PROCEEDINGS OF THE ROYAL SOCIETY A

rspa.royalsocietypublishing.org

Cumulative distribution function solutions of advection–reaction equations with uncertain parameters

F. Boso, S. V. Broyda and D. M. Tartakovsky

Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Research

Cite this article: Boso F, Broyda SV, Tartakovsky DM. 2014 Cumulative distribution function solutions of advection–reaction equations with uncertain parameters. *Proc. R. Soc. A* **470**: 20140189.
http://dx.doi.org/10.1098/rspa.2014.0189

Received: 9 March 2014
Accepted: 12 March 2014

Subject Areas:
mathematical modelling, applied mathematics, mathematical physics

Keywords:
hyperbolic conservation law, uncertainty quantification, random coefficient, stochastic modelling

Author for correspondence:
D. M. Tartakovsky
e-mail: dmt@ucsd.edu

We derive deterministic cumulative distribution function (CDF) equations that govern the evolution of CDFs of state variables whose dynamics are described by the first-order hyperbolic conservation laws with uncertain coefficients that parametrize the advective flux and reactive terms. The CDF equations are subjected to uniquely specified boundary conditions in the phase space, thus obviating one of the major challenges encountered by more commonly used probability density function equations. The computational burden of solving CDF equations is insensitive to the magnitude of the correlation lengths of random input parameters. This is in contrast to both Monte Carlo simulations (MCSs) and direct numerical algorithms, whose computational cost increases as correlation lengths of the input parameters decrease. The CDF equations are, however, not exact because they require a closure approximation. To verify the accuracy and robustness of the large-eddy-diffusivity closure, we conduct a set of numerical experiments which compare the CDFs computed with the CDF equations with those obtained via MCSs. This comparison demonstrates that the CDF equations remain accurate over a wide range of statistical properties of the two input parameters, such as their correlation lengths and variance of the coefficient that parametrizes the advective flux.

1. Introduction

Hyperbolic conservation laws, also known as advection–reaction equations (AREs), are ubiquitous in many fields of science and engineering [1, §1.2]. They are used to describe phenomena as diverse as migration of reactive contaminants in the environment [2], immiscible

Royal Society Publishing

multi-phase flow in oil reservoirs [3], haemodynamics [4], river flows [5], car traffic [6], transport of plasma blobs in tokamaks [7] and population dynamics [8]. These (typically nonlinear) hyperbolic partial differential equations (PDEs) describe the evolution in the d-dimensional space ($\mathbf{x} \in \mathbb{R}^d$) and time ($t \in \mathbb{R}^+$) of an N-dimensional set of state variables $\mathbf{c}(\mathbf{x}, t) = \{c_n(\mathbf{x}, t)\}_{n=1}^N$, such that $\mathbf{c}(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^N$.

Let $\mathbf{J}(\mathbf{x}, t, \mathbf{c}) = \{\mathbf{J}_n(\mathbf{x}, t, \mathbf{c})\}_{n=1}^N$ (with each $\mathbf{J}_n : \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}^d$) and $\mathbf{f}(\mathbf{x}, t, \mathbf{c}) : \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}^N$ denote the (linear or nonlinear) flux of the state variables $\mathbf{c}(\mathbf{x}, t)$ and the rate of its production/consumption, respectively. Then hyperbolic conservation laws for the state variables $\mathbf{c}(\mathbf{x}, t)$ take the form of an ARE

$$\frac{\partial \mathbf{c}}{\partial t} + \nabla_{\mathbf{x}} \mathbf{J}(\mathbf{x}, t, \mathbf{c}) = \mathbf{f}(\mathbf{x}, t, \mathbf{c}). \tag{1.1}$$

The ubiquity of applications renders such AREs relevant in their own right. They are also important as approximate representations of transport phenomena in the limit of negligible diffusion, e.g. advective transport of conservative [9,10] and reactive [11, §2.6, and references therein] solutes in porous media, and a logistic population growth [8,12].

Equally ubiquitous is uncertainty about spatial variability of system parameters entering AREs (1.1). In a typical application, such parametric uncertainty arises from multi-scale heterogeneity of ambient environments, data scarcity and measurement errors. Examples of the parameters whose spatial distributions are notoriously uncertain include the permeability of a (natural) porous medium [3], the friction coefficient of surface topography affecting overland flow [13] and reaction rates of (bio)chemical reactions [14]. The first two examples affect the flux $\mathbf{J}$, while the last one parametrizes the source term $\mathbf{f}$.

A standard approach to quantification of parametric uncertainty is to treat the under-sampled parameters as random fields, which renders the governing equations stochastic, even though an underlying physical phenomenon is deterministic (e.g. [15–17]). Monte Carlo simulations (MCSs) are often used to solve PDEs with random coefficients. MCSs of this type consist of generating multiple realizations of system parameters, solving the corresponding deterministic PDEs for each of these realizations, and post-processing the resulting solutions to compute the ensemble moments (typically the mean and variance) of the system states. While conceptually straightforward and 'exact', MCSs have a number of potential drawbacks. Their slow convergence often makes MCSs computationally prohibitive, especially when the number of uncertain parameters is large and a phenomenon, e.g. advection–reaction transport described by (1.1), is highly nonlinear. Furthermore, the computational burden of MCSs increases as the correlation length of input parameters decreases. That is because an accurate representation of the random parameter fields requires at least five nodes (elements) per correlation length (e.g. [18,19] and references therein).

Convergence of MCSs can be improved by employing various flavours of quasi-random sampling of a parameter space, including stratified and Latin hypercube sampling. Other numerical approaches that often outperform direct MCSs include polynomial chaos expansions and stochastic collocation on sparse grids. Under certain conditions, especially when random system parameters exhibit short correlation lengths, such approaches can be computationally more demanding than MCSs (e.g. [20], [21, §3.3.3, and references therein]). Application of these numerical techniques to hyperbolic PDEs (1.1) is especially challenging owing to degradation of their convergence rate with time [22–24]. Additionally, like MCSs such approaches provide little or no physical insight into the computed moments of system states.

By deriving deterministic PDEs for ensemble moments (typically, mean and variance) of system states, the moment differential equations (MDEs) approach provides a conceptual alternative to the numerical methods described above. In general, MDEs require a closure approximation, such as perturbation expansions in the (small) variance of system parameters (e.g. [25, and references therein]) or Gaussianity assumptions [26]. These and other closures limit the range of applicability of such analyses. For nonlinear hyperbolic PDEs (1.1), perturbation-based MDE solutions were shown either to exhibit non-physical behaviour [3] or to diverge at

finite times [27]. Even when accurate, MDEs yield only the first few statistical moments of system states rather than their full probability density functions (PDFs).

PDF methods overcome the latter shortcoming as they result in deterministic differential equations for PDFs of system states. They have an added benefit of allowing one to avoid linearization of nonlinear terms in equations similar to (1.1). PDF methods were originally developed to model turbulent flows [28], where it is common to assume that flow domains are infinite, and random parameters (e.g. flow velocity) are statistically homogeneous (stationary) and Gaussian. These (overly restrictive) assumptions were relaxed to relate uncertainty in reaction rate coefficients to PDFs of concentration $c(\mathbf{x}, t)$, which is governed by an ARE with a linear and deterministic flux $\mathbf{J}$ [2,29].

The lack of uniquely defined boundary conditions in the probability space remains a major drawback of PDF methods, which would have to be addressed if they were to be used for uncertainty quantification. In this analysis, we resolve this issue by deriving a deterministic equation satisfied by a cumulative distribution function (CDF) of the state variable $\mathbf{c}(\mathbf{x}, t)$. Specification of boundary conditions for the resulting CDF equation is unique and straightforward. Another contribution of our analysis is to account for parametric uncertainty (randomness) in both the flux $\mathbf{J}$, which is assumed to be linear in the state variable $\mathbf{c}$, and the reaction term $\mathbf{f}(\mathbf{x}, t, \mathbf{c})$.

Section 2 contains a probabilistic formulation of advection–reaction PDEs with uncertain coefficients. Corresponding raw (stochastic) and averaged (deterministic) CDF equations are derived in §3. Numerical algorithms for solving these equations are presented in §4. In §5, we analyse the accuracy of the CDF equations based on a large-eddy-diffusivity (LED) closure approximation by comparing their solutions with those computed in a Monte Carlo framework.

2. Problem formulation

Consider a hyperbolic conservation law (1.1) in which a state variable $c(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^+$ is advected by a linear flux $\mathbf{J}(\mathbf{x}, t, c) \equiv \mathbf{v}(\mathbf{x})c(\mathbf{x}, t)$,

$$\frac{\partial c}{\partial t} + \nabla_{\mathbf{x}} \cdot (\mathbf{v}c) = f(\mathbf{x}, t, c), \quad \mathbf{x} \in D \subset \mathbb{R}^d. \tag{2.1}$$

Here $\mathbf{v}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^d$ is assumed, without the loss of generality, to be divergence free ($\nabla \cdot \mathbf{v} = 0$); ARE (2.1) is defined on domain D; and the source term $f(\mathbf{x}, t, c)$ is parametrized with M time-invariant parameters $\boldsymbol{\kappa}(\mathbf{x}) \equiv \{\kappa_1(\mathbf{x}), \ldots, \kappa_M(\mathbf{x})\} : \mathbb{R}^d \to \mathbb{R}^M$. ARE (2.1) is subjected to an initial condition

$$c(\mathbf{x}, 0) = C_0(\mathbf{x}), \quad \mathbf{x} \in D, \tag{2.2}$$

where $C_0(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^+$ denotes the initial distribution (at time $t = 0$) of the state variable $c(\mathbf{x}, t)$. If the domain D is bounded, (2.1) is supplemented with boundary conditions for $c(\mathbf{x}, t)$ and/or its normal gradient along the boundary $\mathbf{x} \in \partial D$. To be concrete, we consider the nonlinear source term $f(\mathbf{x}, t, c)$ in the form

$$f(\mathbf{x}, t, c) = \kappa(\mathbf{x}) f_\alpha(c), \quad \kappa(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^+, \quad f_\alpha(c) : \mathbb{R}^+ \to \mathbb{R}. \tag{2.3}$$

Examples of such nonlinear terms include $f_\alpha = c(C_{\mathrm{eq}} - c)$ and $f_\alpha = \alpha(C_{\mathrm{eq}}^\alpha - c^\alpha)$ encountered in models of population dynamics [8] and transport in porous media [27], respectively. The latter case, which we use as a computational example, represents an effective (Darcy-scale) description of a heterogeneous reaction that causes solute concentration $c(\mathbf{x}, t)$ to change from its initial value C_0 to equilibrium concentration C_{eq}. In this description, α denotes the solute's stoichiometric coefficient and the reaction rate κ is determined by both a kinetic rate constant and a specific reactive surface. The resulting model (2.1)–(2.3) is an approximation in that it is applicable to kinetically controlled reactive processes in which both diffusion and local dispersion are negligible [11, §2.6, and references therein].

