International Journal for Uncertainty Quantification, 11(2):83–104 (2021)

METHOD OF DISTRIBUTIONS FOR SYSTEMS WITH STOCHASTIC FORCING

Rik J.L. Rutjens,[1,2] Gustaaf B. Jacobs,[1,*] & Daniel M. Tartakovsky[3]

[1]*Department of Aerospace Engineering, San Diego State University, San Diego, California 92182, USA*

[2]*Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands*

[3]*Department of Energy Resources Engineering, Stanford University, Stanford, California 94305, USA*

*Address all correspondence to: Gustaaf B. Jacobs, Department of Aerospace Engineering, San Diego State University, San Diego, California 92182, USA, E-mail: gjacobs@mail.sdsu.edu

Original Manuscript Submitted: 8/28/2019; Final Draft Received: 8/30/2020

The method of distributions is developed for systems that are governed by hyperbolic conservation laws with stochastic forcing. The method yields a deterministic equation for the cumulative distribution function (CDF) of a system state, e.g., for flow velocity governed by an inviscid Burgers equation with random source coefficients. This is achieved without recourse to any closure approximation. The CDF model is verified against Monte Carlo (MC) simulations using spectral numerical approximations. Our analysis demonstrates that the CDF model accurately predicts the mean and standard deviation of the system state for Gaussian, normal, and beta distributions of the random coefficients.

KEY WORDS: *uncertainty quantification, method of distributions, Monte Carlo, spectral collocation, cumulative distribution function*

1. INTRODUCTION

Deterministic predictions of models that couple multiphysics through source terms, such as chemically reactive, and multiphase and/or multimaterial flows, are notoriously difficult. A typical system involves many dependent variables, for which the exact coupling may not be known a priori. The coupling may also be imprecise because of measurement and/or numerical errors, and sparsity of experimental data. To improve upon the accuracy of the model, it is necessary to understand how uncertainty of these critical parameters propagates to a quantity of interest (QoI). Uncertainty quantification has become an integral component of computational models of chemically reacting [1–4, and the references therein] and particle-laden [5–7] flows. Given uncertainty distribution in the parameter space, the areas for improvement of the dependent variables can be identified.

If uncertainty is treated within the probabilistic framework, uncertain predictions of the QoI are fully captured by its probability density function (PDF) or cumulative ditribution function (CDF). The latter can be estimated via Monte Carlo (MC) simulations, but these converge slowly ($\sim 1/\sqrt{N_s}$, with N_s the number of samples) and, hence, are computationally inefficient. More efficient sampling techniques, e.g., multilevel MC [8] and Latin hypercube sampling (LHS) [9], may not alleviate computational cost. For example, the variance of LHS output samples can be larger than that obtained with normal sampling [10, Chap. 10].

Generalized polynomial chaos (gPC) [11] provides a nonsampling alternative to MC. It expresses uncertain parameters in terms of orthogonal polynomials of standard random variables. These expansions can be used to obtain a spectral description of the uncertain parameters and are used in stochastic finite element methods (SFEMs) [12]. SFEMs have been used to model a variety of phenomena, such as transport in porous media, solid mechanics, structural applications, and reacting flow (see [3] for a review). The multielement generalized polynomial chaos method [13,14] can handle discontinuities in the stochastic space. Unfortunately, intrusiveness is a significant downside; i.e., a

standard numerical method requires modifications for gPC that result in a high-dimensional coupled linearized system of equations, which is computationally taxing [15]. In fact, SFEM is computationally more expensive as compared to MC when a large number of random variables are considered.

Stochastic collocation methods (see [16] for an overview) are nonintrusive and require only a few repetitive calls to a deterministic solver, similar to MC. Adaptive sparse grid collocation [17] handles discontinuities in the stochastic space. However, stochastic collocation can be slower than MC, e.g., for nonlinear parabolic equations with random coefficients with high variances [18].

Statistical moment equations and the method of distributions provide yet another alternative to sampling methods. These approaches derive deterministic equations for statistical moments (typically, mean and variance) or PDF/CDF of a system state, respectively. Moment equations are derived through ensemble averaging, but they require a closure approximation. The closure is typically done through perturbation expansions or gradient models, which require empiricism and/or homogenization of higher fidelity data. Either way, major concessions are made to model accuracy through closure. Equally important is the inability of the method of moments to deal with highly non-Gaussian system states, whose dynamics cannot be fully captured with their mean and variance. The method of distributions provides the full (joint) PDF/CDF of the solution and the random inputs, including the tail behavior. PDF methods were first developed for applications in turbulence and combustion (see [19], for a review), but later extended to quantify parametric uncertainty in a variety of problems [20–26]. Applications to Burgers' equation can be found in, e.g., [27,28]. Crucially, PDF methods obviate the need for linearization of nonlinear terms. A drawback is the challenging definition of unique boundary conditions in stochastic space.

We develop a CDF method for systems with stochastic sources as they may appear in multiphysics environments such as particle-laden flow and chemically reacting flow. By way of example, we consider a Burgers' equation with random source, which renders the system stochastic. An equation for the joint CDF of the flow velocity and source coefficients is derived. The marginal PDF for the uncertain velocity can then be extracted from this joint CDF. The main advantages of this method are its simplicity, accuracy, and computational efficiency. Moreover, this method leads to an unambiguous closed system of equations. High dimensionality of the PDF equations would pose a computational challenge to the method's performance. While dealing with this problem lies outside the scope of this study, a potential strategy is to deploy parallel tensor methods [29]. A simplified version of the CDF equation is solved using spectral methods, assuming the source consists of only one random coefficient and a steady smooth source. Solutions to the simplified CDF equation are shown to be in good agreement with MC results.

The paper is structured as follows. Section 2 describes the governing equations and the numerical methods. Special attention is given to the regularization of (singular) source terms that appear in problems with deterministic initial conditions. In Section 3, results are shown and discussed for the simplified CDF equation, assuming a uniform, normal, or beta distribution for the random source coefficient. The influence of relevant parameters (e.g., grid resolution) is considered, and a thorough comparison is made with MC results. Conclusions and directions for future work are given in Section 4.

2. GOVERNING EQUATIONS AND METHODOLOGY

2.1 Burgers Equations with a Stochastic Source

Let a state variable, flow velocity $v(x,t)$, defined on $(x,t) \in [x_{\min}, x_{\max}] \times \mathbb{R}^+$, satisfy an inviscid Burgers equation with stochastic source,

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = g^s(u-v)\cdot(u-v). \tag{1}$$

The source term accounts for the relative velocity difference of the state variable v and a (deterministic or random) background velocity u, and the functional form of $g^s(\cdot)$ is unknown/uncertain. Such a formulation is common in coupled multiphysics systems, where u and v represent the solution for two coupled physics environments, respectively (e.g., particle/gas flow, chemistry/gas flow).

Equation (1) is subject to a deterministic initial condition $v(x,0) = v_{\text{in}}(x)$ and a deterministic boundary condition $v(x_{\min},t) = v_0(t)$. The unknown (random) function $g^s(\cdot)$ is represented via a polynomial with orthogonal basis functions $\mathrm{T_i}(\cdot)$,

$$g^s(\cdot) = \sum_{i=0}^{\infty} a_i \mathrm{T_i}(\cdot) \approx \sum_{i=0}^{N_g} a_i \mathrm{T_i}(\cdot). \tag{2}$$

We assume that g^s has a compact support and choose $\mathrm{T_i}(\cdot)$ to be Chebyshev polynomials of the first kind, scaled from the interval $[-1, 1]$ to $[x_{\min}, x_{\max}]$. The polynomial coefficients are uncertain and treated as (correlated or not) random variables with the (joint) PDF $f_{\mathbf{a}}(A_1, \ldots, A_{N_g})$. Combining (1) and (2) yields a PDE with random coefficients,

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = (u - v)\sum_{i=0}^{N_g} a_i \mathrm{T_i}(u - v), \tag{3}$$

whose solution is given in terms of $f_v(V; x, t)$, the PDF of the random state variable $v(x, t)$. Equation (3) can be solved with MC simulations, i.e., by repeatedly sampling the random coefficients and solving the corresponding deterministic PDEs.

