AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2016WR018745

The method of distributions for dispersive transport in porous media with uncertain hydraulic properties

Francesca Boso[1] and Daniel M. Tartakovsky[1]

[1]Department of Mechanical and Aerospace Engineering, University of California, San Diego, California, USA

Abstract Predictions of solute transport in subsurface environments are notoriously unreliable due to aquifer heterogeneity and uncertainty about the values of hydraulic parameters. Probabilistic framework, which treats the relevant parameters and solute concentrations as random fields, allows for quantification of this predictive uncertainty. By providing deterministic equations for either probability density function or cumulative distribution function (CDF) of predicted concentrations, the method of distributions enables one to estimate, e.g., the probability of a contaminant's concentration exceeding a safe dose. We derive a deterministic equation for the CDF of solute concentration, which accounts for uncertainty in flow velocity and initial conditions. The coefficients in this equation are expressed in terms of the mean and variance of concentration. The accuracy and robustness of the CDF equations are analyzed by comparing their predictions with those obtained with Monte Carlo simulations and an assumed beta CDF.

Key Points:
- A unique uncertainty quantification tool
- The method yields full PDF, not moments
- The method outperforms MCS

Supporting Information:
- Supporting Information S1

Correspondence to:
D. M. Tartakovsky,
dmt@ucsd.edu

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

Received 6 FEB 2016
Accepted 21 MAY 2016
Accepted article online 26 MAY 2016
Published online 19 JUN 2016

1. Introduction

Advection-dispersion equations (ADEs) are routinely used by hydrogeologists to model solute transport in the subsurface. Just as ubiquitous as their applications is uncertainty about the values of the coefficients that parameterize ADEs, e.g., flow velocity, dispersivity, and initial and boundary conditions. This parametric uncertainty, which arises from subsurface heterogeneity and limited amounts of hydrologic data, renders ADE-based predictions of solute concentrations fundamentally uncertain. (Another source of uncertainty—the veracity of ADEs stemming from their purported failure to capture non-Fickian transport in heterogeneous formations [*Cushman and O'Malley*, 2015; *Neuman and Tartakovsky*, 2009; *Berkowitz et al.*, 2006]—lies outside the scope of the present study.)

This pervasive uncertainty raises practical questions of how to act upon such predictions and how to communicate them to nonspecialists. Faced with similar questions, the climate change community has adopted the language in Table 1, which we propose to use for groundwater modeling. This language implies the necessity of relying on the probabilistic framework when dealing with uncertainty, such that a prediction "solute concentration $C(\mathbf{x}, t)$ at the location $\mathbf{x}$ and time t is *likely* to stay below a safe limit c" is equivalent to estimating the probability $\mathbb{P}[C(\mathbf{x}, t) \leq c] > 0.66$. In other words, rather than computing a single prediction of concentration one has to find its *cumulative distribution function* (CDF), $F_C(c; \mathbf{x}, t) \equiv \mathbb{P}[C(\mathbf{x}, t) \leq c]$.

This goal is achieved by treating input parameters in an ADE as random fields and solving the stochastic version of this equation, i.e., by propagating the parametric uncertainty through the modeling process. Monte Carlo simulations (MCS) are often used for this purpose, but they are computationally expensive, and often prohibitively so. This is because an accurate estimation of the CDF F_C requires a large number of solves of the ADE (realizations) in order to obtain a representative histogram of the solution. Much of research in stochastic hydrogeology is driven by the need to develop computationally efficient alternatives to MCS.

Moment differential equations (MDEs) provide one such alternative by deriving deterministic equations governing the dynamics of statistical moments of solute concentration. Most, if not all, MDEs are limited to the first two statistical moments, i.e., concentration mean $\langle C(\mathbf{x}, t)\rangle \equiv \int c \, \mathrm{d}F_C(c; \mathbf{x}, t)$ and variance $\sigma_C^2(\mathbf{x}, t) \equiv \int c^2 \, \mathrm{d}F_C(c; \mathbf{x}, t) - \langle C(\mathbf{x}, t)\rangle^2$ [*Winter et al.*, 1984; *Dagan*, 1989; *Cushman*, 1997; *Dagan and Neuman*, 1997; *Kapoor and Kitanidis*, 1998]. (A reason is that while derivation of MDEs for higher statistical moments is relatively straightforward, solving them numerically is computationally more demanding than MCS.) They provide only a limited probabilistic information, with $\langle C(\mathbf{x}, t)\rangle$ serving as the "best" prediction and $\sigma_C^2(\mathbf{x}, t)$ providing a measure of predictive uncertainty. They cannot be used to determine whether a particular

Table 1. A Common Language Description of Probabilistic Occurrence of an Event, Suggested by the 2007 Intergovernmental Panel on Climate Change

Description	Probability of Occurrence
Virtually certain	>99%
Extremely likely	>95%
Very likely	>90%
Likely	>66%
More likely than not	>50%
Unlikely	<33%
Very unlikely	<10%
Extremely unlikely	<5%

prediction is "likely" or "almost certain" (see Table 1) in all but a few limited scenarios, e.g., when $C(\mathbf{x},t)$ can be approximated as a multivariate Gaussian random field. Yet such information is what is needed for probabilistic risk assessment of subsurface contamination [e.g., *Tartakovsky*, 2013], which typically requires estimates of probability of occurrence of rare (e.g., "extremely unlikely") events.

The method of distributions [*Tartakovsky and Gremaud*, 2016] overcomes this limitation by deriving deterministic equations for the concentration's CDF F_C (CDF equations) or PDF $f_C \equiv \mathrm{d}F_C/\mathrm{d}c$ (probability density function or PDF) equations, rather than for its moments. PDF equations originated in the statistical theory of turbulence [*Lundgren*, 1967] and have since been modified to deal with uncertain flow velocity and/or reaction rate constants in advection-reaction transport problems [*Lichtner and Tartakovsky*, 2003; *Shvidler and Karasaki*, 2003; *Tartakovsky et al.*, 2009; *Venturi et al.*, 2013]. Like their PDF counterparts, CDF equations treat nonlinear reaction terms in the transport equations exactly, without resorting to linearization required by MDEs [*Boso et al.*, 2014; *Tartakovsky and Gremaud*, 2016]. Other advantages of CDF equations include the uniqueness of boundary conditions in the event space, e.g., $F_C(c=0;\mathbf{x},t)=0$ for any $\mathbf{x}$ and t, and the numerical efficiency of having to find a solution that is a smooth function in c, which increases from 0 and 1.

The presence of a Laplace operator, i.e., diffusion or dispersion terms, in the ADE complicates the derivation of PDF/CDF equations by requiring a closure approximation. Examples of such closures include a Gaussianity assumption for the dynamics of a plume's center of mass [*Dentz and Tartakovsky*, 2010; *Dentz*, 2012], perturbation expansions [*de Barros and Fiori*, 2014], and the interaction by exchange with the mean (IEM) closure [*Pope*, 2000]. The original IEM formulation fails to capture the correct space-time behavior of both the mean and the variance of concentration unless dispersion is negligible or these statistics are constant in space. Subsequent IEM modifications [e.g., *Raman et al.*, 2005] capture the mean behavior, but fail to match the concentration variance.

We develop a new closure for the method of distributions that, unlike the original IEM procedure and its subsequent generalizations [*Haworth*, 2010], captures the correct behavior of both the mean and the variance of solute concentration. Section 2 contains the derivation of a general CDF equation for the ADE with uncertain (random) flow velocity. This equation is solved for transport in a stratified flow in section 3 to obtain a complete probabilistic description of solute concentration at any space-time point $(\mathbf{x},t)$; the latter is used to compute maps of probability of exceedance of a safe concentration limit c^*, $\mathbb{P}[C(\mathbf{x},t)>c^*]=1-F_C(c^*;\mathbf{x},t)$. Section 4 provides a comparison of the CDFs $F_C(c;\mathbf{x},t)$ computed with our method and MCS, as well as with an assumed β-CDF $F_C^{\beta}(c;\mathbf{x},t)$ that is often used as a surrogate of the actual CDF $F_C(c;\mathbf{x},t)$ [*Fiori*, 2001; *Bellin and Tonina*, 2007; *Bellin et al.*, 2011]. Major conclusions drawn from our study are summarized in section 5.