In a typical application, the spatially varying coefficients $\mathbf{v}(\mathbf{x})$ and $\kappa(\mathbf{x})$ are sampled at a few locations throughout their domain of definition D and their measurements might be

corrupted by experimental and/or interpretive errors. Additionally, the initial distribution $C_0(\mathbf{x})$ and boundary functions are often uncertain. Uncertainty about parameter values at points $\mathbf{x} \in D$ where measurements are not available can be quantified by treating these parameters as random fields, whose ensemble statistics are inferred from available data. Thus, the uncertain functions $\mathbf{v}(\mathbf{x})$, $\kappa(\mathbf{x})$ and $C_0(\mathbf{x})$ are replaced with corresponding (possibly cross-correlated) random fields $\mathbf{v}(\mathbf{x}, \omega_v)$, $\kappa(\mathbf{x}, \omega_\kappa)$ and $C_0(\mathbf{x}, \omega_c)$. Here, $\omega_a \in \Omega_a$ $(a = v, \kappa, c)$ denotes a random realization drawn from a complete probability space $(\Omega_a, \mathcal{A}_a, \mathcal{P}_a)$, whose event space Ω_a generates its σ-algebra $\mathcal{A}_a \subset 2^{\Omega_a}$ and is characterized by a probability measure $\mathcal{P}_a$.

The ensemble means of the input parameter fields, $\langle \mathbf{v} \rangle \equiv \mathbb{E}[\mathbf{v}]$ and $\langle \kappa \rangle \equiv \mathbb{E}[\kappa]$, serve to non-dimensionalize (2.1)–(2.3). Let L and V denote a characteristic length (e.g. a typical size of the domain D) and a characteristic velocity (e.g. the magnitude of a spatially averaged $\langle \mathbf{v} \rangle$), respectively. Then non-dimensional independent variables and input parameters, including the Damköhler number Da (the ratio of the advection and reaction time scales), are defined as

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{L}, \quad \hat{t} = \frac{tV}{L}, \quad \hat{\mathbf{v}} = \frac{\mathbf{v}}{V}, \quad \hat{\kappa} = \frac{\kappa}{\langle \kappa \rangle}, \quad \hat{C}_0 = \frac{C_0}{C_{\mathrm{eq}}} \quad \text{and} \quad \mathrm{Da} = \frac{\langle \kappa \rangle L C_{\mathrm{eq}}^{\alpha-1}}{V}. \tag{2.4}$$

The non-dimensional state variable $\hat{c} = c/C_{\mathrm{eq}}$ satisfies a rescaled ARE

$$\frac{\partial \hat{c}}{\partial \hat{t}} + \nabla_{\hat{\mathbf{x}}} \cdot (\hat{\mathbf{v}} \hat{c}) = \hat{f}(\hat{\mathbf{x}}, \hat{t}, \hat{c}), \quad \hat{f} = \mathrm{Da}\, \hat{\kappa} \hat{f}_\alpha, \quad \hat{f}_\alpha = \alpha(1 - \hat{c}^\alpha), \quad \hat{\mathbf{x}} \in \hat{D} \subset \mathbb{R}^d, \tag{2.5}$$

subjected to the non-dimensionalized initial condition

$$\hat{c}(\hat{\mathbf{x}}, 0) = \hat{C}_0(\hat{\mathbf{x}}), \quad \hat{\mathbf{x}} \in \hat{D} \tag{2.6}$$

and the appropriate boundary condition on $\partial \hat{D}$. In the following, we drop the hats identifying dimensionless quantities.

A solution of (2.5) and (2.6) is given in terms of a PDF or, equivalently, a CDF of the random state variable $c(\mathbf{x}, t, \omega)$. Let $p(C; \mathbf{x}, t) : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^+$ and $F(C; \mathbf{x}, t) : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^+$ denote, respectively, a single-point PDF and CDF of the state variable c at a point $(\mathbf{x}, t)$. Our goal is to derive a deterministic boundary value problem (BVP) for the CDF $F(C; \mathbf{x}, t)$. This is done for a general form of the source term $f(\mathbf{x}, t, c)$, while the computational examples are presented for $f(\mathbf{x}, t, c) = \mathrm{Da}\, \alpha \kappa(\mathbf{x}, \omega)(1 - \hat{c}^\alpha)$.

3. Cumulative distribution function method

Let $\mathcal{H}(\cdot) : \mathbb{R} \to \mathbb{R}$ denote the Heaviside step function, and consider a function $\Pi(c, C; \mathbf{x}, t) : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$,

$$\Pi(c, C; \mathbf{x}, t) \equiv \mathcal{H}[c(\mathbf{x}, t) - C]. \tag{3.1}$$

The ensemble mean of Π over all possible realizations of c at a point $(\mathbf{x}, t)$ is the single-point CDF,

$$\langle \Pi(c, C; \mathbf{x}, t) \rangle \equiv \int_0^\infty \mathcal{H}(\mathcal{C} - C) p(\mathcal{C}; \mathbf{x}, t)\, \mathrm{d}\mathcal{C} = 1 - F(C; \mathbf{x}, t). \tag{3.2}$$

An equation for $F(C; \mathbf{x}, t)$ is derived by noting that

$$\frac{\partial \Pi}{\partial t} = \frac{\partial \Pi}{\partial c} \frac{\partial c}{\partial t} = -\frac{\partial \Pi}{\partial C} \frac{\partial c}{\partial t}, \quad \nabla \cdot (\mathbf{v} \Pi) = \frac{\partial \mathbf{v} \Pi}{\partial c} \cdot \nabla c = -\frac{\partial \Pi}{\partial C} \mathbf{v} \cdot \nabla c \tag{3.3a}$$

and

$$f(\cdot, \cdot, c) \frac{\partial \Pi}{\partial C} = -f(\cdot, \cdot, c) \delta(c - C) \equiv -f(\cdot, \cdot, C) \delta(c - C) = f(\cdot, \cdot, C) \frac{\partial \Pi}{\partial C}. \tag{3.3b}$$

Multiplying (2.5) with $\partial \Pi / \partial C$ and accounting for (3.3) yields

$$\frac{\partial \Pi}{\partial t} + \nabla \cdot (\mathbf{v} \Pi) = -f(C) \frac{\partial \Pi}{\partial C}. \tag{3.4}$$

Let us introduce a four-dimensional space $\tilde{\mathbf{x}} = (x_1, x_2, x_3, x_4 \equiv C)^{\mathrm{T}}$ in which the gradient and 'velocity' are defined as

$$\tilde{\nabla} \equiv \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}, \frac{\partial}{\partial x_4} \equiv \frac{\partial}{\partial C} \right)^{\mathrm{T}} \quad \text{and} \quad \tilde{\mathbf{v}} = (v_1, v_2, v_3, v_4 \equiv f)^{\mathrm{T}}. \tag{3.5}$$

Then (3.4) can be written in the form of an advection equation

$$\frac{\partial \Pi}{\partial t} = -\tilde{\mathbf{v}} \cdot \tilde{\nabla} \Pi \quad \text{and} \quad \tilde{\mathbf{x}} \in \tilde{D} \equiv D \times (C_0, 1), \tag{3.6}$$

which describes the advection of Π in the random velocity field $\tilde{\mathbf{v}}$. Two comments regarding the nature of this advection equation are in order. First, the velocity field $\tilde{\mathbf{v}}$ is no longer divergence free; instead, $\tilde{\nabla} \cdot \tilde{\mathbf{v}} = \mathrm{Da}\, \mathrm{d}f_\kappa / \mathrm{d}x_4$. Second, $\Pi(c, x_4; \mathbf{x}, t)$ is defined on the bounded segment $C_0 \leq c(\mathbf{x}, t; \omega) \leq 1$ and $C_0 \leq x_4 \leq 1$. As written, $C_0 < 1$; the case of $C_0 > 1$ can be treated identically.

Stochastic averaging of advective transport equations with random velocity has been the subject of numerous studies (e.g. [9,10] and references therein). Most approaches start by employing the Reynolds decomposition, $\mathcal{A} = \langle \mathcal{A} \rangle + \mathcal{A}'$, to represent the random quantities in (3.6) as the sum of their ensemble means $\langle \mathcal{A} \rangle$ and zero-mean fluctuations about the mean $\mathcal{A}'$. Then, taking the ensemble average of (3.6) yields an unclosed CDF equation

$$\frac{\partial F}{\partial t} = -\langle \tilde{\mathbf{v}} \rangle \cdot \tilde{\nabla} F + \langle \tilde{\mathbf{v}}' \cdot \tilde{\nabla} \Pi' \rangle. \tag{3.7}$$

This equation contains the unknown cross-correlation term $Q = \langle \tilde{\mathbf{x}}' \cdot \tilde{\nabla} \Pi' \rangle$ and, hence, requires a closure approximation. A phenomenological closure described below gives rise to a diffusive term that is referred to as macrodispersion [29].