2.2 CDF Equations

2.2.1 Positive Source

In Appendix A we show that the joint CDF $F_{\mathbf{a}v}(\mathbf{A}, V; x, t)$ of the set of input parameters $\mathbf{a} = \{a_0, \ldots, a_{N_g}\}$ and the state variable v at any space-time point (x, t) satisfies a deterministic integro-differential equation

$$\frac{\partial F_{\mathbf{a}v}}{\partial t} + V\frac{\partial F_{\mathbf{a}v}}{\partial x} = -(u - V)\sum_{i=0}^{N_g} T_i(u - V)\frac{\partial}{\partial V}\left[A_i F_{\mathbf{a}v} - \int_{-\infty}^{A_i} F_{\mathbf{a}v}(\mathbf{A}\backslash A_i, A_i', V; x, t)\mathrm{d}A_i'\right], \tag{4}$$

where

$$\mathbf{A}\backslash A_i = \big(A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_{N_g}\big). \tag{5}$$

The CDF of random velocity v at point (x, t), i.e., $F_v(V; x, t)$, is the marginal of $F_{\mathbf{a}v}$,

$$F_v(V; x, t) = F_{\mathbf{a}v}(\mathbf{A}_{\max}, V; x, t). \tag{6}$$

Likewise,

$$F_{a_i v}(A_i', V; x, t) = F_{\mathbf{a}v}(\mathbf{A}_{\max}\backslash A_i, A_i', V; x, t). \tag{7}$$

Hence, the marginal F_v satisfies a CDF equation,

$$\frac{\partial F_v}{\partial t} + V\frac{\partial F_v}{\partial x} = -(u - V)\sum_{i=0}^{N_g} T_i(u - V)\frac{\partial}{\partial V}\left[A_{\max,i} F_v - \int_{-\infty}^{A_{\max,i}} F_{a_i v}(A_i', V; x, t)\mathrm{d}A_i'\right]. \tag{8}$$

However, we found it more convenient to solve the equation for the joint CDF. For now, we assume the initial velocity to be deterministic. This leads to the following initial conditions:

$$F_{a_i v}(A_i, V; x, 0) = F_{a_i}(A_i)F_v(V; x, 0) = F_{a_i}(A_i)\mathcal{H}(V - v_{\mathrm{in}}(x)), \quad i = 0, \ldots, N_g, \tag{9}$$

where $\mathcal{H}$ denotes the Heaviside function. Basic properties of probability yield boundary conditions

$$\begin{aligned} F_{a_i v}(A_{\min,i}, V; x, t) &= 0, \qquad F_{a_i v}(A_{\max,i}, V; x, t) = F_v(V; x, t), \\ F_{a_i v}(A_i, V_{\min}; x, t) &= 0, \qquad F_{a_i v}(A_i, V_{\max}; x, t) = F_{a_i}(A_i), \end{aligned} \tag{10}$$

where $i = 0, \ldots, N_g$. The support of $\mathbf{A}$ and V can be either infinite or finite. In numerical simulations, the infinite support has to be restricted to sufficiently small finite intervals in order to reduce the computational cost. The support for V should be sufficiently large to ensure that (i) the initial regularized boundary condition lies within the interval, and (ii) the nonzero part of the solution does not advect out of the interval during the simulation time.

2.2.2 Negative Source

While the CDF equation is applicable for arbitrary smooth sources with compact support, the numerical method described in the next section turns out to be unstable for negative sources. This is likely due to undershoots leading to negative CDFs and steepening of the CDF. As a workaround, we solve the CDF equation,

$$\frac{\partial G_{av}}{\partial t} + V\frac{\partial G_{av}}{\partial x} = -(u-V)\frac{\partial}{\partial V}\left[AG_{av} - \int_{-\infty}^{A} G_{av}(A', V; x, t)\mathrm{d}A'\right], \tag{11}$$

where

$$G_{av}(A, V; x, t) = F_a(A) - F_{av}(A, V; x, t). \tag{12}$$

Here we have used the notation $A = A_1$ for simplicity. Following the same steps as in Appendix A, but with $\tilde{\Pi}(A, a; V, v) = \mathcal{H}(A-a) - \Pi(A, a; V, v)$, one can prove that the function G, which is not a CDF, does satisfy the CDF equation. It is subject to the following adjusted initial and boundary conditions:

$$G_{av}(A, V; x, 0) = F_a(A)(1 - \mathcal{H}(V-1)) = F_a(A)\mathcal{H}(1-V), \tag{13}$$

and

$$\begin{aligned} &G_{av}(A_{\min}, V; x, t) = 0, && G_{av}(A_{\max}, V; x, t) = 1 - F_v(V; x, t), \\ &G_{av}(A, V_{\min}; x, t) = F_a(A), && G_{av}(A, V_{\max}; x, t) = 0, \end{aligned} \tag{14}$$

while G_{av} is not a CDF; it is, for example, no longer nondecreasing in each of its variables, and the corresponding PDF can be determined as follows:

$$f_v(V; x, t) = -\frac{\partial^2}{\partial A \partial V} G_{av}(A_{\max}, V; x, t). \tag{15}$$

For the sake of simplicity, we consider one random coefficient a, instead of a random vector **a**. The extension of the result to multiple coefficients is straightforward.

2.3 Numerical Methods

Numerical solutions of the CDF equation are tested against solutions of Eq. (3) obtained by MC simulations. Both Eqs. (3) and (4) admit solutions with singularities for several reasons. First, the deterministic initial condition (9) contains a Heaviside function. Second, the Burgers equation is known to steepen solutions leading to discontinuities. Finally, certain reduced-physics models can have singular sources [30]. To obtain accurate solutions and consistency between MC and the CDF equation, we must be careful in selecting the numerical methods that we use to approximate the governing systems. The spectral approximation is superior to lower-order alternatives for a formulation based on a method of moments [5]. The solution of the CDF formulation is expected to also benefit from the low dispersion and diffusion characteristic of the spectral methods. Therefore, we rely on a low-dispersive and low-diffusive single domain Chebyshev collocation method and use some of the recently developed filtering and regularization techniques to capture shocks and regularize sources [30,31]. In the following, we briefly summarize the Chebyshev collocation method and the regularization techniques. For a detailed discussion the interested reader is referred to [30–33].

2.3.1 Chebyshev Collocation Method and Time Integration

The collocation method is based on polynomial interpolation of a function $u(x)$,

$$u_{N_x}(x) = \sum_{j=0}^{N_x} u(x_j) l_j(x), \quad l_j(x) = \prod_{k=0, k\neq j}^{N_x} \frac{x - x_k}{x_j - x_k}, \quad j = 0, \ldots, N_x, \tag{16}$$

where x_j, with $j = 0, \ldots, N_x$, are collocation points; and $l_j(x)$ are the Lagrange interpolation polynomials of degree N_x. The derivative, $u'(x_i)$, of the function $u(x)$ at collocation points x_i is expressed in terms of the derivative of the Lagrange interpolating polynomials as

$$\frac{\partial u}{\partial x}(x_i) \approx \sum_{j=0}^{N_x} u(x_j) l_j'(x_i), \tag{17}$$

or, written compactly in the matrix-vector multiplication form, as

$$\mathbf{u}' = \mathbf{D}\mathbf{u}, \tag{18}$$

where the components of the differentiation matrix $\mathbf{D}$ are $D_{i,j} = l_j'(x_i)$. For the Chebyshev collocation method, the collocation points are chosen at the Gauss-Lobatto quadrature points,

$$\xi_i = -\cos(i\pi/N_x), \quad i = 0, \ldots, N_x, \tag{19}$$

such that the L_∞ norm of the interpolant is minimized on its interval [–1,1]. Combining (3) and (18) yields a system of ordinary differential equations (ODEs) on the collocation points $\mathbf{x} \subseteq [x_{\min}, x_{\max}]$:

$$\frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t} = \mathrm{diag}(\mathbf{u} - \mathbf{v}(t)) \sum_{k=0}^{N_g} a_k \mathrm{T_k}(\mathbf{u} - \mathbf{v}(t)) - \mathrm{diag}(\mathbf{v}(t))\mathbf{D}_x\mathbf{v}(t), \tag{20}$$

in which

$$\mathbf{u} = [u(x_0), \ldots, u(x_{N_x})]^\top, \quad \mathbf{v}(t) = [v(x_0, t), \ldots, v(x_{N_x}, t)]^\top,$$
$$\mathrm{T_k}(\mathbf{u} - \mathbf{v}(t)) = [\mathrm{T_k}((u(x_0) - v(x_0, t)), \ldots, \mathrm{T_k}((u(x_{N_x}) - v(x_{N_x}, t))]^\top,$$

