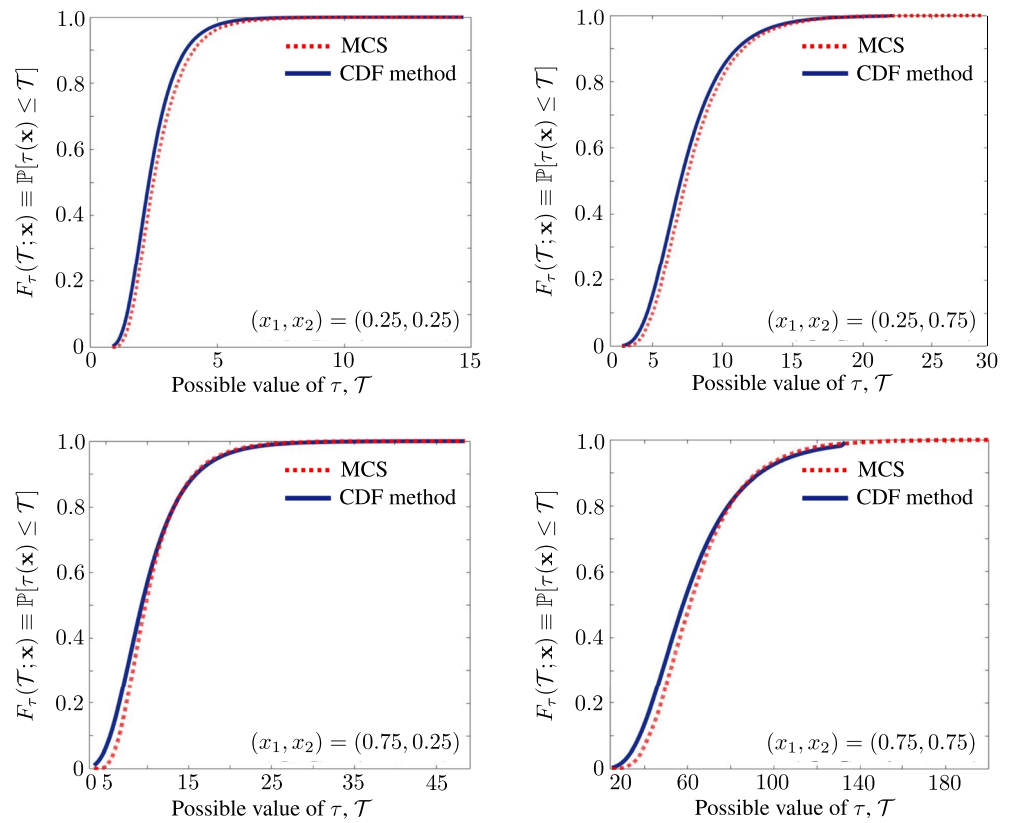
Figure 4. Travel-time cumulative distribution function (CDF), $F_\tau(\mathcal{T}; \mathbf{x})$, computed with Monte Carlo simulations (MCS) and the CDF method at selected locations $\mathbf{x} = (x_1, x_2)^\top$ in the flow domain for the mean uniform flow (top row) and the quarter-five spot configuration (bottom row).

discrepancy between the reference MCS and the CDF-FROST method is largely due to the approximation error of the SMEs (Figure 7).