(a) Large-eddy-diffusivity approximation

Following [9] and others, we demonstrate in appendix A that the cross-correlation $Q(\tilde{\mathbf{v}}, t)$ satisfies exactly a non-local (unclosed) equation

$$\begin{aligned} Q(\tilde{\mathbf{x}}, t) = &\int_0^t \int_{\tilde{D}} \tilde{\nabla} F(\tilde{\mathbf{y}}, \tau) \cdot \langle \tilde{\mathbf{v}}'(\tilde{\mathbf{y}}) [\tilde{\mathbf{v}}'(\tilde{\mathbf{x}}) \cdot \tilde{\nabla}_{\tilde{\mathbf{x}}} \mathcal{G}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau)] \rangle \, \mathrm{d}\tilde{\mathbf{y}} \, \mathrm{d}\tau \\ &+ \int_0^t \int_{\tilde{D}} \langle \tilde{\mathbf{v}}'(\tilde{\mathbf{x}}) \cdot \tilde{\nabla}_{\tilde{\mathbf{x}}} \mathcal{G}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau) \rangle Q(\tilde{\mathbf{y}}, \tau) \, \mathrm{d}\tilde{\mathbf{y}} \, \mathrm{d}\tau. \end{aligned} \tag{3.8}$$

Here, $\mathcal{G}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau)$ is the random Green's function for (3.6) defined as a solution of

$$\frac{\partial \mathcal{G}}{\partial \tau} + \tilde{\nabla}_{\tilde{\mathbf{y}}} \cdot (\tilde{\mathbf{v}} \mathcal{G}) = -\delta(\tilde{\mathbf{x}} - \tilde{\mathbf{y}}) \delta(t - \tau), \tag{3.9}$$

subjected to the homogeneous initial and boundary conditions. Evaluating the kernel and source term in integral equation (3.8) requires a closure. Here, we employ the so-called LED closure, which replaces the random Green's function $\mathcal{G}$ with its 'mean-field approximation' G. The latter is given by a solution of (3.9) with the average velocity $\langle \tilde{\mathbf{v}} \rangle$ used in place of its random counterpart $\tilde{\mathbf{v}}$. Assuming that the CDF gradient $\tilde{\nabla} F(\tilde{\mathbf{y}}, t)$ varies slowly in space and time, we obtain a closed-form expression for the cross-correlation term Q (appendix A),

$$\begin{aligned} Q(\tilde{\mathbf{x}}, t) \approx &\frac{\partial}{\partial \tilde{x}_i} \left[\frac{\partial F(\tilde{\mathbf{x}})}{\partial \tilde{x}_j} \int_0^t \int_{\tilde{D}} G \langle \tilde{v}_j'(\tilde{\mathbf{y}}) \tilde{v}_i'(\tilde{\mathbf{x}}) \rangle \, \mathrm{d}\tilde{\mathbf{y}} \, \mathrm{d}\tau \right] \\ &- \frac{\partial F(\tilde{\mathbf{x}})}{\partial \tilde{x}_j} \int_0^t \int_{\tilde{D}} G \left\langle \tilde{v}_j'(\tilde{\mathbf{y}}) \frac{\partial \tilde{v}_i'(\tilde{\mathbf{x}})}{\partial \tilde{x}_i} \right\rangle \mathrm{d}\tilde{\mathbf{y}} \, \mathrm{d}\tau, \end{aligned} \tag{3.10}$$

where the Einstein notation is used to indicate summation over repeated indices.

Substituting (3.10) into (3.7) yields a closed CDF equation,

$$\frac{\partial F}{\partial t} = -\tilde{u}_i \frac{\partial F}{\partial \tilde{x}_i} + \frac{\partial}{\partial \tilde{x}_i}\left(\tilde{D}_{ij}\frac{\partial F}{\partial \tilde{x}_j}\right), \quad i,j = 1,\ldots,d+1 \tag{3.11}$$

in which the components of the macrodispersion tensor $\tilde{\mathbf{D}}$ and drift velocity $\tilde{\mathbf{u}}$ are given by

$$\tilde{D}_{ij}(\tilde{\mathbf{x}},t) \approx \int_0^t \int_{\tilde{D}} G(\tilde{\mathbf{x}},\tilde{\mathbf{y}},t-\tau)\langle \tilde{v}_i'(\tilde{\mathbf{x}})\tilde{v}_j'(\tilde{\mathbf{y}})\rangle \,\mathrm{d}\tilde{\mathbf{y}}\,\mathrm{d}\tau \tag{3.12a}$$

and

$$\tilde{u}_i(\tilde{\mathbf{x}},t) \approx \langle \tilde{v}_i(\tilde{\mathbf{x}})\rangle + \mathrm{Da}\int_0^t \int_{\tilde{D}} G(\tilde{\mathbf{x}},\tilde{\mathbf{y}},t-\tau)\langle \tilde{v}_i'(\tilde{\mathbf{y}})\kappa'(\mathbf{x})\rangle \frac{\mathrm{d}f_\alpha(x_4)}{\mathrm{d}x_4}\,\mathrm{d}\tilde{\mathbf{y}}\,\mathrm{d}\tau. \tag{3.12b}$$

Note that the coefficients $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{u}}$ are expressed in terms of (cross-)correlations of the input parameters and, hence, are computable. If the input parameters $\mathbf{v}(\mathbf{x},\omega)$ and $\kappa(\mathbf{x},\omega)$ are mutually uncorrelated, the covariance tensor simplifies to

$$\langle \tilde{v}_i'(\tilde{\mathbf{x}})\tilde{v}_j'(\tilde{\mathbf{y}})\rangle = \langle v_i'(\mathbf{x})v_j'(\mathbf{y})\rangle \equiv C_{v_{ij}}(\mathbf{x},\mathbf{y}), \quad i,j = 1,2,3 \tag{3.13a}$$

and

$$\langle \tilde{v}_4'(\tilde{\mathbf{x}})\tilde{v}_4'(\tilde{\mathbf{y}})\rangle = \mathrm{Da}^2\, C_\kappa(\mathbf{x},\mathbf{y}) f_\alpha(x_4) f_\alpha(y_4), \tag{3.13b}$$

where $C_\kappa(\mathbf{x},\mathbf{y})$ represents the two-point covariance of the random coefficient $\kappa(\mathbf{x},\omega)$.

(b) Auxiliary conditions for the cumulative distribution function equation

The initial condition for $F(\tilde{\mathbf{x}},t)$,

$$F(C;\mathbf{x},t=0) = F_{\mathrm{in}}(C;\mathbf{x}), \tag{3.14}$$

depends on the degree of uncertainty about $C_0(\mathbf{x})$ in (2.6). If $C_0(\mathbf{x})$ is known with certainty, then

$$F(C;\mathbf{x},0) = F_{\mathrm{in}}(C;\mathbf{x}) = \int_{C_0}^{C} \delta[\mathcal{C} - C_0(\mathbf{x})]\,\mathrm{d}\mathcal{C} = \mathcal{H}[C - C_0(\mathbf{x})]. \tag{3.15}$$

If $C_0(\mathbf{x})$ is uncertain, then $F_{\mathrm{in}}(C;\mathbf{x})$ is given by the prescribed CDF of $C_0(\mathbf{x},\omega)$.

Boundary conditions for $F(C;\mathbf{x},t)$ along the boundary ∂D of the physical domain D are given by CDFs of the boundary conditions for $c(\mathbf{x},t;\omega)$ at $\mathbf{x}\in\partial D$. Boundary conditions for $F(C;\mathbf{x},t)$ at $C = C_0$ and $C = 1$ are uniquely defined by the definition of a CDF,

$$F(C_0;\mathbf{x},t) = 0 \quad \text{and} \quad F(1;\mathbf{x},t) = 1. \tag{3.16}$$

This is in contrast to PDF equations (e.g. [29]) for which the specification of boundary conditions at $C = C_0$ and $C = 1$ is more ambiguous, except in a few special cases. For example, one can postulate that the PDF $p(C;\mathbf{x},t) = \partial F(C;\mathbf{x},t)/\partial C$ has zero gradient at the boundary surfaces $C = C_0$ and $C = 1$, i.e. prescribe the boundary condition $\partial p(C;\mathbf{x},t)/\partial C = 0$ at $C = C_0$ and 1 for all $(\mathbf{x},t)$. Alternatively, one can extend the domain of definition of the random state variable $c(\mathbf{x},t;\omega)$ into an unphysical range, e.g. $-\infty < c < +\infty$, and specify the boundary conditions for the PDF as $p(C\to\pm\infty;\mathbf{x},t) = 0$.

4. Computational examples

In order to quantify the accuracy and robustness of CDF equation (3.11), i.e. to assess the validity of the LED closure, we compare numerical solutions of (3.11) with the CDF computed via high-resolution Monte Carlo numerical simulations of (3.6).