where $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with entries x_i, $i = 0, \ldots, N_x$, and $\mathbf{D}_x = \partial\xi/\partial x \times \mathbf{D} = 2/(x_{\max} - x_{\min}) \times \mathbf{D}$ is a scaled version of $\mathbf{D}$ to account for the mapping of the spatial domain from the Chebyshev quadrature nodes ξ to the spatial domain $\mathbf{x}$. To integrate the system of ODEs in time, we employ the fourth-order Runge-Kutta scheme [34] for the MC equation. Using orthogonality of the Chebyshev polynomial, the CDF equation is similarly discretized on a tensorial Gauss-Lobatto grid in (x, V), and given for every point $\tilde{\mathbf{A}}$ on the tensorial uniform $\mathbf{A}$-grid $\subseteq [A_{\min}, A_{\max}]^{N_g}$ by

$$\begin{aligned}\frac{\mathrm{d}\mathbf{F}_{\mathbf{a}v}}{\mathrm{d}t}(t) = &-\mathbf{D}_x\mathbf{F}_{\mathbf{a}v}(t)\mathrm{diag}(\mathbf{V}) \\ &- \sum_{k=0}^{N_g} \mathbf{D}_V \left[\tilde{A}_i \mathbf{F}_{\mathbf{a}v}(t) - \int_{-\infty}^{\tilde{A}_i} \mathbf{F}_{\mathbf{a}v}(\tilde{\mathbf{A}} \backslash \tilde{A}_i, \tilde{A}_i', t) \mathrm{d}\tilde{A}_i' \right] \mathrm{diag}[\mathrm{diag}(\mathbf{u} - \mathbf{V})\mathrm{T_k}(\mathbf{u} - \mathbf{V})],\end{aligned} \tag{21}$$

with $\mathbf{V} = [V_0, \ldots, V_{N_V}]^\top$ the grid along the V direction and $\mathbf{D}_V = 2/(V_{\max} - V_{\min}) \times \mathbf{D}$ is a scaled differentiation matrix. The $(N_x + 1) \times (N_V + 1)$-matrix $\mathbf{F}_{\mathbf{a}v}(t)$ is given by $\mathbf{F}_{\mathbf{a}v}^{i,j}(t) = \mathbf{F}_{\mathbf{a}v}(V_j, x_i, t)$. We have taken an equal amount of grid points in the directions x and V directions, i.e., $N_x = N_V$. Since we use a tensorial grid, the integral on the RHS of Eq. (21) is evaluated along lines in the A-coordinate direction. We found that the trapezoid rule is sufficiently accurate.

Because of inherent sharp gradients in the solution of the CDF equations, the third-order total variation diminishing (TVD) Runge-Kutta scheme [35],

$$\begin{aligned} u^{(1)} &= u^n + \Delta t L(u^n), \\ u^{(2)} &= \frac{3}{4}u^n + \frac{1}{4}u^{(1)} + \frac{1}{4}\Delta t L(u^{(1)}), \\ u^{n+1} &= \frac{1}{3}u^n + \frac{2}{3}u^{(2)} + \frac{2}{3}\Delta t L(u^{(2)}), \end{aligned} \tag{22}$$

is used to reduce numerical oscillations induced in time. In Eq. (22), L denotes the discrete spatial derivative operator.

2.3.2 Filtering

Following [31], we filter the solution,

$$\tilde{u}(x) = \int_{x-\epsilon}^{x+\epsilon} u(\tau)\delta_\epsilon^{m,k}(x-\tau)\mathrm{d}\tau, \tag{23}$$

using a kernel that regularizes the Dirac delta function with a class of high-order, compactly supported piecewise polynomial [30],

$$\delta_\epsilon^{m,k}(x) = \begin{cases} \frac{1}{\epsilon}P^{m,k}\left(\frac{x}{\epsilon}\right), & x \in [-\epsilon, \epsilon] \\ 0, & \text{otherwise} \end{cases}, \tag{24}$$

where $\epsilon > 0$ is the support width or scaling parameter. The polynomial $P^{m,k}$ controls the number of vanishing moments m, and the number of continuous derivatives at the end points of the compact support k. The filter based on the Dirac-delta approximation $\delta_\epsilon^{m,k}$ converges according to $\mathcal{O}\left(\epsilon^{m+1}\right)$ in smooth solution regions away from regularization areas [30].

Filtering of the interpolant u_N (16) leads to

$$\tilde{u}_{N_x}(x) = \int_{x-\epsilon}^{x+\epsilon} \left[\sum_{i=0}^{N_x} u(x_i)l_i(\tau)\right]\delta_\epsilon^{m,k}(x-\tau)\mathrm{d}\tau = \sum_{i=0}^{N_x} u(x_i)S_i(x), \tag{25}$$

after interchanging summation and integration, where the filtering function S_i is given by

$$S_i(x) = \int_{x-\epsilon}^{x+\epsilon} l_i(\tau)\delta_\epsilon^{m,k}(x-\tau)\mathrm{d}\tau. \tag{26}$$

On the discrete collocation points the convolution reduces to a matrix-vector product,

$$\tilde{\mathbf{u}} = \mathbf{S}\mathbf{u}, \tag{27}$$

where the $(N_x+1)\times(N_x+1)$ filtering matrix $\mathbf{S}$ has the elements

$$S_{i,j} = \int_{x_j-\epsilon}^{x_j+\epsilon} l_i(\tau)\delta_\epsilon^{m,k}(x_j-\tau)\mathrm{d}\tau. \tag{28}$$

The filtering matrix $\mathbf{S}$ can be precomputed. In two dimensions, the filtering operation is written as $\tilde{\mathbf{U}} = \mathbf{S}_x\mathbf{U}\mathbf{S}_y^{\mathrm{T}}$. For the solution of the CDF equation, the filtering procedure is applied twice to regularize the Heaviside function in the initial joint CDF $F_{\mathbf{a}v}(A,V;x,t=0) = F_{\mathbf{a}}(A)\mathcal{H}(V-1)$ at every A-grid point. Applying it once is sufficient for obtaining a high-order resolution result, but applying it twice makes the solution visually more pleasing as it is smoother. Favorable convergence properties remain.

2.3.3 Exponential Filter

The marginal CDF solution increases sharply at $V = V_{\max}$ (Fig. 1). At this location no boundary condition is specified because the solution moves out of the domain along its characteristic. The boundary spike does not always lead to instability and does not affect the meaningful CDF solution in the center of the domain. It is not entirely clear what the cause for this increase is, but the global nature of the collocation approximation and, hence, global sensitivity of the solution can yield this type of change at the boundary. We found that an exponential filter applied at every time step to the solution in the regions close to boundary, where the CDF is near constant, suppresses the undesired boundary behavior. Following [33,36], we use the filter

$$\tilde{F}_{av} = S_{\exp}F_{av}, \tag{29}$$

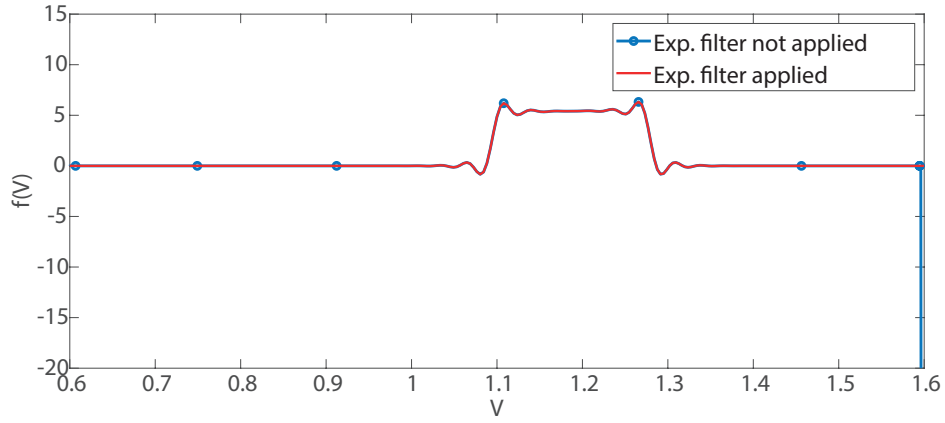


FIG. 1: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by solving the CDF equation (35) with and without applying a (fifth order) exponential filter, assuming $a \sim \mathcal{U}([0.5, 1.5])$. The regularization filter corresponding to $k = 8, m = 13, N_d = 50$ is used. Further, $N_x = 400$ and $N_a = 10$.

where the elements of the pth order $(N_x + 1) \times (N_x + 1)$ exponential filter matrix S_{exp} are given by

$$S^{i,j}_{\text{exp}} = \frac{2}{c(j)N_x} \sum_{k=0}^{N_x} \frac{1}{c(k)} \left[\mathrm{e}^{-\alpha(k/N_x)^p} \cos\left(\frac{ik\pi}{N_x}\right) \cos\left(\frac{jk\pi}{N_x}\right) \right]. \tag{30}$$

Here, **c** is the $(N_x + 1)$-dimensional vector $\mathbf{c} = [2, 1, 1, \ldots, 1, 2]$ and $\alpha = -\ln(10^{-16})$. In general $p = 5$ suppresses the spike (Fig. 1).