Figure 8 elaborates this point further by comparing the exceedance probability maps computed with MCS, CDF-FROST with reference moments, and CDF-FROST with SMEs. As expected, the more accurate estimation of the moments leads to the better agreement between CDF-FROST and the reference MCS at the small flow velocity areas. The CDF estimates at several points $\mathbf{x} \in \Omega$ demonstrate the high accuracy of CDF-FROST.
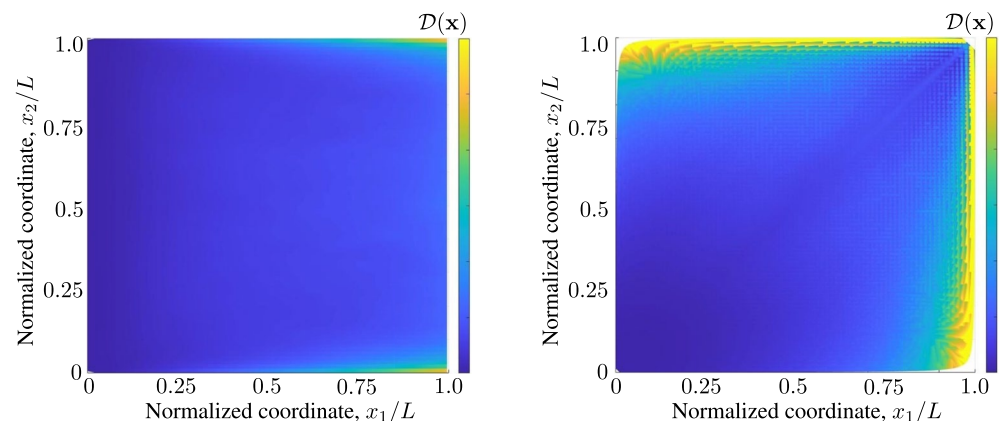
Figure 5. Spatial maps of the normalized Wasserstein distance $\mathcal{D}(\mathbf{x})$ between the hydraulic head cumulative distribution functions (CDFs) computed with the CDF method and Monte Carlo simulations for mean uniform flow (left) and quarter-five spot configuration (right).
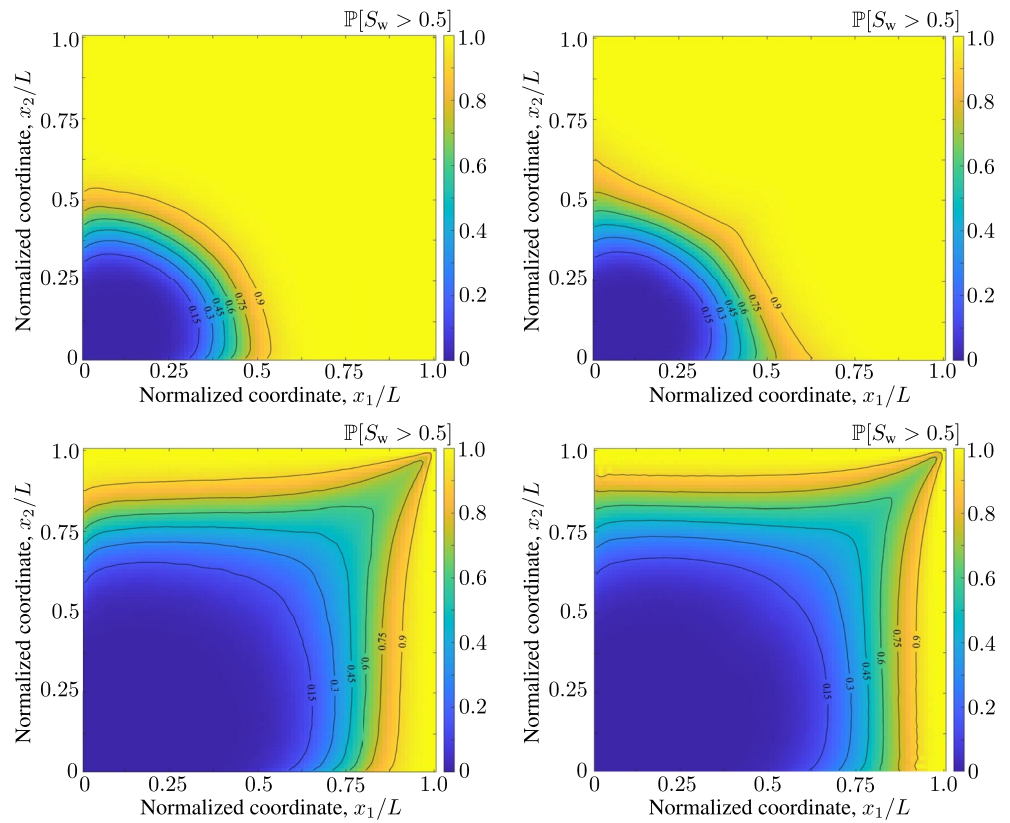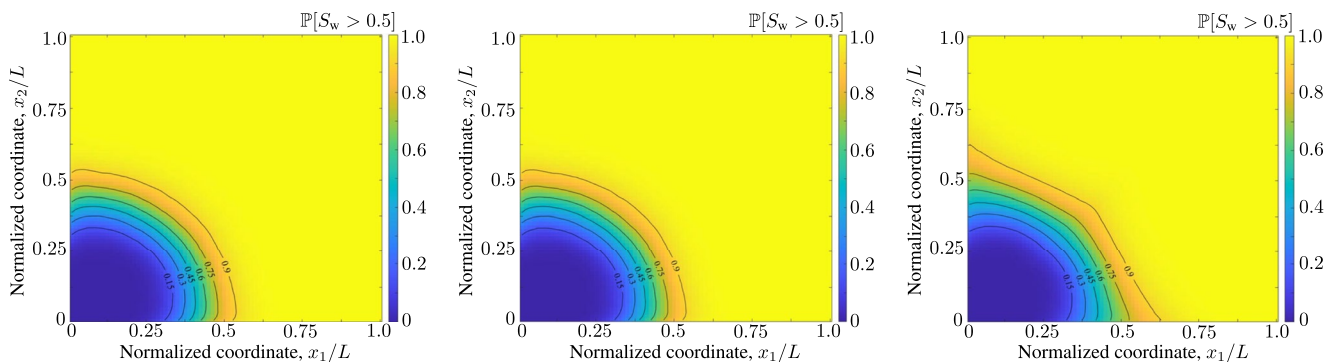
Figure 6. Temporal snapshots of the saturation's exceedance probability map $\mathbb{P}[S_w(\mathbf{x}, t) > 0.5] = 1 - F_{S_w}(0.5; \mathbf{x}, t)$ at times $t = 100$ days (top row) and $t = 500$ days (bottom row) obtained with reference Monte Carlo simulations (MCS) (left column) and cumulative distribution function-frozen streamline method (CDF-FROST) (right column) for the quarter-five spot configuration.

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 (Equation 27), while the latter requires solving a large number of d-dimensional nonlinear coupled PDEs Equations 3–7. CDF-FROST is also expected to be more efficient than MCS-FROST (Ibrahima et al., 2015, 2018), since it obviates the need for any MCS. Table 1 confirms these expectations. The CPU times are reported for the same discrepancy level, defined by the average Wasserstein distance $\mathcal{D}_{ave}$ between the saturation CDFs computed with either MCS, MCS-FROST, or CDF-FROST and the

Figure 7. Temporal snapshots of the saturation's exceedance probability map $\mathbb{P}[S_w(\mathbf{x}, t) > 0.5] = 1 - F_{S_w}(0.5; \mathbf{x}, t)$ at time $t = 100$ days obtained with reference Monte Carlo simulations (MCS) (left), cumulative distribution function-frozen streamline method (CDF-FROST) with reference moments (center), and CDF-FROST with statistical moment equations (SMEs) (right) for the quarter-five spot configuration.
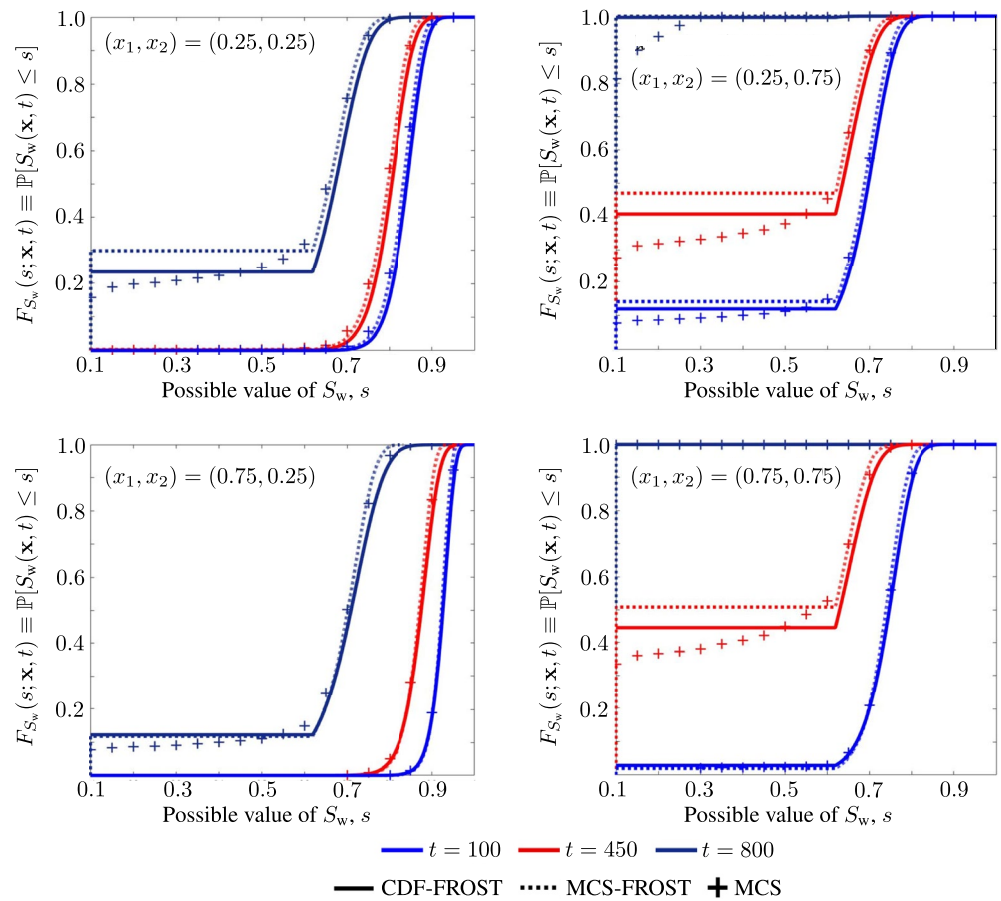
Figure 8. Saturation cumulative distribution functions (CDFs) $F_{S_w}(s; \cdot)$ computed with Monte Carlo simulations (MCS), MCS-frozen streamline method (FROST) (Ibrahima et al., 2015, 2018), and CDF-FROST for four locations $\mathbf{x} = (x_1, x_2)^\top$ at times $t = 100$ days, $t = 500$ days, and $t = 800$ days. The top and bottom rows correspond to the mean uniform flow and the quarter-five spot configuration, respectively.