2. CDF Equation

Following *Dagan* [1987], *Cushman et al.* [2002], *Morales-Casique et al.* [2006], *Neuman and Tartakovsky* [2009] and many others, we postulate the existence of a spatial scale ω on which an ADE

$$\frac{\partial C}{\partial t}=\nabla\cdot(\mathbf{D}\nabla C)-\nabla\cdot(\mathbf{v}C) \tag{1}$$

is valid for any "point" $\mathbf{x}\in\Omega$ inside a computational domain Ω and any time $t>0$. At this scale, the second-order tensor $\mathbf{D}$ is a local diffusion/dispersion coefficient, which is independent of the uncertain divergence-free flow velocity $\mathbf{v}(\mathbf{x},t)$ that is treated as a random vector field with prescribed (or obtained from flow simulations) statistics. The ADE is supplemented by initial and Dirichlet boundary conditions

$$C(\mathbf{x},0)=C_0(\mathbf{x}),\quad \mathbf{x}\in\Omega;\qquad C(\mathbf{x},t)=\phi(\mathbf{x},t),\quad \mathbf{x}\in\partial\Omega, \tag{2}$$

where $\partial\Omega$ is the boundary of the simulation domain Ω, and $C_0(\mathbf{x})$ and $\phi(\mathbf{x},t)$ are uncertain auxiliary functions that are treated as mutually independent random fields with given distributions. Randomness of $\mathbf{v}$, C_0,

and ϕ renders a solution of this transport problem, $C(\mathbf{x},t)$, random as well. Instead of developing deterministic equations for the concentration moments $\langle C(\mathbf{x},t)\rangle$ and $\sigma_C^2(\mathbf{x},t)$, we derive a deterministic equation that governs the space-time evolution of the full CDF $F_C(c;\mathbf{x},t)$.

We introduce a random function $\Pi(c,C)=\mathcal{H}[c-C(\mathbf{x},t)]$, where $\mathcal{H}[\cdot]$ is the Heaviside function. The ensemble average of $\Pi(c,C)$ over all possible values of C at the space-time point $(\mathbf{x},t)$ yields $F_C(c;\mathbf{x},t)$. Indeed, if $f_C(c;\mathbf{x},t)$ is the PDF of C at $(\mathbf{x},t)$, then

$$\langle\Pi(c,C)\rangle=\int_{C_{\min}}^{C_{\max}}\mathcal{H}(c-\mathcal{C})f_C(\mathcal{C};\mathbf{x},t)\mathrm{d}\mathcal{C}=\int_{C_{\min}}^{c}f_C(\mathcal{C};\mathbf{x},t)\mathrm{d}\mathcal{C}=F_C(c;\mathbf{x},t). \tag{3}$$

This property suggests a two-step procedure for deriving a CDF equation. First, derive a stochastic equation for $\Pi(c,C)$ from (1). Second, average the result to obtain an equation for $F_C(c;\mathbf{x},t)$. After a series of approximations, this procedure leads to the CDF equation (see Appendix A for details)

$$\frac{\partial F_C}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla F_C-\left(\frac{\chi}{2}+\frac{\mathbf{D}\nabla\langle C\rangle\cdot\nabla\langle C\rangle}{\sigma_C^2}\right)(c-\langle C\rangle)\frac{\partial F_C}{\partial c}=\nabla\cdot(\mathbf{D}_{\mathrm{eff}}\nabla F_C), \tag{4}$$

where χ is a closure parameter in the equation for the concentration variance σ_C^2 (see equation (B10)),

$$\mathbf{D}_{\mathrm{eff}}=\mathbf{D}+\mathcal{D}_m,\qquad \mathcal{D}_m(\mathbf{x},t)=\int_0^t\int_\Omega\mathbf{C}_{\mathbf{v}}(\mathbf{x},t;\boldsymbol{\xi},\tau)G(\mathbf{x},t;\boldsymbol{\xi},\tau)\mathrm{d}\boldsymbol{\xi}\mathrm{d}\tau, \tag{5}$$

$G(\mathbf{x},t;\boldsymbol{\xi},\tau)$ is the mean-field Green's function for the transport problem (1) and (2), defined in Appendix A, and $\mathbf{C}_{\mathbf{v}}(\mathbf{x},t;\boldsymbol{\xi},\tau)$ is a given space-time auto-covariance matrix of velocity $\mathbf{v}$.

The CDF equation (4) is subject to the initial condition

$$F_C(c;\mathbf{x},0)=F_{C_0}(c;\mathbf{x}),\quad C_{\min}<c<C_{\max},\quad \mathbf{x}\in\Omega, \tag{6a}$$

the boundary conditions along the computational domain's boundary

$$F_C(c;\mathbf{x},t)=F_\phi(c;\mathbf{x},t),\quad C_{\min}<c<C_{\max},\quad \mathbf{x}\in\partial\Omega,\quad t>0, \tag{6b}$$

and the boundary conditions for the limiting values of the concentration's variation

$$F_C(c=C_{\min};\mathbf{x},t)=0,\quad F_C(c=C_{\max};\mathbf{x},t)=1,\quad \mathbf{x}\in\Omega,\quad t>0, \tag{6c}$$

which follow from the very definition of the CDF. Here F_{C_0} and F_ϕ are the CDFs of the initial and boundary functions C_0 and ϕ, respectively. If any of these inputs is deterministic, then its CDF is the Heaviside function centered on the prescribed deterministic value.

The CDF equation (4) ensures its solution F_C has the same mean $\langle C\rangle$ and variance σ_C^2 as those predicted with the MDEs, and share the same closure parameters: the variance dissipation rate χ and the macrodispersion tensor $\mathcal{D}_m$. It captures both the propagation and dissipation effects of dispersion on uncertainty through an explicit dependence on $\mathbf{D}$ and χ, respectively. The classic IEM approach [*Villermaux and Falk*, 1994],

$$\frac{\partial F_C}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla F_C-\frac{\chi}{2}(c-\langle C\rangle)\frac{\partial F_C}{\partial c}=\nabla\cdot(\mathcal{D}_m\nabla F_C), \tag{7}$$

accounts for uncertainty dissipation and dispersion only through χ; in the turbulence literature, χ is a calibration parameter that is usually set to be inversely proportional to the timescale of turbulence [*Pope*, 2000]. This is only adequate when dispersion has a negligible role (high Reynolds number flows) or when the concentration is statistically homogeneous (has constant mean and variance). Subsequent modifications of (7) introduced dependencies on $\nabla^2\langle C\rangle$ [*Raman et al.*, 2005] and the velocity [*Haworth*, 2010, and the references therein], but did not solve the issue of moment reproduction in dispersion-dominated conditions.

3. Computational Example

We consider solute transport in a two-dimensional domain of infinite extent, $\Omega=\{(x,z):-\infty<x<\infty,-\infty<z<\infty\}$, which takes place in a stratified flow, $\mathbf{v}=(v,0)^\top$ with $v=v(z,t)$, induced by a

mean uniform pressure gradient applied in the horizontal (x) direction. Assuming that transverse dispersion, which contributes to dilution in the vertical (z) direction, is negligible, the transport problem (1)–(2) reduces to

$$\frac{\partial C}{\partial t}+v(z,t)\frac{\partial C}{\partial x}=D\frac{\partial^2 C}{\partial x^2},\qquad C(x,z,t=0)=C_0(x,z). \tag{8}$$

Both the flow velocity $v(z, t)$ and the initial condition $C_0(x,z)$ are uncertain and treated as random fields. To be specific, we consider the initial condition in the form of a localized source with maximum concentration $C_{\max}$,

$$C_0(x)=C_{\max}\exp\left[-\frac{(x-x_0)^2}{2I^2}\right], \tag{9}$$

that is centered around an uncertain location x_0, and whose width is controlled by the parameter I. Rewriting (8) in terms of dimensionless quantities

$$\tilde{x}=\frac{x}{I},\quad \tilde{t}=\frac{t\langle v\rangle}{I},\quad \tilde{v}=\frac{v}{\langle v\rangle},\quad \tilde{C}=\frac{C}{C_{\max}},\quad \mathrm{Pe}=\frac{\langle v\rangle I}{D} \tag{10}$$

yields a transport problem for dimensionless concentration $\tilde{C}(\tilde{x},\tilde{z},\tilde{t})$,

$$\frac{\partial \tilde{C}}{\partial \tilde{t}}+\tilde{v}(\tilde{z},\tilde{t})\frac{\partial \tilde{C}}{\partial x}=\frac{1}{\mathrm{Pe}}\frac{\partial^2 \tilde{C}}{\partial \tilde{x}^2},\qquad \tilde{C}(\tilde{x},\tilde{z},0)=\mathrm{e}^{-(\tilde{x}-\tilde{x}_0)^2/2}. \tag{11}$$

In the following, we omit the tilde˜. The random variable x_0 is assumed to be statistically independent from the random field $v(z, t)$. Similar problem formulations for solute transport in stratified flows have been a subject of numerous investigations [e.g., *Jarman and Tartakovsky*, 2013; *Dentz and Carrera*, 2007; *Li and McLaughlin*, 2002; *McLaughlin and Majda*, 1996]. In the simulations reported below, we treat the source location x_0 as a Gaussian variable with mean $\langle x_0\rangle=0$ and variance σ_0^2, and the flow velocity $v(z, t)$ as either a steady or time-dependent lognormal field with constant mean $\langle v\rangle$ and variance σ_v^2 and no spatial correlation along z. In the time-dependent case, $\ln v$ has an exponential temporal covariance function with correlation time λ. Appendix C summarizes the expressions presented by *Jarman and Tartakovsky* [2013] for lognormal velocity fields $v(z, t)$, without and with time dependence. Unless specified otherwise, the subsequent results are for uncorrelated lognormal velocity field $v(z)$ (section C1) with Péclet number Pe=0.1.