2.3.4 Sampling the PDF

To compare MC and the CDF model, we must determine a PDF from a set of samples. For this, we use the built-in Matlab kernel density estimator `ksdensity` [37]. A kernel distribution is a nonparametric representation of the PDF of a random variable. It is used when a parametric distribution cannot properly describe the data, or when one wants to avoid making assumptions about the distribution of the data. A kernel distribution is defined by a smoothing kernel and a bandwidth value, which control the smoothness of the resulting density curve. We refer the interested reader to [38] for more information on density estimation. The kernel density estimator's formula is given by

$$\hat{f}_{Bw}(\xi) = \frac{1}{N_s Bw} \sum_{i=1}^{N_s} K\left(\frac{\xi - \xi_i}{Bw}\right), \tag{31}$$

where $\xi_1, \xi_2, \ldots, \xi_{N_s}$ are random samples from an unknown distribution, N_s is the sample size, $K(\cdot)$ is the kernel smoothing function, and Bw is the bandwidth.

Here, we choose the density estimate produced by `ksdensity` to be based on a normal (Gaussian) kernel function. Other kernels, notably the box, triangle, or Epanechnikov kernel, can also be used. Details about the used parameter values are given in Section 3.1.

2.4 Setup for Numerical Tests

To verify consistency between MC and the CDF model, we consider only the first random coefficients in $\mathbf{a} = \{a_0, \ldots, a_{N_g}\}$; i.e., **a** is replaced with a_0. Further, the source is a steady Gaussian source,

$$u(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - x_a)^2}{2\sigma^2}\right], \tag{32}$$

where x_a denotes the center of the source and σ is a measure of the width of its support. This results in a Burgers test model,

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = a\left(\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-x_a)^2/2\sigma^2} - v\right), \tag{33}$$

which we define on the spatial interval $x \in [0, 0.06]$. The source is placed in the center of the interval, $x_a = 0.03$, and the spread of the source is given by $\sigma = 5 \times 10^{-3}$. The initial velocity v is deterministic, satisfying boundary and initial conditions,

$$v(x, 0) = 1, \quad v(0, t) = 1. \tag{34}$$

The initial-boundary value problem (33) and (34) is solved using MC simulation.

The CDF equation (4) takes the form

$$\frac{\partial F_{av}}{\partial t} + V\frac{\partial F_{av}}{\partial x} = -\left(\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-x_a)^2/2\sigma^2} - V\right)\frac{\partial}{\partial V}\left[AF_{av} - \int_{-\infty}^{A} F_{av}(A', V; x, t)\mathrm{d}A'\right]. \tag{35}$$

It is defined on $(x, V, A) \in [0, 0.06] \times [V_{\min}, V_{\max}] \times [A_{\min}, A_{\max}]$; and is subject to initial condition

$$F_{av}(A, V; x, 0) = F_a(A)F_v(V; x, 0) = F_a(A)\mathcal{H}(V - v_{\text{in}}(x)), \tag{36}$$

with $v_{\text{in}}(x) = 1$, and boundary conditions

$$\begin{aligned} F_{av}(A_{\min}, V; x, t) &= 0, \qquad F_{av}(A_{\max}, V; x, t) = F_v(V; x, t), \\ F_{av}(A, V_{\min}; x, t) &= 0, \qquad F_{av}(A, V_{\max}; x, t) = F_a(A). \end{aligned} \tag{37}$$

We set $V_{\min} = 0.6$, $V_{\max} = 1.6$, $A_{\min} = 0.5$, and $A_{\max} = 1.5$. To test consistency for a number of CDF distributions, we consider a uniform ($a \sim \mathcal{U}([0.5, 1.5])$), normal ($a \sim \mathcal{N}(1, 0.15)$), and beta distribution ($a \sim \mathcal{B}(2, 5) + 0.5$) for the random source coefficient a. The distribution parameters are chosen such that the support of the distribution equals $[0.5, 1.5]$. Therefore, the beta distribution—defined on $[0, 1]$—is translated by 0.5 to the right, but to improve readability we will denote it by $a \sim \mathcal{B}(2, 5)$ in what follows. The normal distribution has infinite support, but with the chosen mean and variance, 99.9% of its mass lies in $[0.5, 1.5]$. By taking this set of distributions, we test the CDF method on discontinuous, smooth, and skewed density functions.

Following numerical experiments (see next section), we take $N_V = 100$ grid intervals (hence 101 grid points) in the V direction and $N_s = 20{,}000$ samples of the random coefficient a to solve the MC equation (33), while in the CDF routine we consider $N_x = N_V = 400$ grid intervals in both the x and V directions, and $N_A = 10$ grid intervals in the A direction to solve (35). We use a regularization filter corresponding to $k = 8, m = 13$, and $N_d = 50$, which gives stable and filter-independent results up to 400 intervals in the V direction for the CDF equation. Regularization is not needed in the MC routine, since all functions and solutions are smooth there. In the CDF routine, an exponential filter of order $p = 5$ is applied after every time step to the 50 rightmost V–grid points (see next chapter). The integral in Eq. (35) is approximated using the trapezoidal rule.

In all simulations, $t = 2.5 \times 10^{-3}$ is taken as the final integration time. That falls within the limits dictated by the shock-forming nature of the Burgers equation; our initial numerical study (not presented in this paper) using the method of characteristics indicates that integration up to $t = 1.5 \times 10^{-2}$ (given $a \in [0.5, 1.5]$) proceeds without risking shock formation. As a becomes larger, this upper bound on the integration time naturally decreases.

3. RESULTS

In this section we test consistency between MC and the CDF model. First, the effect of different parameter values (grid resolution, sample size, etc.) are considered for both MC and CDF. Subsequently, the MC solution with the highest number of samples and optimal bandwidth and CDF solution with the finest grid are directly compared. All results are obtained with Matlab, 2018a.

3.1 Monte Carlo Simulations

The accuracy of the MC solution depends on several factors, including the number of spatial grid intervals N_x, the number of samples N_S, time step Δt, and the input parameters of the density estimator (type of smoothing kernel, number of grid points N_{ks}, and bandwidth). Below we discuss the effect of each of these parameters on the solution.

Figure 2 shows the effect of the number of grid points N_{ks} in the V direction on the density estimation; we set $a \sim \mathcal{U}([0.5, 1.5])$, $N_x = 100$, $N_S = 20{,}000$ and the optimal bandwidth (see below). The default number of grid points $N_{\text{ks}} = 100$ provides sufficient accuracy. Likewise, Fig. 3 depicts the effect of the x-grid resolution, for $a \sim \mathcal{U}([0.5, 1.5])$, $N_x = 50$, 100, or 150; $N_S = 20{,}000$; and the optimal bandwidth. We conclude that $N_x = 100$ gives sufficiently accurate results. Therefore, all MC results correspond to $N_x = 100$ and $N_{\text{ks}} = 100$. To ensure stability, the time step Δt is a function of the number of spatial grid points: $\Delta t = \lambda(x_{\max} - x_{\min})/(N_x + 1)^2$, where $\lambda = 1.5$ is the CFL condition number.