reference MCS. It takes $N_{MCS} = 1{,}510$ MC realizations to achieve $\mathcal{D}_{ave} \approx 0.03$ for the mean uniform flow, and $N_{MCS} = 1860$ MC realizations to achieve $\mathcal{D}_{ave} \approx 0.05$ for the quarter-five well configuration. To achieve the same accuracy, MCS-FROST requires $N_{MCS} = 3{,}820$ MC realizations for the mean uniform flow, and $N_{MCS} = 4{,}630$ MC

Table 1
CPU Time and Accuracy of Monte Carlo Simulations (MCS), MCS-Frozen Streamline Method (FROST), and Cumulative Distribution Functions (CDF)-FROST

Flow regime	Method	Error $\mathcal{D}_{ave}$	CPU time (min)
Mean uniform flow	CDF-FROST	$3.22 \cdot 10^{-2}$	$9.55 \cdot 10^0$
	MCS-FROST with $N_{MCS} = 3{,}820$	$3.22 \cdot 10^{-2}$	$6.01 \cdot 10^1$
	MCS with $N_{MCS} = 1{,}510$	$3.22 \cdot 10^{-2}$	$1.38 \cdot 10^2$
	MCS with $N_{MCS} = 1 \cdot 10^4$	0	$8.25 \cdot 10^2$
Convergent flow	CDF-FROST	$4.58 \cdot 10^{-2}$	$1.22 \cdot 10^1$
	MCS-FROST with $N_{MCS} = 1{,}860$	$4.58 \cdot 10^{-2}$	$8.37 \cdot 10^1$
	MCS with $N_{MCS} = 4{,}630$	$4.58 \cdot 10^{-2}$	$4.65 \cdot 10^2$
	MCS with $N_{MCS} = 4 \cdot 10^4$	0	$2.19 \cdot 10^3$

Note. For the same discrepancy level, our method is about five and 10 times faster than MCS-FROST and MCS, respectively.

realizations for the quarter-five well configuration. For the same discrepancy level, CDF-FROST is about five and 10 times faster than MCS-FROST and MCS, respectively.

7. Conclusions

We developed a sampling-free CDF-FROST method for probabilistic forecast of immiscible two-phase flow in heterogeneous porous media with uncertain permeability. By employing a fixed-streamline assumption, the original FROST method of Ibrahima et al. (2015) outperforms MCS in computing saturation CDFs. However, it has a computational bottleneck, the MCS computation of the distribution of travel time and the statistics of the EIT. CDF-FROST overcomes this limitation by developing a deterministic equation for the CDF of travel time. The equation is obtained using 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, a 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 10 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.

Our CDF method is formally valid for mildly-to-moderately heterogeneous porous media, because it relies on perturbation expansions and closure approximations (Likanapaisal et al., 2012; Tchelepi & Li, 2004; Yang et al., 2019). One possible way to alleviate this limitation is to combine it with the random domain decomposition (RDD); Yang et al. (2020) demonstrated that doing so extends the applicability of the CDF method to highly heterogeneous media whose conductivity PDF might be non-stationary and multi-modal.

Another direction of future research is to incorporate the gravity and capillary-pressure effect into our method by deploying the operator splitting (Berenblyum et al., 2003) and CDF equation for advection-diffusion processes (Boso & Tartakovsky, 2016).

Appendix A: Derivation of Moment-Consistent Closures

We construct the closure variables α and β in a way that ensures that the CDF Equation 26 is consistent with the moments $\bar{\tau}$ and σ_τ^2 computed by moment equations or MCS. 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

$$\bar{\tau} = \tau_{\max} - \int_0^{\tau_{\max}} F_\tau(\mathcal{T}; r) \mathrm{d}\mathcal{T}, \quad \sigma_\tau^2 = \tau_{\max}^2 - 2\int_0^{\tau_{\max}} \mathcal{T} F_\tau(\mathcal{T}; r) \mathrm{d}\mathcal{T} - \bar{\tau}^2, \tag{A1}$$

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 Equation 26 over $\mathcal{T}$ and employing Equation A1 and Equation 29, we can derive the equation for $\bar{\tau}$

$$q_{\mathrm{tot}} \frac{\partial \bar{\tau}}{\partial r} - \beta(r) - \phi = 0. \tag{A2}$$

Similarly, multiplying both sides of Equation 26 by $\mathcal{T}$ and integrating the resulting equation over $\mathcal{T}$ yields

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

or, accounting for Equation A2,

$$q_{\text{tot}} \frac{\partial \sigma_\tau^2}{\partial r} - 2\alpha(\mathbf{x})\sigma_\tau^2 = 0. \tag{A3}$$

Imposition of the equivalency between the mean $(\bar{\tau})$ and variance (σ_τ^2) equations derived by the CDF method, Equation A2 and Equation A3, and the moments computed by SMEs or MCS yields expressions for the closure variables $\alpha(r)$ and $\beta(r)$ in Equation 30.