(a) Test cases

Consider a one-dimensional version of (2.5) with $f_\alpha = \alpha(1 - x_4^\alpha)$ defined on the semi-infinite domain $D = \mathbb{R}^+$. We treat the uncertain coefficients $v(x)$ and $\kappa(x)$ as mutually uncorrelated multivariate lognormal random fields $v(x, \omega)$ and $\kappa(x, \omega)$, with $\langle v \rangle = 1$ and exponential covariance functions $C_v(x, y)$ and $C_\kappa(x, y)$, respectively. The one-dimensional version of (2.5) is subjected to the initial and boundary conditions

$$c(x, 0) = 0, \quad c(x = 0, t) = 0 \quad \text{and} \quad c(x \to \infty, t) = 1. \tag{4.1}$$

We analyse both linear ($\alpha = 1$) and nonlinear ($\alpha = 2$) sources f_α.

(b) Cumulative distribution function equations

For the problem under consideration, CDF equation (3.11) reduces to

$$\frac{\partial F}{\partial t} = -\frac{\partial F}{\partial x} - u_4 \frac{\partial F}{\partial x_4} + \frac{\partial}{\partial x_1}\left(D_{11}\frac{\partial F}{\partial x_1}\right) + \frac{\partial}{\partial x_4}\left(D_{44}\frac{\partial F}{\partial x_4}\right) \tag{4.2}$$

with

$$u_4(\tilde{\mathbf{x}}, t) = \mathrm{Da} f_\alpha(x_4) + \mathrm{Da}^2 \frac{\mathrm{d}f_\alpha(x_4)}{\mathrm{d}x_4} \int_0^t \int_{\tilde{D}} C_\kappa(x, y) f_\alpha(y_4) G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T)\, \mathrm{d}\tilde{\mathbf{y}}\, \mathrm{d}T, \tag{4.3a}$$

$$D_{11}(\tilde{\mathbf{x}}, t) = \int_0^t \int_{\tilde{D}} G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T) C_v(x, x - T)\, \mathrm{d}\tilde{\mathbf{y}}\, \mathrm{d}T \tag{4.3b}$$

and

$$D_{44}(\tilde{\mathbf{x}}, t) = \mathrm{Da}^2 f_\alpha(x_4) \int_0^t \int_{\tilde{D}} G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T) C_\kappa(x, y) f_\alpha(y_4)\, \mathrm{d}\tilde{\mathbf{y}}\, \mathrm{d}T. \tag{4.3c}$$

CDF equation (4.2) is subjected to the initial condition

$$F(\tilde{\mathbf{x}}, t = 0) = 1, \quad \tilde{\mathbf{x}} \in \tilde{D}, \tag{4.4}$$

and the boundary conditions

$$F(x_4; x = 0, t) = 1 \quad \text{and} \quad F(x_4; x \to \infty, t) = 0, \quad x_4 \in (0, 1), \quad t > 0 \tag{4.5a}$$

and

$$F(x_4 = 0; x, t) = 0, \quad F(x_4 = 1; x, t) = 1, \quad x \in D, \quad t > 0. \tag{4.5b}$$

For operational reasons, the boundary condition at $x \to \infty$ is replaced by $\partial F / \partial x = 0$.

The BVP (4.2)–(4.5) is solved with an explicit higher order finite difference method for hyperbolic equations with variable coefficients [30]; the diffusive operator is handled with a semi-implicit Crank–Nicolson algorithm.

(c) Monte Carlo simulations

No approximations have been made to derive stochastic equation (3.6), which governs the dynamics of Π. Consequently, we use this equation, rather than original equation (2.5), in our MCSs. This is advantageous for two reasons. First, (3.6) is always linear even if (2.5) is highly nonlinear. Second, the ensemble mean of the solution to (3.6) yields the full CDF of the solution to (2.5), $\langle \Pi \rangle = 1 - F$.

For the problem under consideration, we perform MCSs on the two-dimensional version of (3.6),

$$\frac{\partial \Pi}{\partial t} + v(x,\omega)\frac{\partial \Pi}{\partial x} + \mathrm{Da}\,\alpha\kappa(x,\omega)(1-x_4^{\alpha})\frac{\partial \Pi}{\partial x_4} = 0, \tag{4.6}$$

subjected to the initial and boundary conditions

$$\Pi(x, x_4, t=0) = 0, \quad \Pi(x=0, x_4, t) = 0 \quad \text{and} \quad \Pi(x, x_4=0, t) = 1. \tag{4.7}$$

To solve this hyperbolic BVP with discontinuous initial conditions by means of MCSs, we generate multiple realizations of characteristic lines along which the solution is determined semi-analytically. Then, the single-realization results are averaged to compute $F = 1 - \langle \Pi \rangle$. A set of Monte Carlo realizations are characterized by a probability distribution of the uncertain parameters $\kappa(x)$ and $v(x)$; each realization represents a sample with given multi-variate statistics. To properly capture the spatial variability of these fields, we construct a mesh which consists of four or five equally spaced nodes per correlation length [18].

As before, we use the notation $\tilde{\mathbf{x}} = (x, x_4)^{\top}$ and $\tilde{\mathbf{v}} = [v, \mathrm{Da}\,\alpha\kappa(x,\omega)(1-x_4^{\alpha})]^{\top}$. Then equations of characteristics for (4.6), $\mathrm{d}\tilde{\mathbf{x}}/\mathrm{d}t = \tilde{\mathbf{v}}$, take the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v(x), \quad x(t=0) = \xi \quad \text{and} \quad \frac{\mathrm{d}x_4}{\mathrm{d}t} = \mathrm{Da}\,\alpha\kappa(x)(1-x_4^{\alpha}), \quad x_4(t=0) = \eta. \tag{4.8}$$

Along these characteristics,

$$\frac{\mathrm{d}\Pi}{\mathrm{d}t} = 0, \quad \Pi(t=0) = 0. \tag{4.9}$$

The characteristic lines propagate boundary or initial conditions in the three-dimensional semi-infinite space $\tilde{D} \times \mathbb{R}^{+}$. Their random realizations correspond to that of the coefficients $\kappa(x;\omega)$ and $v(x;\omega)$.

To facilitate numerical integration, we rewrite (4.8) as

$$t = \int_0^x \frac{\mathrm{d}s}{v(s;\omega)} - \int_0^{\xi} \frac{\mathrm{d}s}{v(s;\omega)}, \quad \frac{\mathrm{d}x_4}{1-x_4^{\alpha}} = \mathrm{Da}\,\alpha\frac{\kappa(x;\omega)}{v(x;\omega)}\,\mathrm{d}x, \tag{4.10a}$$

subjected to

$$x_4(t=0) = x_4(x=\xi) = \eta. \tag{4.10b}$$

Along the characteristics that cross the $x=0$ boundary, (4.9) takes the form

$$\frac{\mathrm{d}\Pi}{\mathrm{d}t} = 0, \quad \Pi(x=0) = 0, \tag{4.11}$$

while the characteristics crossing the $x_4 = 0$ boundary give rise to

$$\frac{\mathrm{d}\Pi}{\mathrm{d}t} = 0, \quad \Pi(x_4=0) = 1. \tag{4.12}$$

Equation (4.9) has the solution $\Pi(x, x_4, t) = \Pi(t=0) = 0$ whenever $(\xi, \eta) \in \tilde{D}$. Otherwise, if $x_4(x=0) = \eta \in (0,1)$, then $\Pi(x, x_4, t) = \Pi(x_4=0)$ in (4.12); else, the solution is $\Pi(x, x_4, t) = \Pi(x=0)$ in (4.11).

Finally, $F = 1 - \langle \Pi \rangle$ where the ensemble average $\langle \Pi \rangle$ is obtained by averaging over an adequate number of realizations N_{r}. In the present setting, the convergence was achieved with $N_{\mathrm{r}} = 10\,000$ for all parameter sets.

5. Results and discussion

In order to quantify the impact of the employed assumptions and the predictive capability of (3.11), we compare the CDFs $F(C;x,t)$ computed by solving the deterministic CDF equation and via MCSs. The comparison is performed for a number of different scenarios, which include both linear ($\alpha = 1$) and nonlinear ($\alpha = 2$) sources f_{α}, and different correlation models for the random coefficients $v(x;\omega)$ and $\kappa(x;\omega)$. The effective parameters for these different scenarios are summarized in appendix C.

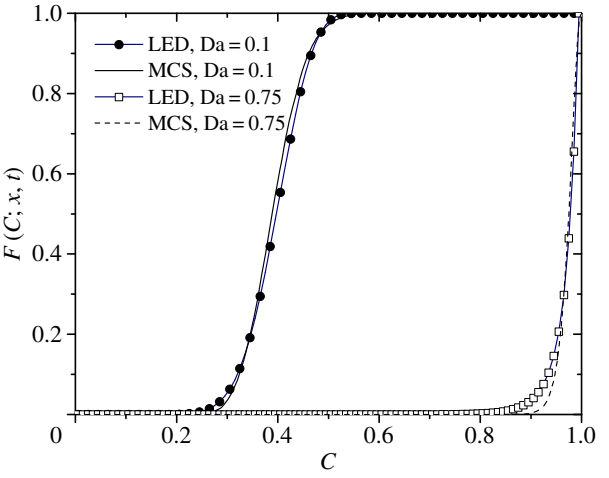

Figure 1. CDF profiles at the space–time point $(x, t) = (25.0, 5.0)$ computed with the LED closure and MCSs for the linear source ($\alpha = 1$), deterministic $v(x) \equiv 1$ and uncertain $\kappa(x)$. The latter is modelled as an exponentially correlated lognormal random field with variance $\sigma_\kappa^2 = 0.1$, correlation length $\lambda_\kappa = 1.0$, and the mean that corresponds to the Damköhler number Da = 0.1 or 0.75.