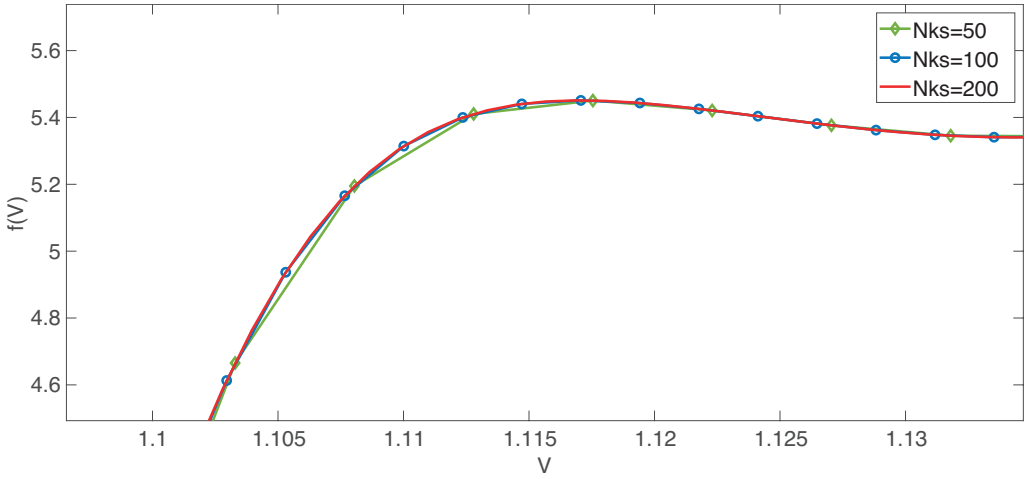


FIG. 2: Closeup of PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by MC simulation, for $a \sim \mathcal{U}([0.5, 1.5])$. In this figure, $N_x = 100$; $N_S = 20{,}000$; and $N_{\text{ks}} = 50$, 100, or 200. The optimal bandwidth is used.

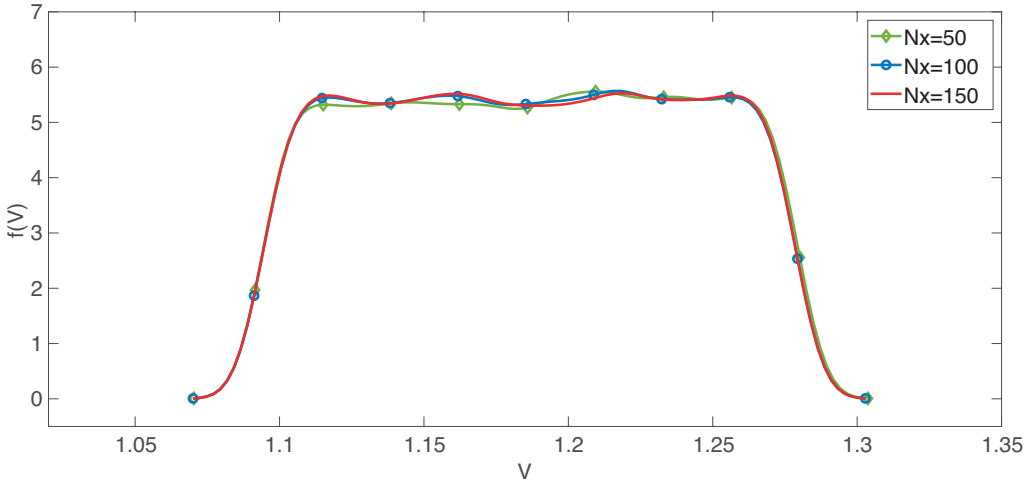


FIG. 3: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by MC simulation, for $a \sim \mathcal{U}([0.5, 1.5])$. In this figure, $N_S = 20{,}000$; $N_{\text{ks}} = 100$; and $N_x = 50$, 100, or 150. The optimal bandwidth is used.

Figures 4 and 5 show the effect of the sample size on the solution to the problem with a uniform and beta distribution, respectively. For $N_S = 20{,}000$, the PDF $f_v(V)$ is sufficiently resolved to be used as a yardstick for the subsequent comparison.

The bandwidth has a significant impact on the accuracy of the PDF determination. Too large of a value produces a significant overlap between the individual kernels. In accordance with Eq. (31), being the sum of almost identical Gaussians, this would result in a nearly Gaussian, thereby obscuring the underlying behavior. If the bandwidth is too small, the support of most individual kernels—especially those belonging to sample outliers—will be isolated, causing the estimate to look like a collection of multiple steep peaks. In other words, sample outliers are not ignored in this case, but influence the density estimate significantly. We empirically determine an optimal bandwidth for each distribution at each point in time and space, by taking the smallest bandwidth that yields a unimodal PDF. For the uniform distribution, this is not possible. Instead, a bandwidth is chosen such that the support of the PDF resembles the support of the actual sample, while avoiding too many peaks in the center part. Figure 6 shows the difference

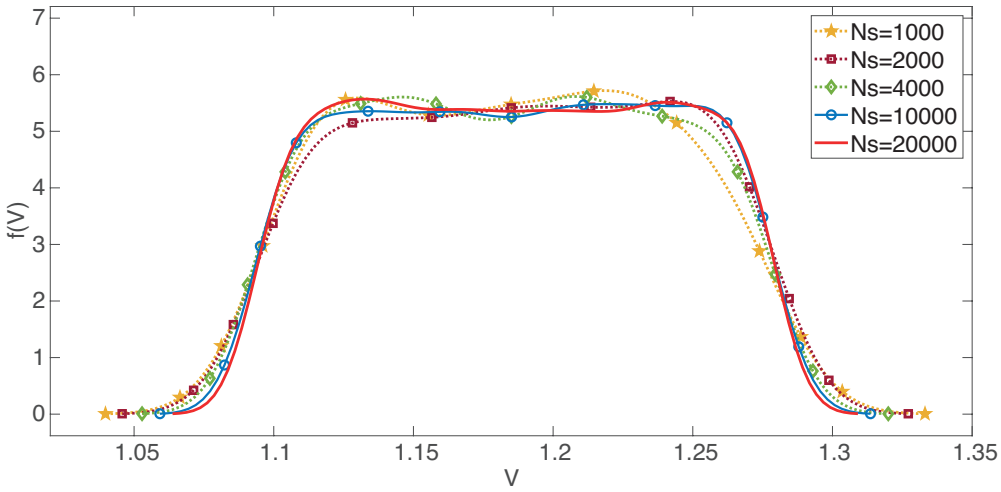


FIG. 4: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by MC simulation, for $a \sim \mathcal{U}([0.5, 1.5])$. In this figure, $N_x = 100$ and the number of samples N_S is either 1000, 2000, 4000, 10,000, or 20,000. The default bandwidth is used.

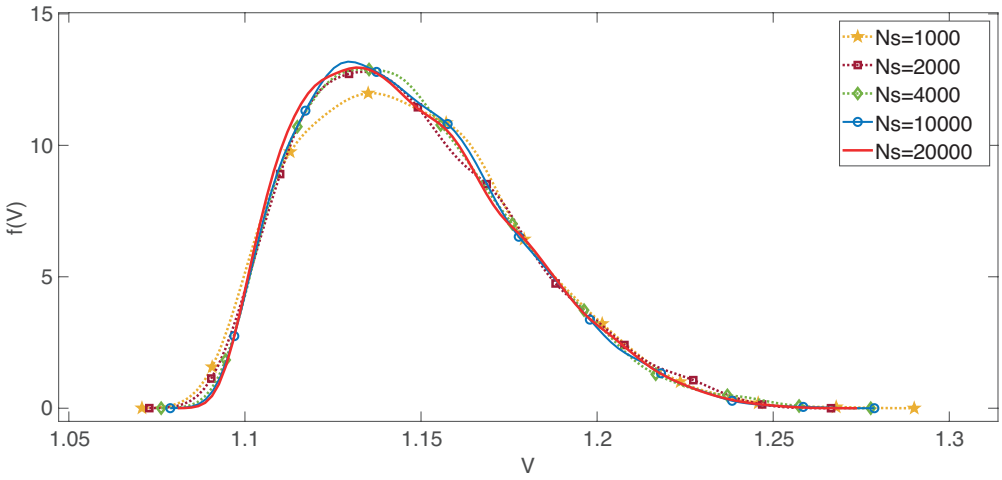


FIG. 5: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by MC simulation, for $a \sim \mathcal{B}(2, 5)$. In this figure, $N_x = 100$ and the number of samples N_S is either 1000, 2000, 4000, 10,000, or 20,000. The default bandwidth is used.