Acronyms

CDF	Cumulative density function
FROST	Frozen streamline method
MCS	Monte Carlo simulation
PDE	Partial differential equation
PDF	Probability distribution function
SME	Statistical moment equation
TOF	Time of flight

Data Availability Statement

There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

References

Aziz, K. (1979). *Petroleum reservoir simulation* (476 pp.). Applied Science Publishers.

Barajas-Solano, D. A., & Tartakovsky, D. M. (2016). Stochastic collocation methods for nonlinear parabolic equations with random coefficients. *SIAM/ASA Journal on Uncertainty Quantification*, *4*(1), 475–494. https://doi.org/10.1137/130930108

Berenblyum, R. A., Shapiro, A. A., Jessen, K., Stenby, E. H., & Orr, F. M. (2003). Black oil streamline simulator with capillary effects. In *SPE annual technical conference and exhibition*.

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

Boso, F., Marzadri, A., & Tartakovsky, D. M. (2018). Probabilistic forecasting of nitrogen dynamics in hyporheic zone. *Water Resources Research*, *54*(7), 4417–4431. https://doi.org/10.1029/2018WR022525

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

Boso, F., & Tartakovsky, D. M. (2020). Data-informed method of distributions for hyperbolic conservation laws. *SIAM Journal on Scientific Computing*, *42*(1), A559–A583. https://doi.org/10.1137/19M1260773

Coats, K. H. (2000). A note on IMPES and some IMPES-based simulation models. *SPE Journal*, *5*(03), 245–251. https://doi.org/10.2118/65092-pa

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

Datta-Gupta, A., & King, M. J. (2007). *Streamline simulation: Theory and practice* (Vol. 11). Society of Petroleum Engineers. https://doi.org/10.2118/9781555631116

Dentz, M., & Tartakovsky, D. M. (2010). Probability density functions for passive scalars dispersed in random velocity fields. *Geophysical Research Letters*, *37*(24), L24406. https://doi.org/10.1029/2010GL045748

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

Fuks, O., Ibrahima, F., Tomin, P., & Tchelepi, H. A. (2019). Analysis of travel time distributions for uncertainty propagation in channelized porous systems. *Transport in Porous Media*, *126*(1), 115–137. https://doi.org/10.1007/s11242-018-1052-z

Fuks, O., Ibrahima, F., Tomin, P., & Tchelepi, H. A. (2020). Uncertainty propagation for compositional flow using a probability distribution method. *Transport in Porous Media*, *132*(3), 113–133. https://doi.org/10.1007/s11242-020-01382-4

Hewett, T. A., & Yamada, T. (1997). Theory for the semi-analytical calculation of oil recovery and effective relative permeabilities using streamtubes. *Advances in Water Resources*, *20*(5–6), 279–292. https://doi.org/10.1016/s0309-1708(96)00049-8

Ibrahima, F., Meyer, D. W., & Tchelepi, H. A. (2015). Distribution functions of saturation for stochastic nonlinear two-phase flow. *Transport in Porous Media*, *109*(1), 81–107. https://doi.org/10.1007/s11242-015-0503-z

Ibrahima, F., Tchelepi, H. A., & Meyer, D. W. (2018). An efficient distribution method for nonlinear two-phase flow in highly heterogeneous multidimensional stochastic porous media. *Computational Geosciences*, *22*(1), 389–412. https://doi.org/10.1007/s10596-017-9698-0

Lax, P. D. (1973). *Hyperbolic systems of conservation laws and the mathematical theory of shock waves*. SIAM.

Likanapaisal, P., Li, L., & Tchelepi, H. A. (2012). Dynamic data integration and quantification of prediction uncertainty using statistical-moment equations. *SPE Journal*, *17*(1), 98–111. https://doi.org/10.2118/119138-pa

Müller, F., Jenny, P., & Meyer, D. W. (2012). 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*.

Pollock, D. W. (1988). Semianalytical computation of path lines for finite-difference models. *Groundwater*, *26*(6), 743–750. https://doi.org/10.1111/j.1745-6584.1988.tb00425.x

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, *40*(2), 99–121. https://doi.org/10.1023/a:1026543900054

Severino, G., & De Bartolo, S. (2015). Stochastic analysis of steady seepage underneath a water-retaining wall through highly anisotropic porous media. *Journal of Fluid Mechanics*, *778*, 253–272. https://doi.org/10.1017/jfm.2015.369

Shahvali, M., Mallison, B., Wei, K., & Gross, H. (2012). An alternative to streamlines for flow diagnostics on structured and unstructured grids. *SPE Journal*, *17*(03), 768–778. https://doi.org/10.2118/146446-pa

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

Tartakovsky, D. M., & Broyda, S. (2011). PDF equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of Contaminant Hydrology*, *120*, 129–140. https://doi.org/10.1016/j.jconhyd.2010.08.009

Tartakovsky, D. M., & Gremaud, P. A. (2016). Method of distributions for uncertainty quantification. In *Handbook of uncertainty quantification* (pp. 1–22). https://doi.org/10.1007/978-3-319-11259-6_27-1

Tartakovsky, D. M., Guadagnini, A., & Riva, M. (2003). Stochastic averaging of nonlinear flows in heterogeneous porous media. *Journal of Fluid Mechanics*, *492*, 47–62. https://doi.org/10.1017/S002211200300538X

Tartakovsky, D. M., Neuman, S. P., & Lu, Z. (1999). Conditional stochastic averaging of steady state unsaturated flow by means of Kirchhoff transformation. *Water Resources Research*, *35*(3), 731–745. https://doi.org/10.1029/1998wr900092

Tartakovsky, D. M., & Winter, C. L. (2008). Uncertain future of hydrogeology. *ASCE Journal of Hydrologic Engineering*, *13*(1), 37–39. https://doi.org/10.1061/(asce)1084-0699(2008)13:1(37)

Taverniers, S., Bosma, S. B. M., & Tartakovsky, D. M. (2020). Accelerated multilevel Monte Carlo with kernel-based smoothing and Latinized stratification. *Water Resources Research*, *56*(9), e2019WR026984. https://doi.org/10.1029/2019WR026984

Taverniers, S., & Tartakovsky, D. M. (2020). Estimation of distributions via multilevel Monte Carlo with stratified sampling. *Journal of Computational Physiscs*, *419*, 109572. https://doi.org/10.1016/j.jcp.2020.109572

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

Thiele, M. R. (2001). Streamline simulation. In *6th international forum on reservoir simulation* (pp. 3–7).