All results are presented as profiles of the CDF $F(C; x, t)$ over the whole range of $C \in [0, 1]$ at a specific space–time location (x, t). The theoretical considerations discussed above dictate that, for deterministic initial (C_0) and equilibrium ($C_{eq} = 1$) conditions, $F(C; x, t = 0) = \mathcal{H}(C - C_0) = \mathcal{H}(0)$ and $F(C; x, t \to \infty) = \mathcal{H}(C - 1) = \mathcal{H}(1)$. The rate of transition between these two deterministic states depends on the space–time location (x, t) and the degree of uncertainty in the input parameters $v(x)$ and $\kappa(x)$. Close to the initial and equilibrium states, the profiles of $F(C; x, t)$ are expected to be close to the Heaviside function, reflecting relatively low levels of predictive uncertainty. These profiles are expected to be smoother in intermediate regimes, reflecting increased degrees of predictive uncertainty.

(a) Sources of uncertainty: $\kappa(x)$

We start by considering deterministic $v(x) \equiv 1$ and treating $\kappa(x)$ as the sole source of uncertainty. If the random parameter $\kappa(x; \omega)$ is spatially uncorrelated (white noise), then the system state PDF, $p(C; \mathbf{x}, t) = \mathrm{d}F(C; \mathbf{x}, t)/\mathrm{d}C$, satisfies exactly the Fokker–Planck equation in [2] and the CDF equation (3.11) or (4.2) are exact. In other words, the cross-correlation term $Q(\tilde{\mathbf{x}}, t)$ defined by (3.8) does not require any closure approximation, yielding exactly the macrodispersion tensor $\tilde{\mathbf{D}}$ and drift velocity $\tilde{\mathbf{u}}$ in (3.12) or (4.3).

If the random parameter $\kappa(x; \omega)$ exhibits a spatial correlation, then the CDF equations (3.11) or (4.2) are based on the LED closure (3.12) or (4.5). Figure 1 demonstrates the accuracy of this approximation for $\kappa(x; \omega)$ with the exponential covariance function $C_\kappa(|x - y|) = \sigma_\kappa^2 \exp(-|x - y|/\lambda_\kappa)$ with $\sigma_\kappa^2 = 0.1$ and $\lambda_\kappa = 1.0$. The semi-analytical MCSs serve as an 'exact' solution in this comparison. Figure 1 also reveals that the CDF $F(C; x, t)$ is sensitive to the Damköhler number Da. The predictive uncertainty is higher for slow reactions (Da = 0.1) than for faster ones (Da = 0.75), as evidenced by the widths of the transition zone between $F = 0$ and $F = 1$. Higher values of Da not only increase the predictive uncertainty. They also accelerate the transition of the system towards its equilibrium state $F = \mathcal{H}(C - 1)$, resulting in the sharper CDF profiles shifted to the right (figure 1).

(b) Sources of uncertainty: $\kappa(x)$ and $v(x)$

In this section, we analyse the combined effect of uncertainty in the two input parameters, $v(x)$ and $\kappa(x)$, on uncertainty in predictions of the state variable $c(x, t)$. The two uncertain parameters a

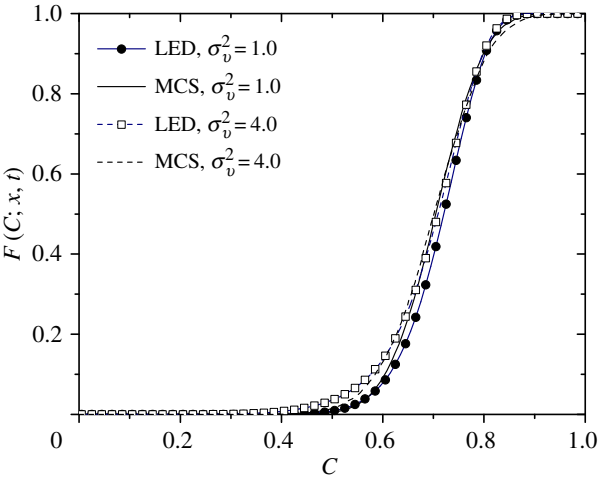

Figure 2. CDF profiles at the space–time point $(x, t) = (10.0, 5.0)$ computed with the LED closure and MCSs for the linear source ($\alpha = 1$) and uncertain $v(x)$ and $\kappa(x)$. The two parameters are modelled as exponentially correlated lognormal random fields with variances $\sigma_v^2 = 0.1$ or 4.0 and $\sigma_\kappa^2 = 0.1$, and correlation lengths $\lambda_v = \lambda_\kappa = 1.0$. The Damköhler number is set to $\mathrm{Da} = 0.25$.

modelled as mutually uncorrelated lognormal random fields, each of which is characterized by an exponential correlation function. Comparison of figures 2 and 3 reveals the relative importance of these two sources of uncertainty, as quantified by their respective variances σ_v^2 and σ_κ^2. For a fixed value of σ_κ^2, the predictive uncertainty captured by the CDF $F(C; x, t)$ is relatively insensitive to the magnitude of σ_v^2 (figure 2), with the fourfold increase in σ_v^2 translating into a slight increase in predictive uncertainty (the width of the transition zone between $F = 0$ and 1). Note that $F(C; x, t)$ displays an asymmetric behaviour with respect to the median, which indicates that the state variable $c(x, t; \omega)$ is non-Gaussian. This behaviour stems from the dependence of the effective diffusion coefficient D_{44} in (4.2) on the phase–space coordinate $x_4 \equiv C$. For the example under consideration, the non-monotonic dependence of D_{44} on x_4 is given explicitly by (C 5*b*). That gives rise to longer tails of the CDF profiles for small C, indicating a non-negligible probability of encountering small values of C owing to the presence of low-κ regions. The agreement between the CDFs computed with the LED closure and MCSs demonstrates the accuracy and robustness of the CDF equation (4.2) with respect to σ_v^2. Figure 2 also demonstrates that the predictive uncertainty (the CDF F) is relatively insensitive to the level of uncertainty in the input parameter $v(x)$, as quantified by its variance σ_v^2.

The predictive uncertainty is more sensitive to the level of uncertainty in the other input parameter, $\kappa(x)$ (figure 3). The magnitude of its variance, σ_κ^2, affects not only the CDF profile $F(C; x, t)$ but also the accuracy of the LED approximation that underpins CDF equation (4.2). This approximation remains accurate for σ_κ^2 up to 0.1. As σ_κ^2 increases, the LED closure underestimates the predictive uncertainty, i.e. leads to the CDF profiles that are sharper than their counterparts computed with the MCSs. These MCSs also demonstrate that the asymmetry of the CDF profiles, i.e. the non-Gaussianity of the state variable $c(x, t; \omega)$, increases with the level of uncertainty in the input parameter $\kappa(x)$ (as quantified by its variance σ_κ^2).

Another measure of uncertainty about the input parameter $\kappa(x)$ is λ_κ, the correlation length of the random field $\kappa(x; \omega)$. For the perfectly correlated (random constant) parameter $v(\omega)$, figure 4 shows both that predictive uncertainty (the width of the CDF profiles F) increases with λ_κ and that the accuracy of the LED approximation is relatively insensitive to the magnitude of λ_κ. It also demonstrates that the degree of non-Gaussianity of the state variable $c(x, t; \omega)$ increases with λ_κ.

It is worthwhile emphasizing here that the computational burden of solving the CDF BVP (4.2)–(4.5) is insensitive to the magnitude of λ_κ. This is in contrast to both MCSs and

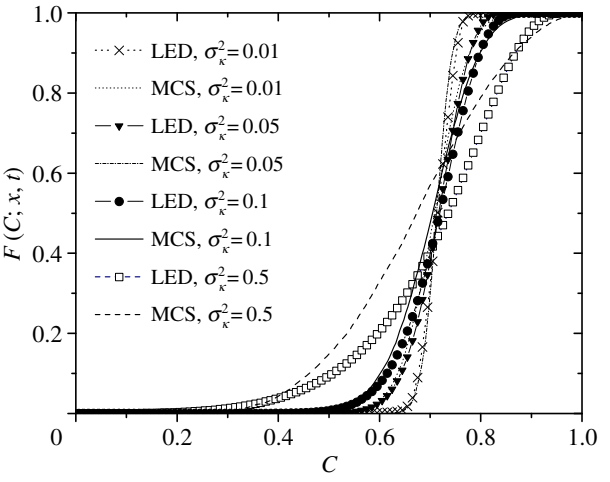

Figure 3. CDF profiles at the space–time point $(x, t) = (10.0, 5.0)$ computed with the LED closure and MCSs for the linear source ($\alpha = 1$) and uncertain $v(x)$ and $\kappa(x)$. The two parameters are modelled as exponentially correlated lognormal random fields with variances $\sigma_v^2 = 1.0$ and $\sigma_\kappa^2 = 0.01$, 0.05, 0.1 or 0.5, and correlation lengths $\lambda_v = \lambda_\kappa = 1.0$. The Damköhler number is set to $\mathrm{Da} = 0.25$.