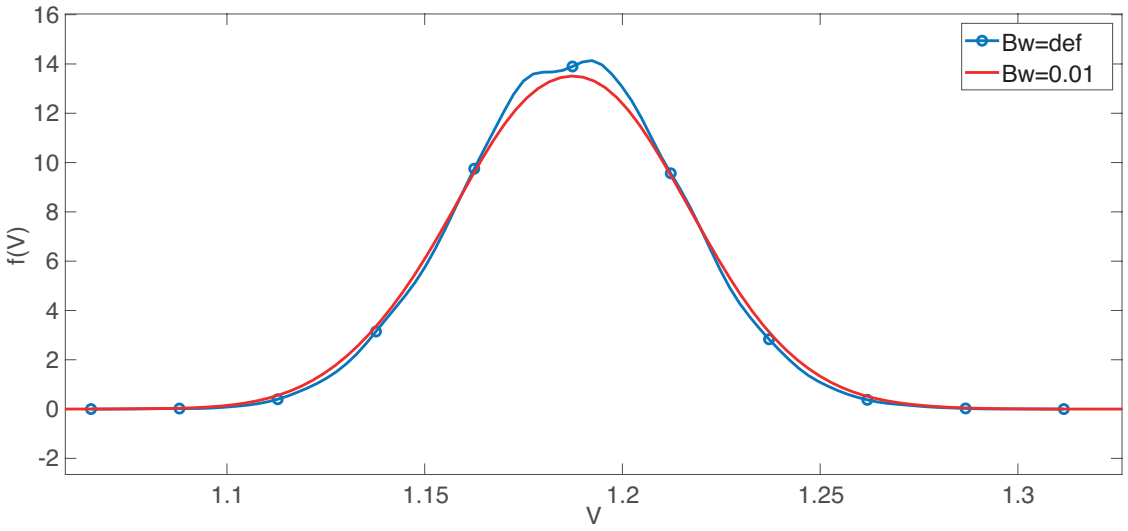


FIG. 6: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by MC simulation, for $a \sim \mathcal{N}(1, 0.15)$. In this figure, $N_x = 100$, $N_s = 20{,}000$, and both the default and optimal bandwidths are shown.

between the default and optimal bandwidth for $a \sim \mathcal{N}(1, 0.15)$. It reveals that the default bandwidth is too small, resulting in a multimodal distribution. By increasing the bandwidth, one obtains a unimodal distribution, without losing the underlying behavior: the location of the maximum and the tail behavior remain very similar. The so-called optimal bandwidths are 0.008, 0.01, and 0.01 for uniform, normal, and beta-distributed a, respectively.

Remark 1. Our numerical experiments reported in Fig. 7 suggest a linear relationship (dependent on position x and time t) between a and v, i.e., $v(x,t) = c(x,t)a$. This implies that v has the same PDF as a, but with unknown distribution parameters, which depend on time and space. Hence, instead of using a less accurate nonparametric density estimator (as we do in the present study), one can fit the velocity PDF at any (x, t) fit to the PDF of a.

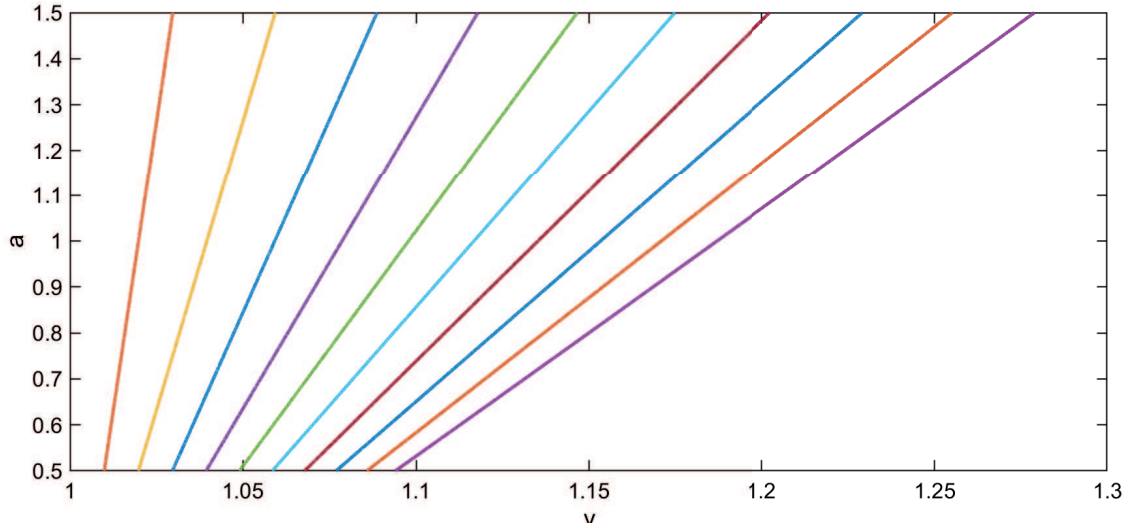


FIG. 7: Evolution through time of deterministic velocity v as a function of a. For all lines $x = 0.03$. From left to right, $t = 2.5k \times 10^{-4}$ with $i = 1, \ldots, 10$.

3.2 CDF Equation

For the CDF equation, the solution accuracy is affected by the grid resolution in the A, x, and V directions, as well as by the regularization and exponential filter settings. Following [31] and the discussion above, we choose the regularization filter with $m = 13$, $k = 8$, $N_d = 50$. An exponential filter of order $p = 5$ is applied locally (i.e., to the 50 rightmost V-grid points) to suppress oscillations near the boundary without impacting the solution at the center of the V domain. This local filtering is only possible because the solution is already approximately zero at the 51st grid point (Fig. 1). The transition from nonfiltered to filtered is thus smooth. This is generally not the case; one should then filter globally, which smears the solution. A finer (higher-order) filter only partly mitigates this issue.

In the A direction, $N_A = 10$ grid intervals is sufficient (Fig. 8), as confirmed by the relatively small value of the loss function at $t = 2.5 \times 10^{-3}$ and $x = 0.03$:

$$\max_V \left| \frac{f_v(V|N_a = 10) - f_v(V|N_a = 40)}{\max_V[f_v(V|N_a = 40)] - \min_V[f_v(V|N_a = 40)]} \right| = 0.0176. \tag{38}$$

In the V direction, the grid resolution has a significant effect. The regularization zone of possible discontinuities in the solution (CDF or PDF) can shrink as the resolution increases (Fig. 9). In the advected regularization zone of the numerical solution, the undershoots are visible, violating the positivity of the PDF. These undershoots are a direct consequence of the approximation of the deterministic Dirac delta initial condition. As discussed in Section 2.3.2, the polynomial approximation of the delta sequence with support width ϵ [30,39,40], $\delta_\epsilon^f(x) \equiv \epsilon^{-1} f(x/\epsilon)$ where f is the generating function, must have m vanishing moments in order to obtain mth-order accuracy. The sequence of regular distributions, $T_\epsilon^f[\phi] \equiv \int_{\mathbb{R}} \delta_\epsilon^f(x)\phi(x)\mathrm{d}x$, must have the property $\lim_{\epsilon\to 0^+} T_\epsilon^f[\phi] = \phi(0)$. To preserve the mth-order convergence of the delta sequence in a kth-order polynomial representation of the sequence of distributions, we must relax the positiveness of the delta sequence. An optimal value for ϵ, which ensures high-order accuracy in the solution of problems with regularized singular sources like the problem described here, was found in [30,40]. The kernel can be applied to the regularization of singularities in spectral solutions if used as a kernel in a convolution filter [31]. It was proven that high-order convergence or resolution is obtained away from the regularization zone according to the number of vanishing moments. Integral to the recovering of this high-order resolution is relaxation of the positivity of the delta function. Hence, even though the undershoot makes the solution visually unsettling, the solution outside of the regularized delta-sequence zone is more accurate. As the grid resolution increases, the solution converges outside the regularization zone according to the theoretical convergence rate. For N_x and $N_V = 400$, the solution outside