Venturi, D., Tartakovsky, D. M., Tartakovsky, A. M., & Karniadakis, G. E. (2013). Exact PDF equations and closure approximations for advective-reactive transport. *Journal of Computational Physics*, *243*, 323–343. https://doi.org/10.1016/j.jcp.2013.03.001

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

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

Yang, H.-J., Boso, F., Tchelepi, H. A., & Tartakovsky, D. M. (2020). Method of distributions for quantification of geologic uncertainty in flow simulations. *Water Resources Research*, *56*(7), e2020WR027643. https://doi.org/10.1029/2020WR027643

References From the Supporting Information

Dagan, G. (1982). Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 1. Conditional simulation and the direct problem. *Water Resources Research*, *18*(4), 813–833. https://doi.org/10.1029/wr018i004p00813

Fowler, K. R., & Kelley, C. T. (2005). Pseudo-transient continuation for nonsmooth nonlinear equations. *SIAM Journal on Numerical Analysis*, *43*(4), 1385–1406. https://doi.org/10.1137/s0036142903431298

Gao, B.-Q., & Gandhi, O. P. (1992). An expanding-grid algorithm for the finite-difference time-domain method (EM deposition in human body). *IEEE Transactions on Electromagnetic Compatibility*, *34*(3), 277–283. https://doi.org/10.1109/15.155842

Gelhar, L. W., & Collins, M. A. (1971). General analysis of longitudinal dispersion in nonuniform flow. *Water Resources Research*, *7*(6), 1511–1521. https://doi.org/10.1029/wr007i006p01511

Kelley, C. T., & Keyes, D. E. (1998). Convergence analysis of pseudo-transient continuation. *SIAM Journal on Numerical Analysis*, *35*(2), 508–523. https://doi.org/10.1137/s0036142996304796

Li, L., Tchelepi, H. A., & Zhang, D. (2003). Perturbation-based moment equation approach for flow in heterogeneous porous media: Applicability range and analysis of high-order terms. *Journal of Computational Physics*, *108*(1), 296–317. https://doi.org/10.1016/s0021-9991(03)00186-4

Neuman, S. P., Tartakovsky, D. M., Wallstrom, T. C., & Winter, C. L. (1996). Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resources Research*, *32*(5), 1479–1480. https://doi.org/10.1029/96wr00489

Roe, P. L. (1981). Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, *43*(2), 357–372. https://doi.org/10.1016/0021-9991(81)90128-5

Tartakovsky, D. M., & Neuman, S. P. (1998a). Transient flow in bounded randomly heterogeneous domains 1. Exact conditional moment equations and recursive approximations. *Water Resources Research*, *34*(1), 1–12. https://doi.org/10.1029/97wr02118

Tartakovsky, D. M., & Neuman, S. P. (1998b). Transient flow in bounded randomly heterogeneous domains 2. Localization of conditional mean equations and temporal nonlocality effects. *Water Resources Research*, *34*(1), 13–20. https://doi.org/10.1029/97wr02117

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

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

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

Contents of this file

S1. Derivation of moment equations

S1.1 Velocity equations

Derivation and analysis of the statistical moment equations (SMEs) for steady-state flow equation (3) and stochastic average macroscopic velocity $\mathbf{v}(\mathbf{x})$ have been a subject of intensive research in stochastic hydrogeology for several decades (e.g., Gelhar & Collins, 1971; Dagan, 1982; Neuman et al., 1996; Dagan & Neuman, 1997; Tartakovsky & Neuman, 1998a,b; Li et al., 2003; Severino & De Bartolo, 2015, among many others). Detailed derivations of SMEs for (3) can be found in Yang et al. (2019). Here, we derive the first- and second-moment equations for the velocity $\mathbf{v}(\mathbf{x})$. For flow in porous media with constant porosity ϕ, the latter is defined by Darcy's law,

$$\mathbf{v} = -\frac{\mathrm{e}^{Y}}{\phi}\nabla h, \tag{S1}$$

----

where $Y(\mathbf{x}) = \ln K$ is the random log-hydraulic conductivity and h is the hydraulic head. 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 $K_G = \exp(\bar{Y})$ the geometric mean of the permeability K, expanding $\exp(Y')$ into a Taylor series around $Y' = 0$, and taking the ensemble mean of the resulting equation leads to

$$\bar{\mathbf{v}} = -\frac{K_G}{\phi} \sum_{n=0}^{\infty} \frac{1}{2n!} (\nabla \bar{h}^{(0)} \sigma_Y^{2n} + \langle Y'^n \nabla h' \rangle). \tag{S2}$$

The first-order approximation of mean velocity, $\tilde{\mathbf{v}}(\mathbf{x})$, is

$$\tilde{\mathbf{v}} = -\frac{K_G}{\phi} [\nabla \bar{h}^{(0)} (1 + \frac{\sigma_Y^2}{2}) + \nabla C_{Yh}(\mathbf{x}, \mathbf{x}))], \tag{S3}$$

where $C_{Yh}(\mathbf{x}, \mathbf{x})$, the covariance between the log-conductivity $Y(\mathbf{x})$ and the hydraulic head $h(\mathbf{x})$, is computed by solving the corresponding SME. An equation for the covariance between the i-th and j-th components of the velocity vector, $C_{v_i v_j}(\mathbf{x}, \boldsymbol{\chi})$, is obtained by subtracting (S2) with (S3), multiplying $v'_j(\mathbf{x})$, and taking ensemble mean,

$$C_{v_i v_j} = \frac{K_G^2}{\phi^2} \left[\frac{\partial \bar{h}}{\partial x_i} \frac{\partial \bar{h}}{\partial \chi_j} C_Y(\mathbf{x}, \boldsymbol{\chi}) + \frac{\partial^2 C_h}{\partial x_i \partial \chi_j} + \frac{\partial \bar{h}}{\partial x_i} \frac{\partial C_{Yh}(\mathbf{x}, \boldsymbol{\chi})}{\partial \chi_j} + \frac{\partial \bar{h}}{\partial \chi_j} \frac{\partial C_{Yh}(\boldsymbol{\chi}, \mathbf{x})}{\partial x_i} \right], \tag{S4}$$

where the auto-covariance of hydraulic head at two points x and χ, $C_h(\mathbf{x}, \boldsymbol{\chi})$, is a solution of the SMEs.