A solution of the CDF equation (4) for this problem is shown in Figure 1 in the form of risk maps of subsurface contamination, i.e., the spatiotemporal evolution of the exceedance probability for a given (e.g., regulatory) concentration value c^*, $\mathbb{P}[C(x,t)>c^*]=1-F_C(c^*;x,t)$. The color scheme in these graphs represents various degrees of certainty in the model predictions (see Table 1). The space-time region over which the predictions are "extremely likely" to be correct increases significantly as the normalized threshold concentration decreases from $c^*=1/50$ (first row) to $c^*=1/100$ (second row).

Figure 1 also provides a comparison between the computed CDF (left column), $F_C(c;x,t)$, and its assumed β-distributed counterpart (right column), $F_C^\beta(c;x,t)=B(c;\alpha_1,\alpha_2)/B(\alpha_1,\alpha_2)$ where $B(c;\alpha_1,\alpha_2)$ and $B(\alpha_1,\alpha_2)$ are the complete and incomplete beta functions, respectively, and the parameters α_1 and α_2 are chosen such that $\langle C\rangle=\alpha_1/(\alpha_1+\alpha_2)$ and $\sigma_C^2=\alpha_1\alpha_2/[(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)]$. While, by construction, F_C and F_C^β have the same mean and variance, their tails differ, resulting in appreciably different risk maps, especially for low threshold concentrations c^*.

4. Accuracy and Robustness of the CDF Method

Two levels of approximation underpin our derivation of the CDF equation (4) in Appendices A and B. One is required to close the formal CDF equation (A2), and the other to close the moment equations (B1) and (B2). We conduct a series of numerical experiments to analyze the accuracy and robustness of these approximations and the resulting CDF equation. This is done by comparing solutions of the CDF equation with MCS of the stochastic transport problem (1)–(2), with (semi)analytical expressions for the corresponding concentration moments $\langle C\rangle$ and σ_C^2, and with the assumed β-CDF F_C^β.

4.1. Comparison of Alternative CDF Solutions

The transport scenario under investigation lends itself to exact analytical treatment in terms of the concentration statistics $\langle C(x,t)\rangle$ and $\sigma_C^2(x,t)$. By eliminating the need for a closure of the MDEs, it facilitates an

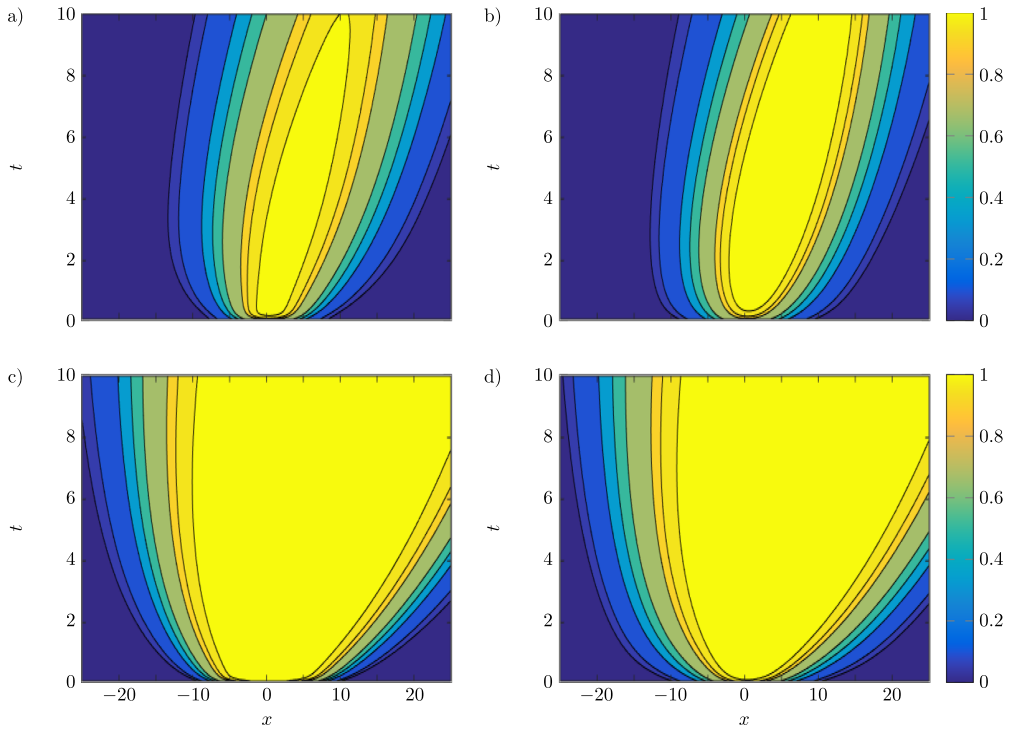

Figure 1. Spatiotemporal maps of the exceedance probability $\mathbb{P}[C(x,t) > c^*] = 1 - F_C(c^*; x, t)$ computed with the CDF equation (4) (left column) and the assumed β-CDF (right column) for concentration threshold values $c^* = 1/50$ (top row) and $c^* = 1/100$ (bottom row). The statistical parameters of the inputs are set to $\langle x_0 \rangle = 0$, $\sigma_0^2 = 10$, $\sigma_v^2 = 0.1$, and Pe = 0.1.

analysis of the accuracy of the approximations used to close the formal CDF equation (A2). The same statistics are used to parameterize the assumed β-CDF F_C^β. To establish the veracity of these two CDF estimates, we compare them with the "exact" CDF computed with MCS, $F_C^{\mathrm{MC}}(c; x, t)$. The latter involved 10,000 realizations, which proved to be sufficient to converge to the exact solutions for $\langle C(x,t) \rangle$ and $\sigma_C^2(x,t)$. The errors of our CDF solution $F_C(c; x, t)$ and the assumed CDF $F_C^\beta(c; x, t)$ are computed as

$$\mathcal{E}_F(x,t) = \int_0^1 |F_C - F_C^{\mathrm{MC}}| \, \mathrm{dc}, \qquad \mathcal{E}_F^\beta(x,t) = \int_0^1 |F_C^\beta - F_C^{\mathrm{MC}}| \, \mathrm{dc}. \tag{12}$$

The numerical quadratures are computed by rectangular integration on the numerical grid points of the CDF method (dc = 0.001).

The errors $\mathcal{E}_F(x,t)$ and $\mathcal{E}_F^\beta(x,t)$ are plotted in Figure 2. The $\mathcal{E}_F(x,t)$ error is confined to the immediate vicinity of the source and to early times, while the $\mathcal{E}_F^\beta(x,t)$ is significantly more persistent in time. The CDF solution error $\mathcal{E}_F(x,t) = \mathcal{E}_F^{\mathrm{m}}(x,t) + \mathcal{E}_F^{\mathrm{n}}(x,t)$ comprises the modeling error $\mathcal{E}_F^{\mathrm{m}}(x,t)$ and the numerical error $\mathcal{E}_F^{\mathrm{n}}(x,t)$. The former stems from the inability of IEM approximations, including ours given (A7) and (A10), to handle the early-time bimodality. The dispersive term in the CDF equation (4) quickly dissipates this bimodality, while allowing the CDF F_C to exhibit intermittency, i.e., to maintain the asymmetry at intermediate times and transition to a Gaussian distribution at later times. The numerical error $\mathcal{E}_F^{\mathrm{n}}(x,t)$ is dominant in space-time regions where the CDF profiles are sharp, i.e., where the concentration variance is small, because of numerical diffusion. At very early times, $\mathcal{E}_F^\beta(x,t) < \mathcal{E}_F(x,t)$, especially in the regions where the CDF is very sharp. However, the performance of F_C^β does not improve with time; it moves towards a symmetric distribution faster than the benchmark MC solution does. In other words, F_C^β fails to capture the asymmetric features of the concentration CDF at intermediate times. Figure S1 in Supporting Information further elucidates the dynamics of $F_C(c; x, t)$, $F_C^\beta(c; x, t)$ and $F_C^{\mathrm{MC}}(c; x, t)$ by exhibiting their temporal snapshots at $x = 0.0$.