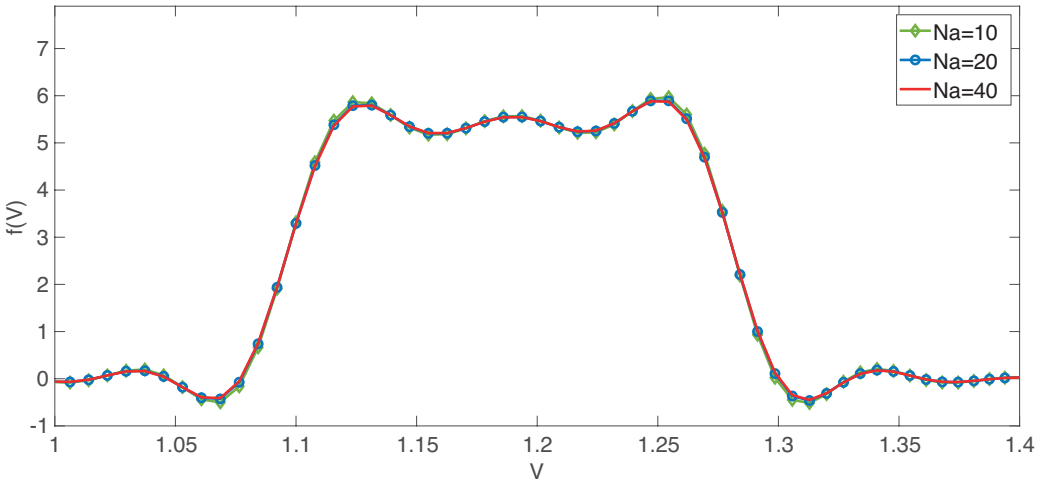


FIG. 8: PDF $f_v(V)$, at $t = 2.5 \times 10^{-3}$ and $x = 0.03$, obtained by solving the CDF equation (35), for $a \sim \mathcal{U}([0.5, 1.5])$. The regularization filter corresponding to $k = 8$, $m = 13$, $N_d = 50$ is used. Further, $N_x = N_v = 200$ and $N_a = 10$, 20, or 40. A fifth-order exponential filter has been applied to the 50 rightmost grid points.

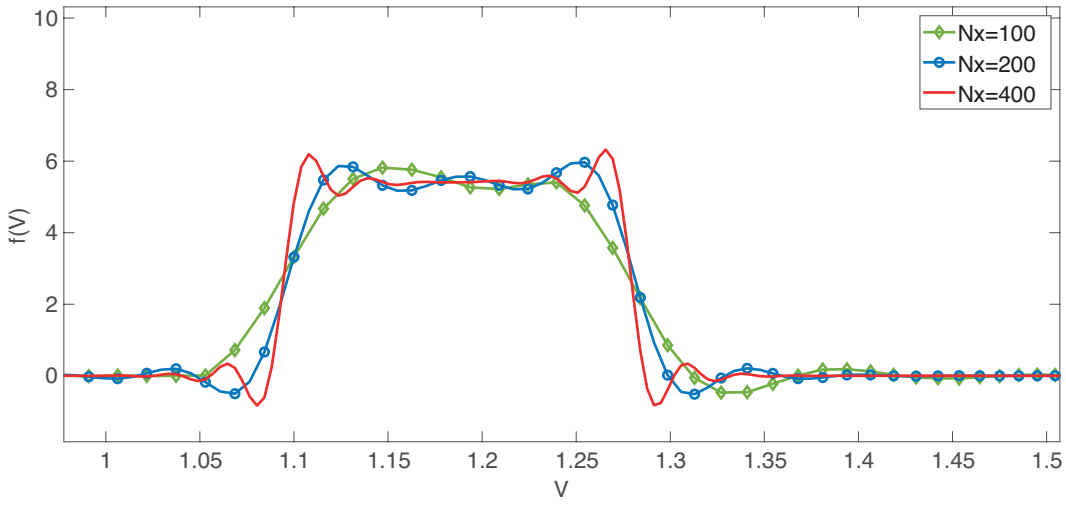


FIG. 9: PDF $f_v(V)$, at $t = 2.5 \times 10^{-3}$ and $x = 0.03$, obtained by solving the CDF equation (35), for $a \sim \mathcal{U}([0.5, 1.5])$. The regularization filter corresponding to $k = 8$, $m = 13$, $N_d = 50$ is used. Further, N_x and N_v are either 100, 200, or 400 and $N_a = 10$. A fifth-order exponential filter has been applied to the 50 rightmost grid points.

the regularization zone has converged (Fig. 9), an observation that is confirmed by the value of the loss function at $t = 2.5 \times 10^{-3}$ and $x = 0.03$:

$$\max_{V\in I}\left|\frac{f_v(V|N_V = 200) - f_v(V|N_V = 400)}{\max_V[f_v(V|N_V = 400)] - \min_V[f_v(V|N_V = 400)]}\right| = 0.0093, \tag{39}$$

where the first maximum runs over the interval $I = [0.6, 1] \cup [1.4, 1.6]$ to avoid the regularization zone; the time step is $\Delta t = 2\lambda(x_{\max} - x_{\min})/(V_{\max}N_x^2)$, where $\lambda = 1.2$ is the CFL condition number.

3.3 Comparison of MC and CDF Solutions

With the grid independence established for both the MC and CDF solutions, we test consistency between the two methods. As a first indicator, we compare their zeroth moment, the mean, standard deviation, and skewness. For MC, the mean, standard deviation, and skewness are calculated with built-in Matlab functions. For the CDF solution, these moments are computed via numerical integration of the computed PDF $f_v(V)$.

Table 1 demonstrates that both the MC and CDF solutions satisfy the property that the PDF integrates to unity. The other moments agree well, except for the skewness for a normal distribution of a, which we expect to be zero

TABLE 1: Statistical moments of the CDF solution $F_v(V)$ with $N_x = N_V = 400$ and the MC solution with $N_x = 100$, $N_s = 20{,}000$, and the optimal bandwidth, both at $t = 2.5 \times 10^{-3}$ and $x = 0.03$

Distr.	Method	$\int f_v$	Mean	Standard deviation	Skewness
Uniform	MC	1.000	1.1867	0.0532	–0.0061
	CDF	1.000	1.1872	0.0530	–0.0080
Gaussian	MC	1.000	1.1870	0.0275	0.0049
	CDF	1.000	1.1868	0.0281	0.0350
Beta	MC	1.000	1.1471	0.0296	0.5904
	CDF	1.000	1.1477	0.0301	0.5681

based on the reasoning in Remark 1 (above). We attribute the deviation from zero to the regularization, which has a more severe effect on the left tail (see Remark 2 below), thereby creating a slightly skewed approximation of a Gaussian. Overall, the agreement between the MC and CDF moments is excellent, verifying consistency between the two methods.

Figures 10 and 11 compare the PDFs determined with MC and CDF. General features like support, position of the maximum (when applicable), and general shape are in excellent agreement, as can be expected from the agreement in the statistical moments. Initial undershoots and overshoots induced by initial filtering with the nonpositive delta kernel nearly disappear at the final time in the case of a normal distributed source coefficient. They are, however, causing small deviations at the tails and tops of the normal and beta distribution [Figs. 10(b) and 10(c)]. In the case of a uniform (hence discontinuous) distribution, the undershoots and overshoots are necessary for regularization and apparent at the discontinuous edges of the PDF. Away from the regularization zones, however, the agreement is very good.

Figure 12 exhibits the PDFs $f_v(V)$ computed with MC and the CDF method for a negative steady Gaussian source,

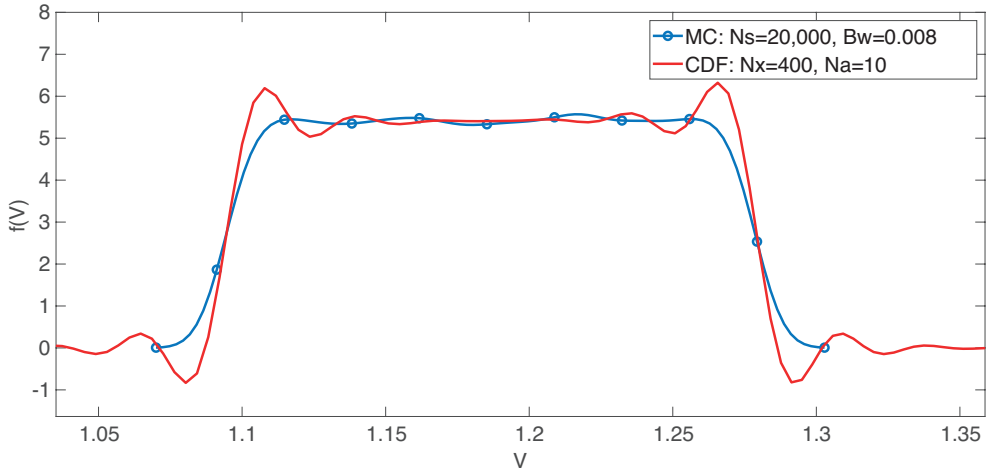
$$u(x) = -\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-x_a)^2/2\sigma^2}, \tag{40}$$

and the beta-distributed a. Other distributions for a give similar results, establishing consistency for negative sources. Only $N_x = 200$ spatial intervals were considered for the CDF equation, and $N_s = 2000$ samples for MC (all with the optimal bandwidth as mentioned before). Nevertheless, this is sufficient for the purpose of verification of the adjusted CDF equation (11).