S1.2 TOF equations

The first two moments of the velocity field obtained by (S3) and (S4) are used to generate a mean streamline. Then, the travel time moments are computed using the perturbation expansion around the mean streamline. TOF along the mean streamline coordinate, τ, is

defined as

$$\tau = \int_0^r \frac{d\chi}{v_r(\chi, \eta(\chi))} \tag{S5}$$

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 used to parameterize the mean streamline. The transverse displacement $\eta(\chi)$, i.e., the displacement perpendicular to the mean streamline, is defined as

$$\eta(r) = \int_0^r \frac{v_\eta(\chi)}{v_r(\chi)} d\chi. \tag{S6}$$

Here, v_η is the velocity component perpendicular to the mean streamline. Considering the definition of mean streamline, the mean traverse displacement $\langle \eta \rangle = 0$. Randomness of traverse displacement η and velocity v_r in (S5) renders TOF τ random as well.

The underlying assumption of our perturbation expansion is that variances of the traverse displacement $\eta(\chi)$ and the velocity along the streamline v_r are relatively small. Taylor expansions of these two quantities are employed to perturb all streamline-related random quantities around the mean streamline. Applying a Taylor expansion to (S5) and neglecting the terms higher than first-order, we derive a truncated equation for TOF τ,

$$\tau(r) = \int_0^r \left[\frac{1}{\langle v_r(\chi, \langle \eta \rangle) \rangle} - \frac{v_r'(\chi, \langle \eta \rangle)}{\langle v_r(\chi, \langle \eta \rangle) \rangle^2} + \frac{\eta'(\chi)}{\langle v_r(\chi, \langle \eta \rangle) \rangle^2} \frac{\partial \langle v_r(\chi, \eta) \rangle}{\partial \eta(\chi)} \bigg|_{\eta=\langle \eta \rangle} \right] d\chi \tag{S7}$$

Taking ensemble average to the (S7) yields the equation of the zeroth-order mean travel-time $\bar{\tau}(r)^{(0)}$,

$$\bar{\tau}(r)^{(0)} = \int_0^r \frac{d\chi}{\langle v_r(\chi, \langle \eta \rangle) \rangle}. \tag{S8}$$

We subtract the mean equation (S8) from (S7) to obtain the first-order fluctuation term,

$$\tau'(r) = \int_0^r \left[v_r'(\chi, \langle \eta \rangle) + \eta'(\chi) \frac{\partial \langle v_r(\chi, \eta) \rangle}{\partial \eta(\chi)} \bigg|_{\eta=\langle \eta \rangle} \right] \frac{d\chi}{\langle v_r(\chi, \langle \eta \rangle) \rangle^2}. \tag{S9}$$

October 12, 2022, 10:37am

Multiplying TOF fluctuations at two different locations and taking ensemble average lead

to the expression for TOF covariance

$$
\begin{aligned}
C_\tau(r_1, r_2) = \int_0^{r_1} \int_0^{r_2} & \left[\langle v'_{r1} v'_{r2} \rangle + \frac{\partial \langle v_{r1} \rangle}{\partial \eta} \bigg|_{\eta_1 = \langle \eta_1 \rangle} \langle v'_{r2} \eta'_1 \rangle + \frac{\partial \langle v_{r2} \rangle}{\partial \eta} \bigg|_{\eta_2 = \langle \eta_2 \rangle} \langle v'_{r1} \eta'_2 \rangle \right. \\
& \left. + \frac{\partial \langle v_{r1} \rangle}{\partial \eta} \bigg|_{\eta_1 = \langle \eta_1 \rangle} \frac{\partial \langle v_{r2} \rangle}{\partial \eta} \bigg|_{\eta_2 = \langle \eta_2 \rangle} \langle \eta'_1 \eta'_2 \rangle \right] \frac{\mathrm{d}\chi_1 \mathrm{d}\chi_2}{\langle v_{r1} \rangle^2 \langle v_{r2} \rangle^2},
\end{aligned}
\tag{S10}
$$

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 $\langle \eta'_1 \eta'_2 \rangle$, are obtained from (S6). By

expanding v_η and $1/v_r$ with a Taylor series expansion around $\langle v_r \rangle$, we write the transverse

displacement fluctuation as

$$
\eta'(r) = \int_0^r \frac{v'_\eta(\chi)}{\langle v_r(\chi) \rangle} \left[1 - \frac{v'_r(\chi)}{\langle v_r(\chi) \rangle} + \frac{v'^2_r(\chi)}{\langle v_r(\chi) \rangle^2} - \cdots \right] \mathrm{d}\chi.
\tag{S11}
$$

With this fluctuation term, we express the transverse displacement covariance as the

expected value of two traverse displacements with first order accuracy,

$$
\langle \eta'_1(r_1) \eta'_2(r_2) \rangle = \int_0^{r_2} \int_0^{r_1} \frac{\langle v'_{\eta,1}(\chi_1) v'_{\eta,2}(\chi_2) \rangle}{\langle v_{r,1} \rangle (\chi_1) \langle v_{r,2} \rangle (\chi_2)} \mathrm{d}\chi_1 \mathrm{d}\chi_2.
\tag{S12}
$$

Similarly, the following first-order equation for covariance between transverse displacement

and velocity along a streamline are derived by multiplying (S11) with v'_r and taking

ensemble average:

$$
\langle v'_r(r_1, \langle \eta_1 \rangle) \eta'(r_2) \rangle = \int_0^{r_2} \frac{\langle v'_\eta(r_1, \langle \eta(r_1) \rangle) v'_\eta(\chi) \rangle}{\langle v_r(\chi) \rangle} \mathrm{d}\chi.
\tag{S13}
$$

In summary, the TOF moments are computed by integrating the velocity moments

derived in Section S1.1. The numerical integration generally requires less computational

time than a linear solver. Hence, the TOF moments are obtained readily from the results

of Section S1.1.