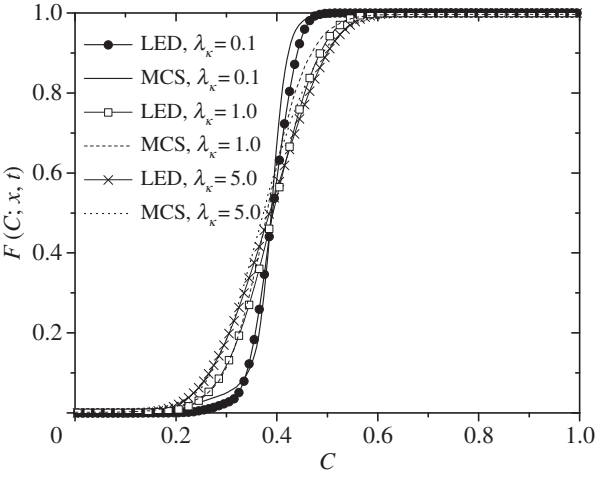

Figure 4. CDF profiles at the space–time point $(x, t) = (10.0, 5.0)$ computed with the LED closure and MCSs for the linear source ($\alpha = 1$), perfectly correlated (random constant) v and exponentially correlated $\kappa(x)$ with $\lambda_\kappa = 0.1$, 1.0 and 5.0. The two parameters have lognormal distributions with variances $\sigma_v^2 = 1.0$ and $\sigma_\kappa^2 = 0.1$. The Damköhler number is set to $\mathrm{Da} = 0.1$.

direct numerical algorithms (e.g. polynomial chaos expansions or stochastic collocation methods), whose computational cost increases as correlation lengths of the input parameters decrease. The cost of the former increases because at least five nodes per correlation length are required to accurately discretize a random input parameter. The latter methods rely on finite-term representations of random fields, e.g. Karhunen–Loève expansions, whose convergence rate decreases as the correlation lengths of the random fields become smaller.

(c) Effects of nonlinearity on predictive uncertainty

In the previous section, we analysed the impact of uncertain parameters $v(x)$ and $\kappa(x)$ on the predictive uncertainty of the linear conservation law expressed by the one-dimensional version of (2.5) and (2.6) with $\alpha = 1$. Here, we explore the effects of nonlinearity of the governing

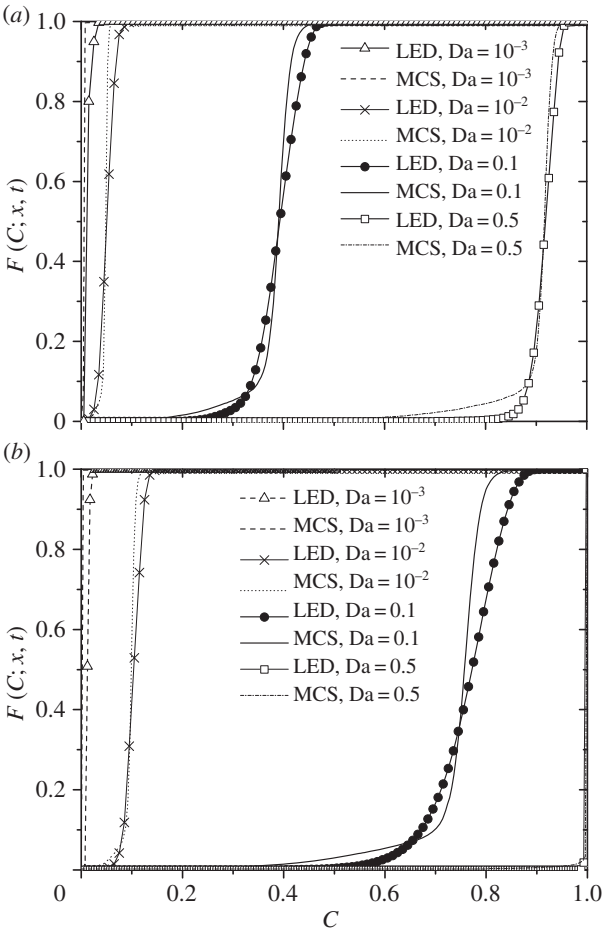

Figure 5. CDF profiles at the space–time point $(x, t) = (10.0, 5.0)$ computed with the LED closure and MCSs for (*a*) the linear ($\alpha = 1$) and (*b*) nonlinear ($\alpha = 2$) sources, perfectly correlated (random constant) v and uncorrelated $\kappa(x)$ with $\lambda_\kappa = 1.0$. The two parameters have lognormal distributions with variances $\sigma_v^2 = 1.0$ and $\sigma_\kappa^2 = 0.05$. (*a*) $\alpha = 1$, (*b*) $\alpha = 2$.

equation by setting $\alpha = 2$ in (2.5). The simulations reported below correspond to the perfectly correlated (random constant) parameter $v(\omega)$ and uncorrelated (white noise) parameter $\kappa(x;\omega)$. Both parameters are lognormal, with respective variances $\sigma_v^2 = 1.0$ and $\sigma_\kappa^2 = 0.1$.

Figure 5 exhibits the effect of the Damköhler number Da on predictive uncertainty (the CDF F) of the linear ($\alpha = 1$) and nonlinear ($\alpha = 2$) systems. In both cases, small Damköhler numbers ($\mathrm{Da} = 10^{-3}$) yield the sharp CDF profiles $F(C)$ centred around small C's, which indicates that the dimensionless state variable $c(x = 10.0, t = 5.0)$ is close to its deterministic initial value $C_0 = 0$. At the same space–time point $(x, t) = (10.0, 5.0)$, large Damköhler numbers ($\mathrm{Da} = 0.5$) give rise to the state variable c that approaches its deterministic equilibrium $C = 1$, resulting in the sharp CDF profiles centred around large C's. For intermediate Da, the state variable $c(x, t)$ is far from its deterministic bounds, giving rise to higher predictive uncertainty (wider CDF profiles). Comparison of figure 5*a*,*b* reveals that the nonlinearity accelerates the transition to equilibrium, shifting the CDF profiles to the right. The nonlinearity also increases the asymmetry (non-Gaussianity) of $F(C)$, enhancing the skewness of the CDF profiles towards low values of C. These long tails are reduced faster with time in the nonlinear than in the linear case (not shown here), as a result of a faster transition towards equilibrium. Another effect of the nonlinearity is to reduce the accuracy of the LED closure at intermediate values of the Damköhler number ($\mathrm{Da} = 0.1$

6. Summary and conclusion

We derived deterministic CDF equations that govern the evolution of CDFs of state variables whose dynamics are described by the first-order hyperbolic conservation laws (AREs) with uncertain input parameters. Uncertainty (randomness) in two input parameters, which parametrize the (linear) advective flux and (nonlinear) reactive terms in hyperbolic conservation laws, was considered. The CDF equations possess the following advantages over other statistical and stochastic approaches to uncertainty quantification.

(i) CDF equations are subjected to uniquely specified boundary conditions in the phase space: if a random state variable $c(\mathbf{x}, t; \omega)$ takes values on an interval $[C_0, C_{\mathrm{eq}}]$, its CDF $F(C; \mathbf{x}, t)$ satisfies the conditions $F(C_0; \mathbf{x}, t) = 0$ and $F(C_{\mathrm{eq}}; \mathbf{x}, t) = 1$. This is in contrast to PDF methods (e.g. [29]), which generally defy unique specification of boundary conditions for the PDF $p(C; \mathbf{x}, t)$ at $C = C_0$ and $C = 1$.

(ii) The computational burden of solving CDF equations is insensitive to the magnitude of the correlation lengths of random input parameters. This is in contrast to both MCSs and direct numerical algorithms (e.g. polynomial chaos expansions or stochastic collocation methods), whose computational cost increases as correlation lengths of the input parameters decrease. The cost of the former increases because at least five nodes per correlation length are required to accurately discretize a random input parameter. The latter methods rely on finite-term representations of random fields, e.g. Karhunen–Loève expansions, whose convergence rate decreases as their correlation lengths become smaller.

The LED closure, which underpins the presented CDF method, consists of two approximations: a perturbation expansion of the third ensemble moments and the assumption that ∇F varies slowly in space and time. To verify the accuracy and robustness of the LED closure, we conducted a set of numerical experiments which compared the CDFs computed with the CDF equations with those obtained via MCSs. This comparison leads to the following major conclusions.

(i) The CDF equations remain accurate over a wide range of statistical properties of the two input parameters, such as their correlation lengths and variance of the coefficient that parametrizes the advective flux.

(ii) The parameter that affects the performance of the LED closure is σ_κ^2, variance of the coefficient that parametrizes the (nonlinear) source term. The CDF equations remain accurate for $\sigma_\kappa^2 \leq 0.1$.

(iii) The order of the reaction does not seem to affect the predictive capabilities of the LED approach, with a good match for a wide range of Da since early times.

Funding statement. This work was supported in part by the Computational Mathematics Program of the Air Force Office of Scientific Research and by the National Science Foundation award EAR-1246315.

Appendix A. Large-eddy-diffusivity closure

The equation for the random fluctuations $\Pi' = \Pi - \langle \Pi \rangle$ can be obtained by subtracting the equation for $\langle \Pi \rangle$ (3.7) from the equation for Π (3.4), both written in terms of space–time coordinates $\tilde{\mathbf{y}}$ and τ, to obtain

$$\frac{\partial \Pi'}{\partial \tau} + \tilde{\mathbf{v}}(\tilde{\mathbf{y}}) \cdot \tilde{\nabla}_{\tilde{\mathbf{y}}} \Pi'(\tilde{\mathbf{y}}, \tau) = -\tilde{\mathbf{v}}'(\tilde{\mathbf{y}}) \cdot \tilde{\nabla}_{\tilde{\mathbf{y}}} \langle \Pi \rangle + \langle \tilde{\mathbf{v}}'(\tilde{\mathbf{y}}) \cdot \tilde{\nabla}_{\tilde{\mathbf{y}}} \Pi'(\tilde{\mathbf{y}}, \tau) \rangle. \tag{A 1}$$

Multiplying (A 1) with the random Green's function $\mathcal{G}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau)$ defined by (3.9) and the corresponding homogeneous initial and boundary conditions and integrating the left-hand side

of the resulting equation in space and time leads to

$$\Pi' = \int_{\tilde{D}} \Pi'_0 \mathcal{G}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t)\, \mathrm{d}\tilde{\mathbf{y}} - \int_0^t \int_{\tilde{\Gamma}} \tilde{v}_n \Pi' \mathcal{G}\, \mathrm{d}\tilde{\mathbf{s}}\, \mathrm{d}\tau - \int_0^t \int_{\tilde{D}} (\tilde{\mathbf{v}}' \cdot \tilde{\nabla} \langle \Pi \rangle - Q) \mathcal{G}\, \mathrm{d}\tilde{\mathbf{y}}\, \mathrm{d}\tau. \tag{A 2}$$

Here, $\Pi'_0(\tilde{\mathbf{y}}) = \Pi'(\tilde{\mathbf{y}}, \tau = 0)$, $Q = \langle \tilde{\mathbf{v}}' \cdot \tilde{\nabla} \Pi' \rangle$, $\tilde{v}_n = \tilde{\mathbf{n}} \cdot \tilde{\mathbf{v}}$ and $\tilde{\mathbf{n}}$ is the outward normal vector to the boundary $\tilde{\Gamma} \equiv \partial \tilde{D}$ of the hyper-domain $\tilde{D} \equiv D \times (C_0, 1)$.