At later times, the system approaches its deterministic equilibrium state, which both F_C and F_C^β capture well. In this regime, any method that accurately reproduces the mean concentration $\langle C(x,t) \rangle$ would perform

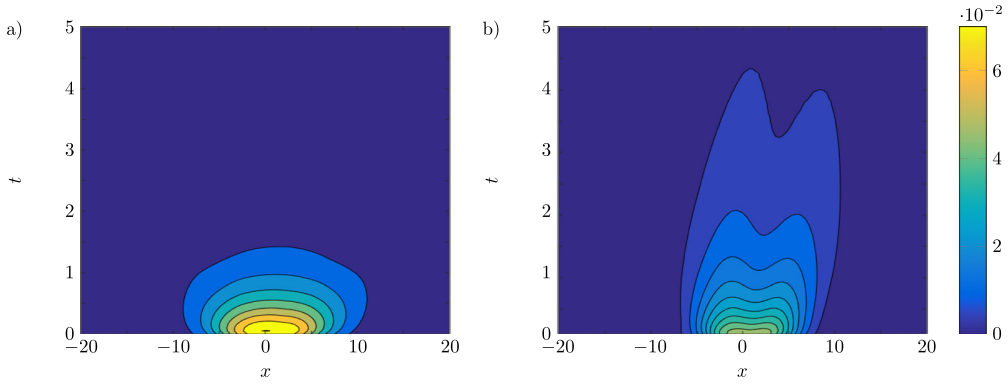

Figure 2. Spatiotemporal maps of the errors of the CDF solution $F_C(c;x,t)$ and the assumed β-CDF $F_C^\beta(c;x,t)$, $\mathcal{E}_F(x,t)$ (left) and $\mathcal{E}_F^\beta(x,t)$ (right), respectively. The statistical parameters of the inputs are set to $\langle x_0\rangle=0$, $\sigma_0^2=10$, $\sigma_v^2=0.1$, and Pe=0.1.

well, since the concentration CDF becomes progressively symmetric and its variance decreases with time. Figure S2 in Supporting Information substantiates this point by showing the error maps for F_C, F_C^β and the Gaussian CDF F_C^G for dimensionless times $t>10$. The error $\mathcal{E}_F$ is practically time-invariant, depending mostly on the accuracy of the numerical solution of the CDF equation (4): $\mathcal{E}_F \sim \mathcal{O}(10^{-3})$, which is the same as the numerical accuracy. The spatial distributions of $\mathcal{E}_F^\beta$ and $\mathcal{E}_F^G$ are more heterogeneous, reaching their respective maxima close to the mean $\langle C(x,t)\rangle$.

Derivation of CDF equations for advection-dominated transport does not require an IEM closure and is often exact [*Boso et al.*, 2014; *Tartakovsky and Gremaud*, 2016]. It is therefore reasonable to expect that the performance of the IEM-based CDF equation (4) depends on the Péclet number Pe. Figure 3 exhibits the errors $\mathcal{E}_F$ and $\mathcal{E}_F^\beta$ for transport regimes with Pe=1.0 and Pe=0.01 (dispersion-dominated). The errors of both PDF

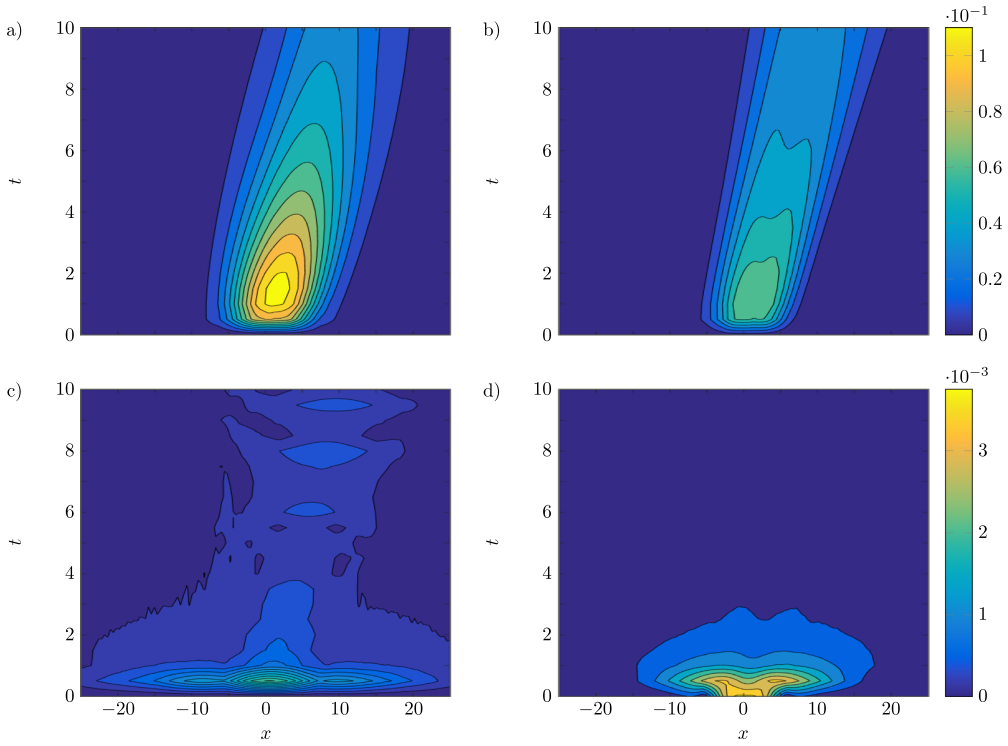

Figure 3. Spatiotemporal maps of the errors $\mathcal{E}_F(x,t)$ (left) and $\mathcal{E}_F^\beta(x,t)$ (right) for transport regimes with Pe=1.0 (top) and Pe=0.01 (bottom). The other statistical parameters of the inputs are set to $\langle x_0\rangle=0$, $\sigma_0^2=10$, and $\sigma_v^2=0.1$.

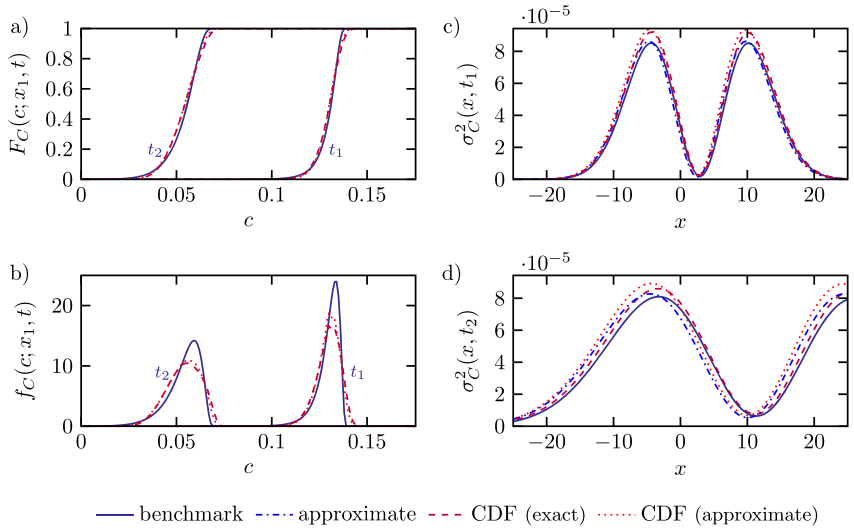

Figure 4. (Left) Profiles of the CDF $F_C(c;x=0.0,t)$ and PDF $f_C(c;x=0.0,t)$ at times $t_1=2.5$ and $t_2=10.0$ computed with the CDF equation (4) (patterned lines) and Monte Carlo simulations (solid lines); $\langle C\rangle$ and σ_C^2 in the CDF equation are given, alternatively, by their exact expressions (dashed lines) and the approximate MDEs (B9) and (B10) (dotted lines). Right column: the concentration variance $\sigma_C^2(x=0.0,t)$ at times $t_1=2.5$ and $t_2=10.0$, computed with its exact expression (solid lines), the approximate MDEs (B9) and (B10) (dash-dotted lines), and integration of $F_C(c;x=0.0,t)$ obtained with the exact (dashed lines) and approximate (dotted lines) expressions for the moments.

estimates decrease by more than one order of magnitude as Pe decreases from 1.0 to 0.01, as more pronounced dispersion acts to homogenize the system leading to a symmetric CDFs. The snapshots of F_C, F_C^β, and F_C^{MC} in Figure S3 in Supporting Information reveal that for Pe=1.0 the bimodality of CDFs/PDFs persists for longer times, causing F_C and F_C^β to deviate more from the MCS benchmark F_C^{MC}. This bimodality is largely absent in the dispersion-dominated regime (Pe=0.01), resulting in significantly reduced approximation errors. For advection-dominated regimes it is advisable to resort to a different closure of the CDF equation [*Boso et al.*, 2014], with closure coefficients that account for random advection.