S2. Numerical methods for CDF equation

S2.1 Pseudo-time stepping

A pseudo-time stepping or pseudo-transient continuation method is particularly appropriate for the steady-state equation with non-smooth coefficients (Kelley & Keyes, 1998; Fowler & Kelley, 2005). It starts with introducing the fictitious or pseudo-time t'. Dropping the subscript τ for simplicity, (27) is transformed into the following time dependent equation

$$\frac{\partial F}{\partial t'} + \frac{\partial (q^{\text{sl}} F)}{\partial r} + \frac{\partial (U F)}{\partial \mathcal{T}} = vF, \quad F(t' = 0) = F_0, \tag{S14}$$

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 (27) regardless of the initial condition.

When (S14) 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

$$F_{i,j}^{n+1} = (1+v)F_{i,j}^n - \lambda_{r,i}(f_{i+1/2,j}^n - f_{i-1/2,j}^n) - \lambda_{T,j}(g_{i,j+1/2}^n - g_{i,j-1/2}^n). \tag{S15}$$

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 $r_{i\pm1/2} = r_i \pm \Delta r_i$ and $T_{j\pm1/2} = T_j \pm \Delta T_j$. f and g are respectively the r and T direction components of numerical flux.

0.1. Flux-limited scheme

Among a plethora of schemes for approximating numerical flux, Roe's first-order up-wind method (Roe, 1981) guarantees non-oscillatory solution near discontinuities. Since $q_{\text{tot}}^{\text{sl}}$ is always greater than 0 along the streamline, flux term f can be expressed as $f_{i+1/2,j}^{n} = q_{i,j}^{\text{sl}} F_{i,j}^{n}$. For the flux term g, the first-order upwind method fails to impose the T space boundary conditions (??) 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 com-bination of low-order (i.e., Roe first-order upwind) and high-order (i.e., Lax-Wendroff) methods,

$$g_{i,j+1/2}^{n} = \begin{cases} U_{i,j} F_{i,j}^{n} + \frac{1}{2} a_{i,j+1/2}(1 - \lambda_T a_{i,j+1/2}) \psi_{i,j}^{+}(F_{i,j+1}^{n} - F_{i,j}^{n}), & a_{i,j+1/2} > 0 \\ U_{i,j+1} F_{i,j+1}^{n} + \frac{1}{2} a_{i,j+1/2}(1 + \lambda_T a_{i,j+1/2}) \psi_{i,j+1}^{-}(F_{i,j+1}^{n} - F_{i,j}^{n}), & a_{i,j+1/2} < 0 \end{cases}$$
(S16)

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

$$a_{i,j+1/2}^{n} = \begin{cases} \frac{U_{i,j+1} F_{i,j+1}^{n} - U_{i,j} F_{i,j}^{n}}{F_{i,j+1}^{n} - F_{i,j}^{n}} & F_{i,j} \neq F_{i,j+1} \\ U_{i,j} & F_{i,j+1} = F_{i,j} \end{cases} .$$
(S17)

Here, flux-limiter $\psi^{\pm}$ is the function of $\theta^{\pm}$ which is the ratio of successive gradients indicating the smoothness of the solution. The minmod flux-limiter function $\psi(\theta) = \max[0, \min(1, \theta)]$ is employed in the present study. The ratio $\theta_{i,j}^{\pm}$ is a measure of smooth-ness of the data near cell (i, j); it is defined as

$$\theta_{i,j}^{+} = \frac{F_{i,j} - F_{i,j-1}}{F_{i,j+1} - F_{i,j}}, \quad \theta_{i,j}^{-} = \frac{F_{i,j+1} - F_{i,j}}{F_{i,j} - F_{i,j-1}}.$$
(S18)

The numerical solution $F_{i,j}^n$ is updated iteratively using (S15) until it converges to the steady-state solution. The convergence criteria is

$$\max_{i,j} |F_{i,j}^{n+1} - F_{i,j}^n| < 10^{-8}. \tag{S19}$$

S2.2 Flux-limited scheme

Among a plethora of schemes for approximating numerical flux, Roe's first-order up-wind method (Roe, 1981) guarantees non-oscillatory solution near discontinuities. Since $q_{\text{tot}}^{\text{sl}}$ is always greater than 0 along the streamline, flux term f can be expressed as $f_{i+1/2,j}^n = q_{i,j}^{\text{sl}} F_{i,j}^n$. For the flux term g, the first-order upwind method fails to impose the T space boundary conditions (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,

$$g_{i,j+1/2}^n = \begin{cases} U_{i,j} F_{i,j}^n + \frac{1}{2} a_{i,j+1/2} (1 - \lambda_T a_{i,j+1/2}) \psi_{i,j}^+ (F_{i,j+1}^n - F_{i,j}^n), & a_{i,j+1/2} > 0 \\ U_{i,j+1} F_{i,j+1}^n + \frac{1}{2} a_{i,j+1/2} (1 + \lambda_T a_{i,j+1/2}) \psi_{i,j+1}^- (F_{i,j+1}^n - F_{i,j}^n), & a_{i,j+1/2} < 0 \end{cases} \tag{S20}$$

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

$$a_{i,j+1/2}^n = \begin{cases} \frac{U_{i,j+1} F_{i,j+1}^n - U_{i,j} F_{i,j}^n}{F_{i,j+1}^n - F_{i,j}^n} & F_{i,j} \neq F_{i,j+1} \\ U_{i,j} & F_{i,j+1} = F_{i,j} \end{cases}. \tag{S21}$$

Here, flux-limiter $\psi^\pm$ is the function of $\theta^\pm$ which is the ratio of successive gradients indicating the smoothness of the solution. The minmod flux-limiter function $\psi(\theta) = \max[0, \min(1, \theta)]$ is employed in the present study. The ratio $\theta_{i,j}^\pm$ is a measure of smooth-

ness of the data near cell (i, j); it is defined as

$$\theta_{i,j}^+ = \frac{F_{i,j} - F_{i,j-1}}{F_{i,j+1} - F_{i,j}}, \quad \theta_{i,j}^- = \frac{F_{i,j+1} - F_{i,j}}{F_{i,j} - F_{i,j-1}}. \tag{S22}$$

The numerical solution $F_{i,j}^n$ is updated iteratively using (S15) until it converges to the steady-state solution. The convergence criteria is

$$\max_{i,j} |F_{i,j}^{n+1} - F_{i,j}^n| < 10^{-8}. \tag{S23}$$

S2.3 Exponential grid spacing

Considering the original definition of TOF (9), $\tau(r)$ increases monotonically along the streamline. Defining the CDF equation (27) along the mean streamline coordinate, every point along the streamline should share the same T-direction grid system. Figure S1 demonstrates that a small grid size ΔT_j is required to accurately describe the CDF when the distance from the origin r is small. A coarser grid could be enough to represent the travel-time CDF at large r. We use this observation to further speed-up the computations by introducing an exponential grid spacing.