Applying the differential operator $\tilde{\mathbf{v}}'(\tilde{\mathbf{x}}) \cdot \nabla_{\tilde{\mathbf{x}}}$ to both sides of (A 2) and taking the ensemble mean yields the expression for $Q(\mathbf{x}, t)$ in (3.8). To render this expression computable, we drop the last term in (3.8) because it is of lower order than the other terms. Then, we assume that $\tilde{\nabla} F$ varies slowly in time and space, which allows us to take the unknown outside the integrals. This approximation is valid at asymptotical times, when F is smooth. Finally, we replace the random Green's function $\mathcal{G}$ with its 'mean-field' (deterministic) counterpart G that is defined as a solution of (3.9) in which the mean $\langle \tilde{\mathbf{v}} \rangle$ is substituted for the random $\tilde{\mathbf{v}}$. This gives (3.10).

Appendix B. Green's functions

If the random parameter $\mathbf{v}(\mathbf{x}; \omega)$ is either statistically homogeneous, $\langle \mathbf{v}(\mathbf{x}; \omega) \rangle = \bar{\mathbf{v}}$, or divergence free, $\nabla \cdot \bar{\mathbf{v}}(\mathbf{x}) = 0$, then the deterministic Green's function $G(\tilde{\mathbf{x}}, \tilde{\mathbf{x}}, t - \tau)$ satisfies

$$\frac{\partial G}{\partial \tau} + \mathbf{v} \cdot \nabla_{\mathbf{y}} G + \mathrm{Da} \frac{\partial f_\alpha(y_4) G}{\partial y_4} = -\delta(\mathbf{x} - \mathbf{y}) \delta(x_4 - y_4) \delta(t - \tau), \quad \tau < t, \tag{B 1}$$

subjected to $G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, 0) = 0$ and $G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t < \tau) = 0$. A solution of (B 1) can be expressed as the product $G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T) = G_{Nd}(\mathbf{x}, \mathbf{y}, T) G_4(x_4, y_4, T)$. Here, $T = t - \tau$; $\mathbf{x}, \mathbf{y} \in D$; $x_4, y_4 \in (C_0, 1)$; and G_{Nd} and G_4 satisfy, respectively,

$$\frac{\partial G_{Nd}}{\partial T} - \bar{\mathbf{v}} \cdot \nabla_{\mathbf{y}} G_{Nd} = \delta(\mathbf{x} - \mathbf{y}) \delta(T), \quad T > 0; \quad G_{Nd}(\mathbf{x}, \mathbf{y}, T = 0) = 0 \tag{B 2}$$

and

$$\frac{\partial G_4}{\partial T} - \mathrm{Da} \frac{\partial f_\alpha(y_4) G}{\partial y_4} = \delta(x_4 - y_4) \delta(T), \quad T > 0; \quad G_4(x_4, y_4, T = 0) = 0. \tag{B 3}$$

Solving (B 2) with the method of characteristics gives

$$G_{Nd}(\mathbf{x}, \mathbf{y}, T) = \delta(\mathbf{x} - \mathbf{y} - \mathbf{v}T), \quad T > 0. \tag{B 4}$$

For $f_\alpha(y_4) = -\alpha(y_4^\alpha - 1)$, the characteristics of (B 3) satisfy a Cauchy problem

$$\frac{\mathrm{d}y_4}{\mathrm{d}T} = -\mathrm{Da} f_\alpha = \alpha \mathrm{Da}(y_4^\alpha - 1), \quad y_4(0) = \xi, \tag{B 5}$$

whose general solution is

$${}_2F_1\left[\frac{1}{\alpha}, 1; 1 + \frac{1}{\alpha}, y_4^\alpha\right] y_4 - {}_2F_1\left[\frac{1}{\alpha}, 1; 1 + \frac{1}{\alpha}, \xi^\alpha\right] \xi = \alpha\, \mathrm{Da}\, T, \tag{B 6}$$

where ${}_2F_1$ is the hypergeometric function. For $\alpha = 1$ and 2, (B 6) reduces to closed-form expressions for the characteristic

$$y_4(T) = (\xi - 1) \mathrm{e}^{\mathrm{Da}\, T} + 1 \tag{B 7}$$

and

$$y_4(\xi, T) = 2 \left(1 - \frac{\xi - 1}{\xi + 1} \mathrm{e}^{4\mathrm{Da}\, T}\right)^{-1} - 1, \tag{B 8}$$

respectively. Along these characteristics

$$\frac{\mathrm{d}G_4}{\mathrm{d}T} - \mathrm{Da} \frac{\partial f_\alpha}{\partial y_4} G_4 = \delta[x_4 - y_4(\xi, T)] \delta(T), \quad T > 0; \quad G_4(x_4, \xi, 0) = 0. \tag{B 9}$$

As $\delta[x_4 - y_4(\xi, T)]\delta(T) = \delta[x_4 - y_4(\xi, 0)]\delta(T) = \delta(x_4 - \xi)\delta(T)$, the solution of (B 9) is

$$G_4(x_4, \xi, T) = \exp\left(-\mathrm{Da}\int_0^T \frac{\partial f_\alpha}{\partial y_4}\mathrm{d}t\right)\mathcal{H}(T)\delta(x_4 - \xi) = \mathcal{H}(T)\frac{f_\alpha(\xi)}{f_\alpha(y_4)}\delta[x_4(\tau) - \xi]$$
$$= \mathcal{H}(T)\frac{f_\alpha(x_4)}{f_\alpha(y_4)}\delta[x_4(\tau) - \xi]. \tag{B 10}$$

Green's function $G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau)$ is obtained by combining (B 4) and (B 10),

$$G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, t - \tau) = \mathcal{H}(T)\delta(\mathbf{x} - \mathbf{y} - \bar{\mathbf{v}}T)\frac{f_\alpha(\xi)}{f_\alpha(y_4)}\delta[x_4(\tau) - \xi]. \tag{B 11}$$

To eliminate ξ, we use either (B 7) if $\alpha = 1$ or (B 8) if $\alpha = 2$. In $N = 1$ spatial dimensions and for $\bar{v} = 1$, this yields

$$G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T) = \mathcal{H}(T)\delta(x - y - T)\mathrm{e}^{-\mathrm{Da}\,T}\delta[x_4 - (y_4 - 1)\mathrm{e}^{-\mathrm{Da}\,T} - 1] \quad \text{if } \alpha = 1 \tag{B 12}$$

and

$$G(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, T) = \mathcal{H}(T)\delta(x - y - T)\frac{4\mathrm{e}^{4\mathrm{Da}\,T}}{[\mathrm{e}^{4\mathrm{Da}\,T}(y_4 + 1) - (y_4 - 1)]^2}$$
$$\times\,\delta\left(x_4 - \frac{\mathrm{e}^{4\mathrm{Da}\,T}(y_4 + 1) + (y_4 - 1)}{\mathrm{e}^{4\mathrm{Da}\,T}(y_4 + 1) - (y_4 - 1)}\right) \quad \text{if } \alpha = 2. \tag{B 13}$$

Appendix C. Coefficients in cumulative distribution function equation (4.2)

Here, we compute the coefficients u_4, D_{11} and D_{44} given by the general expressions (4.3) for the several examples considered in §5.

(a) Sources of uncertainty: $\kappa(x)$

If the parameter v is deterministic, then $D_{11} = 0$. For a given $C_\kappa(x, y)$ and $\alpha = 1$ or 2, analytical expressions for D_{44} and u_4 are computed as follows.

(i) Linear reaction law ($\alpha = 1$)

Green's function G is given by (B 12), and u_4 and D_{44} in (4.3) reduce to

$$u_4 = \mathrm{Da}(1 - x_4) - \mathrm{Da}^2(1 - x_4)\int_0^{t_*} C_\kappa(x, x - T)\mathrm{e}^{\mathrm{Da}\,T}\,\mathrm{d}T \tag{C 1a}$$

and

$$D_{44} = \mathrm{Da}^2(x_4 - 1)^2\int_0^{t_*} C_\kappa(x, x - T)\mathrm{e}^{\mathrm{Da}\,T}\,\mathrm{d}T, \tag{C 1b}$$

where

$$t_* = \min\left\{t, \frac{1}{\mathrm{Da}}\ln\left(\frac{1}{1 - x_4}\right), x\right\}. \tag{C 2}$$

If $\kappa(x; \omega)$ is uncorrelated (white noise), then

$$C_\kappa(x, y) = \sigma_\kappa^2\delta(x - y), \tag{C 3}$$

where δ is the Dirac delta function and σ_κ^2 is the variance of $\kappa(x; \omega)$. This state is reached asymptotically in time by correlated $\kappa(x)$ [2]. Substituting (C 3) into (C 1) yields

$$u_4(x_4) = \mathrm{Da}\left(1 - \frac{\mathrm{Da}\,\sigma_\kappa^2}{2}\right)(1 - x_4) \quad \text{and} \quad D_{44}(x_4) = \frac{\mathrm{Da}^2\sigma_\kappa^2}{2}(1 - x_4)^2. \tag{C 4}$$

These expressions correspond to the approximations obtained in [29] and the theoretical results in [2].