4.2. Sensitivity to Approximation of Moments

The F_C solutions reported above were obtained by solving the CDF equation (4) with the analytical expressions for $\langle C\rangle$ and σ_C^2. In general, one would have to deploy closure approximations to derive the MDEs and, hence, to obtain numerical approximations of $\langle C\rangle$ and σ_C^2. Based on the analysis of *Jarman and Tartakovsky* [2013] we adopted the macrodispersion approximation (see Appendix B). Figure 4 shows $F_C(c;x,t)$ computed with the CDF equation (4) in which we alternatively used the exact expressions for $\langle C\rangle$ and σ_C^2 and their approximations given by solutions of the MDEs (B9) and (B10); the PDFs were computed by numerical differentiation of the corresponding CDFs. While the approximation error for the moments grows with time (only the concentration variance σ_C^2 is shown), the accuracy of the CDF and PDF solutions remains relatively insensitive to it.

4.3. Impact of Increasing Input Uncertainty

The CDF method is expected to loose accuracy as the input variances, σ_0^2 and σ_v^2, increase. To isolate the effect of the input variances on the robustness of the IEM-based closure, we use the semi-analytical expressions for $\langle C\rangle$ and σ_C^2 in the CDF equation (4). A larger uncertainty (variance) in the inputs translates into a large uncertainty in the output: σ_C^2 increases with σ_0^2 and/or σ_v^2 (Figure S4 in Supporting Information). This smooths the CDF profiles and widens the corresponding PDF (Figure S5 in Supporting Information). The overall accuracy of the CDF solution $F_C(c;x,t)$, quantified in terms of the error $\mathcal{E}_F(x,t)$, deteriorates as σ_0^2 and/or σ_v^2 become larger (Figure 5). Yet, $F_C(c;x,t)$ reproduces the asymmetric features of the distributions, which are especially visible in the PDF plots in Figure S5. Figure S5 also reveals that $F_C(c;x,t)$ predicted with the CDF method has longer tails than the exact (MCS-based) CDF, i.e., it provides a conservative estimate of uncertainty over the whole range of possible concentration levels. This is in contrast to the assumed β-CDF $F_C^\beta(c;x,t)$, which underestimates uncertainty at low and high concentration levels.

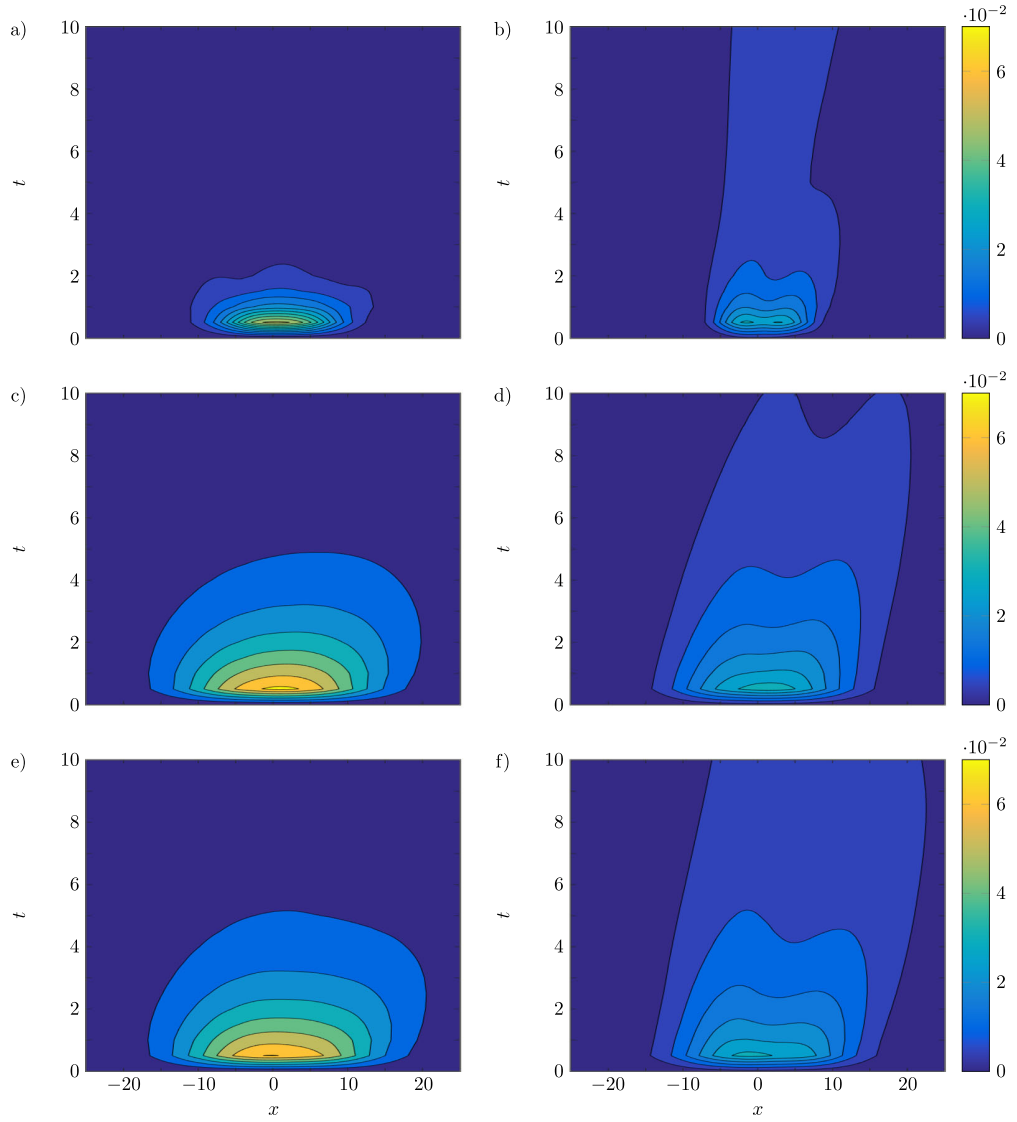

Figure 5. Spatiotemporal maps of the errors $\mathcal{E}_F(x,t)$ (left) and $\mathcal{E}_F^{\beta}(x,t)$ (right) for $\langle x_0 \rangle = 0$, Pe = 0.1 and different combinations of the degrees of uncertainty: $\sigma_0^2 = 10.0$ and $\sigma_v^2 = 0.5$ (first row), $\sigma_0^2 = 50.0$ and $\sigma_v^2 = 0.1$ (second row), and $\sigma_0^2 = 50.0$ and $\sigma_v^2 = 0.5$ (third row).

4.4. Impact of Temporal Correlation

For temporally correlated velocity fields, analytical expressions for the moments $\langle C \rangle$ and σ_C^2 are not available; instead, they are computed by solving the MDEs (B9) and (B10). Figure 6 reveals that $\mathcal{E}_F(x,t)$ slightly decreases as the dimensionless correlation length λ changes from 1 to 100. This effect is also observed in Figure S6 in Supporting Information, which compares the CDF and PDF profiles computed with the CDF method, the β-CDF, and the MCS for $\lambda = 1$ and 100. Increasing λ does not affect much the space-time behavior of the mean $\langle C \rangle$, but its effect on the variance σ_C^2 is more pronounced (Figure S7 in Supporting Information).