In the exponential grid spacing system, the size $\Delta T_j = \Delta T_0 \xi^{j-1}$ of the jth cell increases with a fixed expansion rate ξ, where ΔT_0 is the initial time. The location of the jth cell is $T_j = T_0 + \Delta T_0 (1 - \xi^{j-1})/(1 - \xi)$, where T_0 is the location of first cell $j = 0$. In our numerical experiments, we set $\xi = 1.02$ based on the previous study (Gao & Gandhi, 1992). Figure S2 shows the numerical solutions of (27) obtained by two different grid systems with same grid number ($N_T = 150$). The exponential grid spacing provides the same level of accuracy as the uniform grid at a fraction of the computational cost of the latter.

References

Dagan, G. (1982). Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 1. Conditional simulation and the direct problem. *Water Resources Research*, *18*(4), 813–833.

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

Fowler, K. R., & Kelley, C. T. (2005). Pseudo-transient continuation for nonsmooth nonlinear equations. *SIAM Journal on Numerical Analysis*, *43*(4), 1385–1406.

Gao, B.-Q., & Gandhi, O. P. (1992). An expanding-grid algorithm for the finite-difference time-domain method (EM deposition in human body). *IEEE Transactions on Electromagnetic Compatibility*, *34*(3), 277–283.

Gelhar, L. W., & Collins, M. A. (1971). General analysis of longitudinal dispersion in nonuniform flow. *Water Resources Research*, *7*(6), 1511–1521.

Kelley, C. T., & Keyes, D. E. (1998). Convergence analysis of pseudo-transient continuation. *SIAM Journal on Numerical Analysis*, *35*(2), 508–523.

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

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

Roe, P. L. (1981). Approximate Riemann solvers, parameter vectors, and difference

schemes. *Journal of Computational Physics*, *43*(2), 357–372.

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

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

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

Yang, H.-J., Boso, F., Tchelepi, H. A., & Tartakovsky, D. M. (2019). Probabilistic forecast of single-phase flow in porous media with uncertain properties. *Water Resources Research*, *55*(11), 8631-8645. doi: 10.1029/2019WR026090
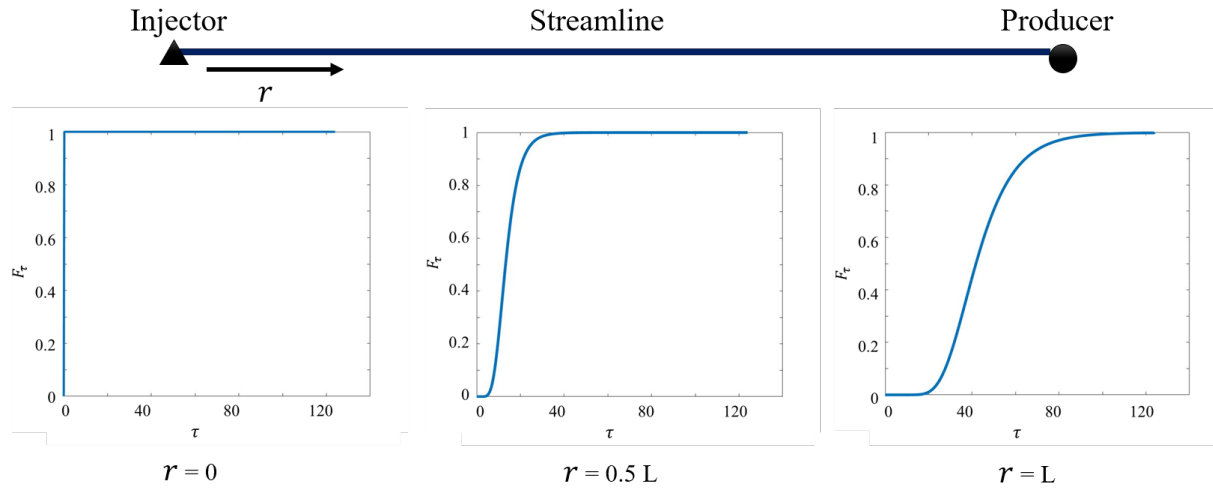
Figure S1.   Evolution of the travel-time CDF with distance r from the starting point

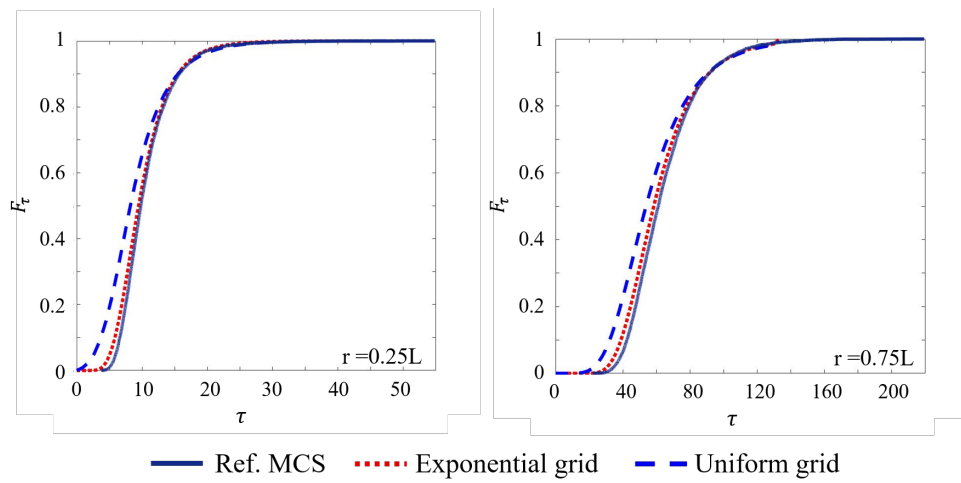of mean streamline. L is the total length of the mean streamline.

Figure S2.   Travel time CDFs F_τ computed with reference MCS and the CDF method with the uniform and exponential grid systems at two selected distance r along the mean streamline for quarter-five spot well configuration. L is the total length of the mean streamline.