If $\kappa(x;\omega)$ has an exponential covariance function $C_\kappa(x,y)=\sigma_\kappa^2\exp(-|x-y|/\lambda_\kappa)$, then u_4 and D_{44} in (C 1) become

$$u_4(x,x_4,t)=\mathrm{Da}(1-x_4)-\frac{\mathrm{Da}^2\,\sigma_\kappa^2(1-x_4)}{(\mathrm{Da}-1/\lambda_\kappa)}[\mathrm{e}^{(\mathrm{Da}-1/\lambda_\kappa)t_*}-1] \tag{C 5a}$$

and

$$D_{44}(x,x_4,t)=\frac{\mathrm{Da}^2\,\sigma_\kappa^2(x_4-1)^2}{(\mathrm{Da}-1/\lambda_\kappa)}[\mathrm{e}^{(\mathrm{Da}-1/\lambda_\kappa)t_*}-1]. \tag{C 5b}$$

(ii) Nonlinear reaction law ($\alpha=2$)

Substituting (B 13) into (C 1) gives

$$u_4=2\mathrm{Da}(1-x_4^2)+32\,\mathrm{Da}^2\frac{x_4(x_4+1)}{x_4-1}\int_0^{t_*}C_\kappa(x,x-T)\frac{\exp(-4\,\mathrm{Da}\,T)}{(g-1)^2}\,\mathrm{d}T \tag{C 6a}$$

and

$$D_{44}=16\,\mathrm{Da}^2(x_4+1)^2\int_0^{t_*}C_\kappa(x,x-T)\frac{\exp(-4\,\mathrm{Da}\,T)}{(g-1)^2}\,\mathrm{d}T, \tag{C 6b}$$

where

$$t_*=\min\left\{t,\frac{1}{4\,\mathrm{Da}}\ln\left(\frac{1+x_4}{1-x_4}\right),x\right\}\quad\text{and}\quad g(x_4,T)=\frac{x_4+1}{x_4-1}\exp(-4\,\mathrm{Da}\,T). \tag{C 7}$$

If $\kappa(x;\omega)$ is uncorrelated, then

$$u_4(x_4)=2\,\mathrm{Da}(1-x_4^2)(1-2\,\mathrm{Da}\,\sigma_\kappa^2x_4)\quad\text{and}\quad D_{44}(x_4)=2\,\mathrm{Da}^2\,\sigma_\kappa^2(x_4^2-1)^2, \tag{C 8}$$

which is exactly as in [2,29].

(b) Sources of uncertainty: $\kappa(x)$ and $v(x)$

We consider two types of the uncertain parameter v: (i) a random variable $v(\omega)$ with variance σ_v^2 and (ii) a correlated random field $v(x;\omega)$ with the exponential covariance function $C_v(x,y)=\sigma_v^2\exp(-|x-y|/\lambda_v)$ and the correlation length λ_v.

For the uncertain v of the first type, combining (4.3*b*) with $C_v\equiv\sigma_v^2$ and G given by either (B 12) for $\alpha=1$ or (B 13) for $\alpha=2$ gives rise to

$$D_{11}=\sigma_v^2t_*, \tag{C 9}$$

where t_* depends on G and thus on α. A similar procedure applied to the uncertain v of the second type yields

$$D_{11}=\sigma_v^2\lambda_v(1-\mathrm{e}^{-t_*/\lambda_v}). \tag{C 10}$$

References

1. LeVeque RJ. 1999 *Numerical methods for conservation laws*, 2nd edn. Basle, Switzerland: Birkhuser.
2. Tartakovsky DM, Dentz M, Lichtner PC. 2009 Probability density functions for advective–reactive transport in porous media with uncertain reaction rates. *Water Resour. Res.* **45**, W07414. (doi:10.1029/2008WR007383)
3. Jarman KD, Russell TF. 2003 Eulerian moment equations for 2-D stochastic immiscible flow. *Multiscale Model. Simul.* **1**, 598–608. (doi:10.1137/S1540345902413176)
4. Cousins W, Gremaud PA, Tartakovsky DM. 2013 A new physiological boundary condition for hemodynamics. *SIAM J. Appl. Math.* **73**, 1203–1223. (doi:10.1137/120895470)
5. Lighthill MJ, Whitham GB. 1955 On kinematic waves. I. Flood movement in long rivers. *Proc. R. Soc. Lond. A* **229**, 281–316. (doi:10.1098/rspa.1955.0088)
6. Lighthill MJ, Whitham GB. 1955 On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. A* **229**, 317–345. (doi:10.1098/rspa.1955.0089)
7. Krasheninnikov SI. 2001 On scrape off layer plasma transport. *Phys. Lett. A* **283**, 368–370. (doi:10.1016/S0375-9601(01)00252-3)

8. Lika K, Hallam TG. 1999 Traveling wave solutions of a nonlinear reaction-advection equation. *J. Math. Biol.* **38**, 346–358. (doi:10.1007/s002850050152)
9. Neuman SP. 1993 Eulerian-Lagrangian theory of transport in space-time nonstationary velocity fields: exact nonlocal formalism by conditional moments and weak approximation. *Water Resour. Res.* **29**, 633–645. (doi:10.1029/92WR02306)
10. Shvidler M, Karasaki K. 2003 Probability density functions for solute transport in random field. *Transp. Porous Media* **50**, 243–266. (doi:10.1023/A:1021129325701)
11. Logan JD. 2001 *Transport modeling in hydrogeochemical systems*. Interdisciplinary Applied Mathematics, vol. 15. New York, NY: Springer.
12. Dal Passo R, de Mottoni P. 1984 Aggregative effects for a reaction–advection equation. *J. Math. Biol.* **20**, 103–112. (doi:10.1007/BF00275865)
13. Wang P, Tartakovsky DM. 2012 Uncertainty quantification in kinematic-wave models. *J. Comput. Phys.* **231**, 7868–7880. (doi:10.1016/j.jcp.2012.07.030)
14. Severino G, Dagan G, van Duijn CJ. 2000 A note on transport of a pulse of nonlinearly reactive solute in a heterogeneous formation. *J. Comput. Geosci.* **4**, 275–286. (doi:10.1023/A:1011568118126)
15. Dagan G, Neuman SP (eds). 1997 *Subsurface flow and transport: a stochastic approach*. New York, NY: Cambridge University Press.
16. Frisch U. 1968 Wave propagation in random media. In *Probabilistic methods in applied mathematics*, vol. 1 (ed. AT Bharucha-Reid), pp. 75–198. New York, NY: Academic Press.
17. Papanicolaou GC. 1973 Stochastic equations and their applications. *Am. Math. Monthly* **80**, 526–545. (doi:10.2307/2319609)
18. Salandin P, Fiorotto V. 1998 Solute transport in highly heterogeneous aquifers. *Adv. Water Resour.* **34**, 949–961. (doi:10.1029/98WR00219)
19. Ye M, Neuman SP, Guadagnini A, Tartakovsky DM. 2004 Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly heterogeneous porous media. *Water Resour. Res.* **40**, W05104. (doi:10.1029/2003WR002099)
20. Foo J, Karniadakis GE. 2010 Multi-element probabilistic collocation method in high dimensions. *J. Comput. Phys.* **229**, 1536–1557. (doi:10.1016/j.jcp.2009.10.043)
21. Xiu D, Tartakovsky DM. 2006 Numerical methods for differential equations in random domains. *SIAM J. Sci. Comput.* **28**, 1167–1185. (doi:10.1137/040613160)
22. Gottlieb D, Xiu D. 2008 Galerkin method for wave equations with uncertain coefficients. *Commun. Comput. Phys.* **3**, 505–518.
23. Lin G, Su C-H, Karniadakis GE. 2006 Predicting shock dynamics in the presence of uncertainties. *J. Comput. Phys.* **217**, 260–276. (doi:10.1016/j.jcp.2006.02.009)
24. Tryoen J, Le Maître O, Ndjing M, Ern A. 2010 Intrusive Galerkin methods with upwinding for uncertain nonlinear hyperbolic systems. *J. Comput. Phys.* **229**, 6485–6511. (doi:10.1016/j.jcp.2010.05.007)
25. Tartakovsky DM, Winter CL. 2001 Dynamics of free surfaces in random porous media. *SIAM J. Appl. Math.* **61**, 1857–1876. (doi:10.1137/S0036139999358180)
26. Dentz M, Tartakovsky DM. 2008 Self-consistent four-point closure for transport in steady random flows. *Phys. Rev. E* **77**, 066307. (doi:10.1103/PhysRevE.77.066307)
27. Lichtner PC, Tartakovsky DM. 2003 Upscaled effective rate constant for heterogeneous reactions. *Stoch. Environ. Res. Risk Assess.* **17**, 419–429. (doi:10.1007/s00477-003-0163-3)
28. Lundgren TS. 1967 Distribution functions in the statistical theory of turbulence. *Phys. Fluids* **10**, 969–975. (doi:10.1063/1.1762249)
29. Tartakovsky DM, Broyda S. 2011 PDF equations for advective-reactive transport in heterogeneous porous media with uncertain properties. *J. Contam. Hydrol.* **120–121**, 129–140. (doi:10.1016/j.jconhyd.2010.08.009)
30. LeVeque RJ (ed.). 2004 *Finite volume methods for hyperbolic problems*. New York, NY: Cambridge University Press.