5. Summary

We developed a deterministic equation for the CDF of solute concentration, whose behavior is governed by an advection-dispersion equation with uncertain (random) flow velocity and initial conditions. This CDF equation is closed by introducing a nonlinear "drift" velocity, which depends on the mean and variance of

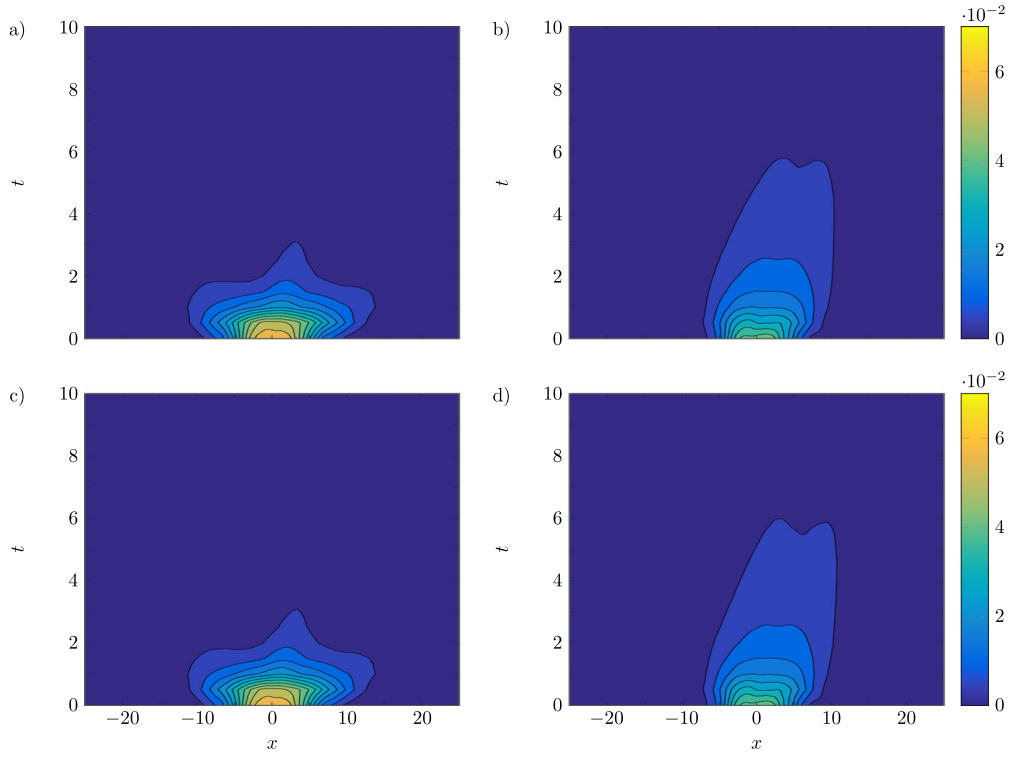

Figure 6. Spatiotemporal maps of the errors $\mathcal{E}_F(x,t)$ (left) and $\mathcal{E}_F^{\beta}(x,t)$ (right) for the correlation time of the velocity field $v(z,t)$ set to $\lambda=1.0$ (first row) and $\lambda=100.0$ (second row). The other input statistics are $\langle x_0\rangle=0$ and Pe=0.1.

solute concentration. In contrast to the classic IEM closure and its modifications, which generally fail to reproduce these statistics, our closure preserves the mean and variance. It does so by adding a correction that is proportional to the scalar dissipation rate, a measure of local mixing. The resulting CDF equation is capable of reproducing asymmetric CDF/PDF profiles at intermediate times and Gaussin-like profiles at late times. At very early times, the CDF method does not capture the bimodality imposed by initial conditions. Validation against (semi-)analytical solutions and Monte Carlo simulations and comparison with the β distribution show that the CDF method is robust with respect to both the correlation lengths of the random parameters and their variances.

The CDFs and PDFs obtained with the proposed method of distributions can be used as priors to be refined through Bayesian update as concentration measurements become available. An efficient numerical implementation of this approach to data assimilation is a subject of our ongoing research.

Appendix A: Closures for the CDF Method

It follows from the definition of Π that $\partial\Pi/\partial t=-(\partial\Pi/\partial c)(\partial C/\partial t)$ and $\nabla\Pi=-(\partial\Pi/\partial c)\nabla C$. Multiplying both sides of (1) with $\partial\Pi/\partial c$ and using these relations yields

$$\frac{\partial\Pi}{\partial t}+\mathbf{v}\cdot\nabla\Pi=\nabla\cdot(\mathbf{D}\nabla\Pi)-(\mathbf{D}\nabla C)\cdot\nabla C\frac{\partial^2\Pi}{\partial c^2}. \tag{A1}$$

Employing a Reynolds decomposition, $\mathbf{v}=\langle\mathbf{v}\rangle+\mathbf{v}'$ and $\Pi=F_c+\Pi'$ where the primed quantities denote zero-mean fluctuations about their respective means, and taking the ensemble average of (A1) yields a formal, i.e., unclosed, equation for $F_C(c;\mathbf{x},t)$

$$\frac{\partial F_C}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla F_C=\nabla\cdot(\mathbf{D}\nabla F_C-\mathbf{Q})-M,\quad \mathbf{Q}\equiv\langle\mathbf{v}'\Pi'\rangle,\quad M\equiv\left\langle\mathbf{D}\nabla C\cdot\nabla C\frac{\partial^2\Pi}{\partial c^2}\right\rangle, \tag{A2}$$

which requires closure approximations to compute $\mathbf{Q}$ and M.

The derivation of an approximation for $\mathbf{Q}(\mathbf{x},t)$ starts with an equation for the fluctuations Π', which is obtained by subtracting (A2) from (A1). Its solution is written in terms of the random Green's function $\mathcal{G}(\mathbf{x},\mathbf{y},t-\tau)$ as

$$\Pi'(\mathbf{x},t)=-\int_0^t\int_\Omega\left(\mathbf{v}'\cdot\nabla F_C+\mathbf{D}\nabla C\cdot\nabla C\frac{\partial^2\Pi}{\partial c^2}-\nabla\cdot\langle\mathbf{v}'\Pi'\rangle-\left\langle\mathbf{D}\nabla C\cdot\nabla C\frac{\partial^2\Pi}{\partial c^2}\right\rangle\right)\mathcal{G}\mathrm{d}\mathbf{y}\mathrm{d}\tau$$

$$+\int_\Omega\Pi_0'\mathcal{G}(\tau=0)\mathrm{d}\mathbf{y}+\int_0^t\int_{\partial\Omega}\Pi_\phi'\mathbf{n}\cdot\mathbf{D}\nabla\mathcal{G}\mathrm{d}\mathbf{y}\mathrm{d}\tau, \tag{A3}$$

where Π_0' and Π_ϕ' represent the initial and boundary fluctuations, and $\mathbf{n}$ is the unit normal vector to the boundary $\partial\Omega$. The Green's function $\mathcal{G}$ is defined as a solution of

$$\frac{\partial\mathcal{G}}{\partial\tau}-\nabla\cdot(\mathbf{v}\mathcal{G})=\nabla\cdot(\mathbf{D}\nabla\mathcal{G})+\delta(\mathbf{x}-\mathbf{y})\delta(t-\tau), \tag{A4}$$

where $\delta(\cdot)$ is the Dirac delta function, subject to the homogeneous initial and boundary conditions. Multiplying (A3) with $\mathbf{v}'(\mathbf{x},t)$ and taking the ensemble mean yields

$$\mathbf{Q}=-\int_0^t\int_\Omega[\langle\mathcal{G}\mathbf{v}'(x,t)\mathbf{v}'(\mathbf{y},\tau)^\top\rangle\cdot\nabla F_C+\langle\mathbf{v}'M'\mathcal{G}\rangle]\mathrm{d}\mathbf{y}\mathrm{d}\tau,\quad M'=\mathbf{D}\nabla C\cdot\nabla C\frac{\partial^2\Pi}{\partial c^2}-M. \tag{A5}$$

The initial and boundary terms are zero because the sources of uncertainty are mutually uncorrelated. In Appendix B we show that approximating (A5) with

$$Q(\mathbf{x},t)\approx-\int_0^t\int_\Omega G\langle\mathbf{v}'(\mathbf{x},t)\mathbf{v}'(\mathbf{y},\tau)^\top\rangle\cdot\nabla F_C\,\mathrm{d}\mathbf{y}\mathrm{d}\tau \tag{A6}$$

leads to a CDF equation that reproduces the mean $\langle C(\mathbf{x},t)\rangle$ predicted with the corresponding MDE. Here $G(\mathbf{x},\mathbf{y},t-\tau)$ is the mean-field approximation of the Green's function $\mathcal{G}$ obtained by replacing the random velocity $\mathbf{v}$ with its $\langle\mathbf{v}\rangle$ in (A4). Finally, localizing (A6), i.e., assuming that the CDF gradient ∇F_C varies slowly in space and time and, hence, can be taken outside the integral, yields a dispersive closure $\mathbf{Q}(\mathbf{x},t)\approx-\mathcal{D}_m\nabla F_C$, where the macrodispersion tensor $\mathcal{D}_m(\mathbf{x},t)$ is defined by (5). This expression for $\mathbf{Q}$ is similar to the large-eddy-diffusivity closure developed by *Boso et al*. [2014] for advection-reaction equations.

Classic perturbation techniques do not allow one to obtain workable closure expressions for M. Inspired by *Villermaux and Falk* [1994], we look for a closure for M in the form

$$M=\mathcal{V}(\mathbf{x},t)(c-\langle C(\mathbf{x},t)\rangle)\frac{\partial F_C}{\partial c}, \tag{A7}$$

and determine the constant of proportionality $\mathcal{V}(\mathbf{x},t)$ by requiring the resulting CDF equation to yield the same mean, $\langle C(\mathbf{x},t)\rangle$, and variance, $\sigma_C^2(\mathbf{x},t)$, as the MDEs,

$$\langle C\rangle=C_{\max}-\int_{C_{\min}}^{C_{\max}}F_C(c;\mathbf{x},t)\mathrm{d}c,\qquad\sigma_C^2=C_{\max}^2-2\int_{C_{\min}}^{C_{\max}}cF_C(c;\mathbf{x},t)\mathrm{d}c-\langle C\rangle^2. \tag{A8}$$

Substituting the macrodispersion closure $\mathbf{Q}=-\mathcal{D}_m\nabla F_C$ into (A2), integrating over $c\in[C_{\min},C_{\max}]$, and accounting for the boundary conditions (6c) yields the MDE for the mean (B9) for any choice of $\mathcal{V}(x,t)$. Hence, our closure (A7) automatically preserves the mean, without constraining $\mathcal{V}$. The same procedure, but applied to (A2) multiplied by c, leads to

$$\frac{\partial\sigma_C^2}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla\sigma_C^2=\nabla\cdot(\mathbf{D}_{\mathrm{eff}}\nabla\sigma_C^2)+2(\mathbf{D}_{\mathrm{eff}}\nabla\langle C\rangle)\cdot\nabla\langle C\rangle+2\mathcal{V}\sigma_C^2, \tag{A9}$$

after accounting for (B9) and setting $\mathbf{D}_{\mathrm{eff}}=\mathbf{D}+\mathcal{D}_m$. Comparing (A9) with the MDE for the variance MDE (B10) defines $\mathcal{V}$ as

$$\mathcal{V}=-\frac{\chi}{2}-\frac{(\mathbf{D}\nabla\langle C\rangle)\cdot\nabla\langle C\rangle}{\sigma_C^2},\qquad\chi=\frac{2\xi}{\sigma_C^2},\qquad\xi=\langle\mathbf{D}\nabla C'(\mathbf{x},t)\cdot\nabla C'(\mathbf{x},t)\rangle. \tag{A10}$$

Substituting these approximations for **Q** and M into (A2) yields (4). Derivation of the closures **Q** and M, and the resulting CDF equation (4), is the main result of our study.

Appendix B: MDEs and Macrodispersion Theory

When applied to (1), a standard procedure for derivation of MDEs [e.g., *Morales-Casique et al*., 2006] yields unclosed equations for the mean and variance of C,

$$\frac{\partial\langle C\rangle}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla C=\nabla\cdot(\mathbf{D}\nabla\langle C\rangle)-\nabla\cdot\langle\mathbf{v}'C'\rangle, \tag{B1}$$

$$\frac{\partial\sigma_C^2}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla\sigma_C^2=\nabla\cdot(\mathbf{D}\nabla\sigma_C^2)-2\langle\mathbf{D}\nabla C'\cdot\nabla C'\rangle-\langle\mathbf{v}'C'\rangle\cdot\nabla\langle C\rangle-\nabla\cdot\langle\mathbf{v}'C'2\rangle. \tag{B2}$$

They require closure expressions to compute the means of the fluctuation products. This can be achieved with different approximation strategies, including recursive perturbation expansions [*Tartakovsky and Neuman*, 1998] and nonlocal models [*Morales-Casique et al*., 2006]. We adopt a closure based on the "macrodispersion theory" [e.g., *Morales-Casique et al*., 2006, and the references therein] because it seems to provide accurate results for a range of test problems [*Jarman and Tartakovsky*, 2013].

Derivation of this and the other closures mentioned above starts with an equation for concentration fluctuations, obtained by subtracting (B1) from (1),

$$\frac{\partial C'}{\partial t}+\nabla\cdot(\mathbf{v}C')=\nabla\cdot(\mathbf{D}\nabla C')-\nabla\cdot(\mathbf{v}'\langle C\rangle)+\nabla\cdot\langle\mathbf{v}'C'\rangle. \tag{B3}$$

Its formal solution is written in terms of the random Green's function $\mathcal{G}(\mathbf{x},\mathbf{y},t-\tau)$ in (A4) as

$$C'(\mathbf{x},t)=-\int_0^t\int_\Omega\nabla\cdot(\mathbf{v}'\langle C\rangle)\mathcal{G}\mathrm{d}\mathbf{y}\mathrm{d}\tau+\int_0^t\int_\Omega\mathcal{G}\nabla\cdot\langle\mathbf{v}'C'\rangle\mathrm{d}\mathbf{y}\mathrm{d}\tau+\text{a.i.t.}, \tag{B4}$$

where "a.i.t." refers to the auxiliary integral terms representing the contribution of the initial and boundary conditions, similar to the last two terms in (A3); contribution of these terms to the subsequent solutions is exactly 0 because random fluctuations of the auxiliary functions C_0' and ϕ' are statistically independent from the velocity fluctuations $\mathbf{v}'$. Multiplying (B4) with $\mathbf{v}'(\mathbf{x},t)$ and taking the ensemble average gives

$$\langle\mathbf{v}'C'\rangle=-\int_0^t\int_\Omega\langle\mathbf{v}'(\mathbf{x},t)\mathbf{v}'(\mathbf{y},\tau)^\top\mathcal{G}\rangle\nabla\langle C\rangle\mathrm{d}\mathbf{y}\mathrm{d}\tau+\int_0^t\int_\Omega\langle\mathbf{v}'(\mathbf{x},\mathrm{t})\mathcal{G}\rangle\nabla\cdot\langle\mathbf{v}'C'\rangle\mathrm{d}\mathbf{y}\mathrm{d}\tau. \tag{B5}$$

The lowest-order approximation of this expression is given by

$$\langle\mathbf{v}'C'\rangle\approx-\int_0^t\int_\Omega\langle\mathbf{v}'(\mathbf{x},t)\mathbf{v}'(\mathbf{y},\tau)^\top\rangle\,G\,\nabla\langle C\rangle\mathrm{d}\mathbf{y}\mathrm{d}\tau, \tag{B6}$$

where G is the mean-field Green's function defined in (A6). Localization, which is based on the assumption that the mean concentration gradient $\nabla\langle C(\mathbf{y},\tau)\rangle$ varies slowly in space and time, yields the macrodispersion closure

$$\langle\mathbf{v}'C'\rangle\approx-\mathcal{D}_m(\mathbf{x},t)\nabla\langle u(\mathbf{x},t)\rangle, \tag{B7}$$

where $\mathcal{D}_m$ is the macrodispersion tensor defined in (5). (B7) can be used to close (B1) and (B2). Similar manipulations with (B4) are used to compute approximate relations for $\sigma_C^2(\mathbf{x},t)=\langle[C'(\mathbf{x},t)]^2\rangle$ and $\xi(\mathbf{x},t)=\langle\mathbf{D}\nabla C'(\mathbf{x},t)\cdot\nabla C'(\mathbf{x},t)\rangle$. Finally, we use the *Kapoor and Kitanidis* [1998] closure

$$\langle\mathbf{v}'C'^2\rangle\approx-\mathcal{D}_m\nabla\sigma_C^2. \tag{B8}$$

Substituting these closures into (B1) and (B2) gives a set of closed MDEs for the mean and variance of solute concentration C,

$$\frac{\partial\langle C\rangle}{\partial t}+\langle\mathbf{v}\rangle\cdot\nabla C=\nabla\cdot(\mathbf{D}_{\mathrm{eff}}\nabla\langle C\rangle), \tag{B9}$$

$$\frac{\partial \sigma_C^2}{\partial t} + \langle \mathbf{v} \rangle \cdot \nabla \sigma_C^2 = \nabla \cdot (\mathbf{D}_{\text{eff}} \nabla \sigma_C^2) - \chi \sigma_C^2 + 2\mathcal{D}_m \nabla \langle C \rangle \cdot \nabla \langle C \rangle, \tag{B10}$$

where $\mathbf{D}_{\text{eff}} \equiv \mathbf{D} + \mathcal{D}_m$ is the effective macrodispersion coefficient, and $\chi = 2\xi/\sigma_C^2$ is the "variance destruction rate." The latter can be either used as a fitting parameter [e.g., *Kapoor and Kitanidis*, 1998] or approximated by using its definition and the closure relations for ξ and σ_C^2. In the simulations reported in this study, we adopt the second approach.

Appendix C: Computational Examples

Solute transport in stratified flow has been used by *Jarman and Tartakovsky* [2013] as a computational testbed for comparison and validation of several closure strategies for the MDEs. We use this setting to analyze the performance of our CDF equation, and extend it by including uncertainty (randomness) in the initial condition.

C1. Stratified Flow With Steady Lognormal Velocity

For steady flow velocity $v(z)$, the ADE (8) admits an analytical solution

$$C_{\text{ex}}(x, z, t) = \frac{1}{\sqrt{2\pi(2\text{Pe}^{-1}t + 1)}} \exp\left(-\frac{(x - x_0 - v(z)t)^2}{4\text{Pe}^{-1}t + 2} \right), \tag{C1}$$

which establishes an explicit relation between solute concentration, $C(x, z, t)$, the center of mass of the initial plume, x_0, and the flow velocity $v(z)$. This enables one to compute, by direct integration, the moments of C from the PDFs of $v(z)$ and x_0, denoted by f_v and f_0, respectively. To be concrete, we take f_v and f_0 to be lognormal and Gaussian, respectively, with constant means, $\langle v \rangle$ and $\langle x_0 \rangle$, and variances, σ_v^2 and σ_0^2. The numerical integration necessary to compute the moments of $C(x, z, t)$ is performed with the trapezoidal rule with dv=0.03. Assuming the lack of spatial correlation of $v(z)$ along z, the macrodispersion coefficient in (5) is computed analytically, $\mathcal{D}_m = \sigma_v^2 t$, as are all the other closure parameters in the MDEs (Appendix B).

C2. Stratified Flow With Time-Dependent Lognormal Distribution

In this setting, flow velocity is a time-dependent random field, $v(z, t)$, uncorrelated in space and correlated in time. The field $Y(\cdot, t) = \ln v$ has an exponential covariance function with correlation time λ, so that the temporal covariance for the velocity at a given $\bar{z}$ is

$$\langle v'(\bar{z}, t) v'(\bar{z}, \tau) \rangle = \sigma_v^2 = \exp\left[2\mu_Y + \sigma_Y^2 + \sigma_Y^2 \exp\left(-|t - \tau|/\lambda\right)\right] - \exp\left(2\mu_Y\right) + \sigma_Y^2, \tag{C2}$$

where μ_Y and σ_Y^2 are the mean and variance of Y, respectively.

A (semi-)analytical solution of the ADE (8) is

$$C_{\text{ex}}(x, z, t) = \frac{1}{\sqrt{2\pi(2\text{Pe}^{-1}t + 1)}} \exp\left[-\frac{1}{4\text{Pe}^{-1}t + 1} \left(x - x_0 - \int_0^t v(z, \tau) \mathrm{d}\tau \right)^2 \right]. \tag{C3}$$

The solution for each realization of $v(z, t)$ and x_0, as well as the moments of C, are computed via numerical integration.

Acknowledgments
This work was supported in part by Defense Advanced Research Projects Agency under the EQUiPS program, and by the National Science Foundation under grants EAR-1246315 and DMS-1522799. 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

Bellin, A., and D. Tonina (2007), Probability density function of non-reactive solute concentration in heterogeneous porous formations, *J. Contam. Hydrol.*, *94*(1–2), 109–125.
Bellin, A., G. Severino, and A. Fiori (2011), On the local concentration probability density function of solutes reacting upon mixing, *Water Resour. Res.*, *47*, W01514, doi:10.1029/2010WR009696.
Berkowitz, B., A. Cortis, M. Dentz, and H. Scher (2006), Modeling non-Fickian transport in geological formations as a continuous time random walk, *Rev. Geophys.*, *44*, RG2003, doi:10.1029/2005RG000178.
Boso, F., S. V. Broyda, and D. M. Tartakovsky (2014), Cumulative distribution function solutions of advection-reaction equations with uncertain parameters, *Proc. R. Soc. A*, *470*(2166), 20140189.
Cushman, J. H. (1997), *The Physics of Fluids in Hierarchical Porous Media: Angstroms to Miles*, Kluwer Academic, N. Y.
Cushman, J. H., and D. O'Malley (2015), Fickian dispersion is anomalous, *J. Hydrol.*, *531*(1), 161–167.
Cushman, J. H., L. Bennethum, and B. X. Hu (2002), A primer on upscaling tools for porous media, *Adv. Water Resour.*, *25*, 1043–1067.
Dagan, G. (1987), Theory of solute transport by groundwater, *Annu. Rev. Fluid Mech.*, *19*, 183–215.

Dagan, G. (1989), *Flow and Transport in Porous Formations*, 461 pp., Springer, Berlin.
Dagan, G., and S. P. Neuman (Eds.) (1997), *Subsurface Flow and Transport: A Stochastic Approach*, Cambridge, N. Y.
de Barros, F. P. J., and A. Fiori (2014), First-order based cumulative distribution function for solute concentration in heterogeneous aquifers: Theoretical analysis and implications for human health risk assessment, *Water Resour. Res.*, *50*, 4018–4037, doi:10.1002/2013WR015024.
Dentz, M. (2012), Concentration statistics for transport in heterogeneous media due to stochastic fluctuations of the center of mass velocity, *Adv. Water Resour.*, *36*, 11–22.
Dentz, M., and J. Carrera (2007), Mixing and spreading in stratified flow, *Phys. Fluids*, *19*(1), 017107.
Dentz, M., and D. M. Tartakovsky (2010), Probability density functions for passive scalars dispersed in random velocity fields, *Geophys. Res. Lett.*, *37*, L24406, doi:10.1029/2010GL045748.
Fiori, A. (2001), The Lagrangian concentration approach for determining dilution in aquifer transport: Theoretical analysis and comparison with field experiments, *Water Resour. Res.*, *37*(12), 3105–3114.
Haworth, D. (2010), Progress in probability density function methods for turbulent reacting flows, *Prog. Energy Combust.*, *36*(11), 168–259.
Jarman, K. D., and A. M. Tartakovsky (2013), A comparison of closures for stochastic advection-diffusion equations, *SIAM J. Uncert. Quant.*, *1*, 319–347.
Kapoor, V., and P. K. Kitanidis (1998), Concentration fluctuations and dilution in aquifers, *Water Resour. Res.*, *34*(5), 1181–1193.
Li, S.-G., and D. McLaughlin (2002), Asymptotic properties of the Eulerian truncation approximation: Analysis of the perfectly stratified transport problem, *Water Resour. Res.*, *38*(8), 18-1–18-7, doi:10.1029/2000WR000040.
Lichtner, P. C., and D. M. Tartakovsky (2003), Stochastic analysis of effective rate constant for heterogeneous reactions, *Stochastic Environ. Res. Risk Assess.*, *17*(6), 419–429.
Lundgren, T. S. (1967), Distribution functions in the statistical theory of turbulence, *Phys. Fluids*, *10*(5), 969–975.
McLaughlin, R. M., and A. J. Majda (1996), An explicit example with non-Gaussian probability distribution for nontrivial scalar mean and fluctuation, *Phys. Fluids*, *8*(2), 536–547.
Morales-Casique, E., S. P. Neuman, and A. Guadagnini (2006), Non-local and localized analyses of non-reactive solute transport in bounded randomly heterogeneous porous media: Theoretical framework, *Adv. Water Resour.*, *29*(8), 1238–1255.
Neuman, S. P., and D. M. Tartakovsky (2009), Perspective on theories of anomalous transport in heterogeneous media, *Adv. Water Resour.*, *32*(5), 670–680.
Pope, S. B. (2000), *Turbulent Flows*, Cambridge Univ. Press, Cambridge, U. K.
Raman, V., H. Pitsch, and R. O. Fox (2005), Hybrid large-Eddy simulation/Lagrangian filtered-density-function approach for simulating turbulent combustion, *Combust. Flame*, *143*(1–2), 56–78.
Shvidler, M., and K. Karasaki (2003), Probability density functions for solute transport in random field, *Transp. Porous Media*, *50*(3), 243–266.
Tartakovsky, D. M. (2013), Assessment and management of risk in subsurface hydrology: A review and perspective, *Adv. Water Resour.*, *51*, 247–260.
Tartakovsky, D. M., and P. A. Gremaud (2016), Method of distributions for uncertainty quantification, in *Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon, and H. Owhadi, Springer, Zurich, Switzerland.
Tartakovsky, D. M., and S. P. Neuman (1998), Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations, *Water Resour. Res.*, *34*(1), 1–12.
Tartakovsky, D. M., M. Dentz, and P. C. Lichtner (2009), Probability density functions for advective-reactive transport in porous media with uncertain reaction rates, *Water Resour. Res.*, *45*, W07414, doi:10.1029/2008WR007383.
Venturi, D., D. M. Tartakovsky, A. M. Tartakovsky, and G. E. Karniadakis (2013), Exact PDF equations and closure approximations for advective-reactive transport, *J. Comput. Phys.*, *243*, 323–343.
Villermaux, J., and L. Falk (1994), A generalized mixing model for initial contacting of reactive fluids, *Chem. Eng. Sci.*, *49*(24), 5127–5140.
Winter, C. L., C. M. Newman, and S. P. Neuman (1984), A perturbation expansion for diffusion in a random velocity field, *SIAM J. Appl. Math.*, *44*(2), 411–424.