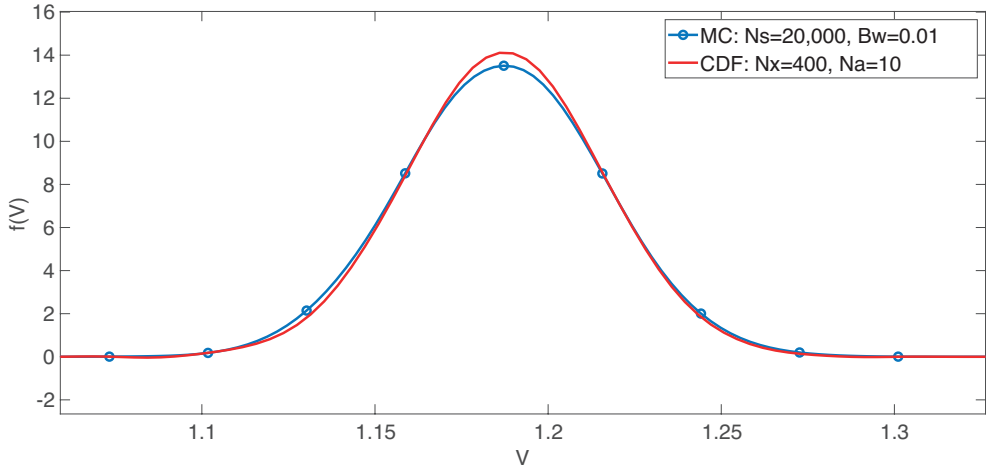
Figures 13 and 14 exhibit the predicted (mean) flow velocity $v(x,t)$ accompanied by the two standard deviations' uncertainty bound. The uncertainty bounds are generated by calculating the standard deviation (in the way explained above) at every spatial grid point. The solutions predicted with MC and the CDF equation are visually indistinguishable when plotted as either a function of x for a fixed t (Fig. 13) or a function of t for a fixed x (Fig. 14). The figures show a linear interpolation of ten uniformly spaced time points. As expected, the predicted solution v increases due to the positive source, and the prediction uncertainty increases with time. The largest uncertainty bounds correspond to uniformly distributed a, which has the most mass towards the A domain boundary of the three distributions. Finally, the mean corresponding to the left-skewed beta-distributed a is lower than the other two, the latter being equal since both have symmetric distributions of a.

Finally, in terms of computational efficiency, a significant reduction in computation time is obtained; the CDF method is 10–15 times faster than MC simulation with 20,000 samples. Considering that ideally one would want to use many more samples for MC when computing a distribution, rather than its low moments, the gain in computational efficiency may be even higher. This finding is consistent with others, which revealed the method of distributions to be up to two orders of magnitude faster than MC [41–43]. These results come with a caveat. The CDF method does not scale well with the number of state variables, because each of these variables adds a dimension to a CDF equation. One can ameliorate this problem by deploying numerical techniques for high-dimensional PDEs, e.g., parallel tensor methods [29]. In this regard, it is worthwhile pointing out that the problem of scalability with the number of random state variables is generic, affecting other uncertainty quantification techniques such as (multilevel) MC, especially when they are used to estimate (joint) distributions, rather than their low moments [44,45].

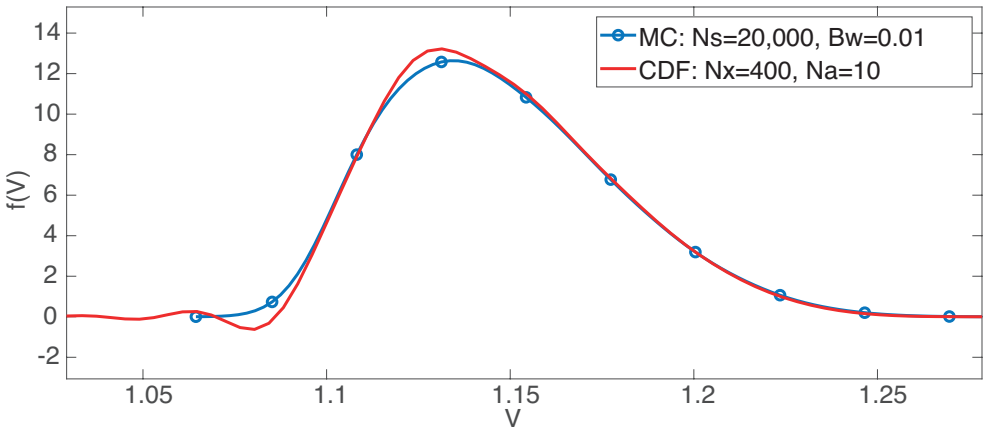
Remark 2. Because of the regularization it is not possible to make rigorous statements about (the order of) convergence of the CDF solution to the MC solution as the number of grid points increases; in the advected regularization zone, the CDF solution does not converge to its MC counterpart, since there the former necessarily has undershoots and overshoots. For the uniformly distributed input, these zones lie at both boundaries of the support of the MC solution [Fig. 10(a)]. For the Gaussian or beta-distributed input, the propagated regularization zone runs from the left support boundary to the maximum of the MC solution [Figs. 10(b) and 10(c)]. Figure 10(c) reveals that the right tail is less prone to under- and overshoot. Moreover, for the uniform input, the estimated kernel density does not have a flat plateau shape (which the analytical solution should have), but shows multiple smaller peaks. We must therefore be satisfied with the visual comparisons, the good agreement that we generally find in the moments, and the PDFs that confirm consistency between MC and CDF.



(a) $a \sim \mathcal{U}([0.5, 1.5])$



(b) $a \sim \mathcal{N}(1, 0.15)$



(c) $a \sim \mathcal{B}(2, 5)$

FIG. 10: PDF $f_v(V)$, at $t = 2.5 \times 10^{-3}$ and $x = 0.03$, obtained by solving the CDF equation (35) and by MC simulation, for (a) uniform, (b) Gaussian, and (c) beta distributions of a

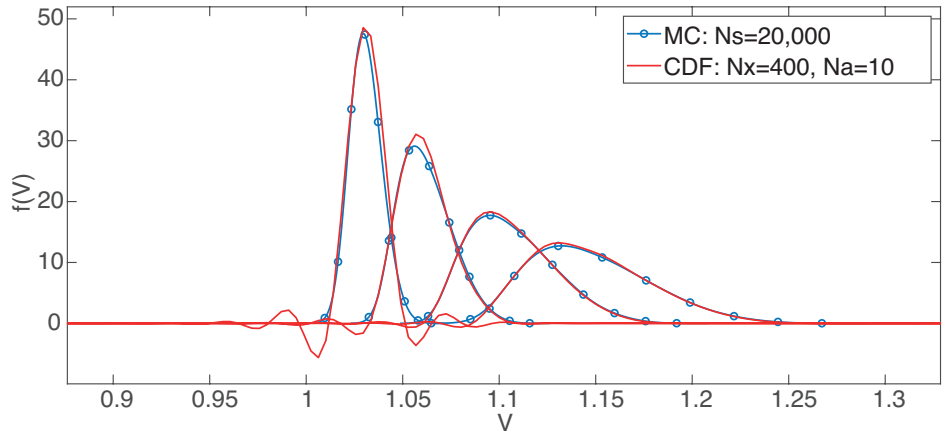


FIG. 11: PDF $f_v(V)$ at (from left to right) $t = 5 \times 10^{-4}$, 1×10^{-3}, 1.75×10^{-3}, 2.5×10^{-3}, and $x = 0.03$ obtained by solving the CDF equation (11) and MC simulation, assuming $a \sim \mathcal{B}(2, 5)$. The used bandwidths are $0.0055, 0.0055, 0.0065$, and 0.01, respectively.

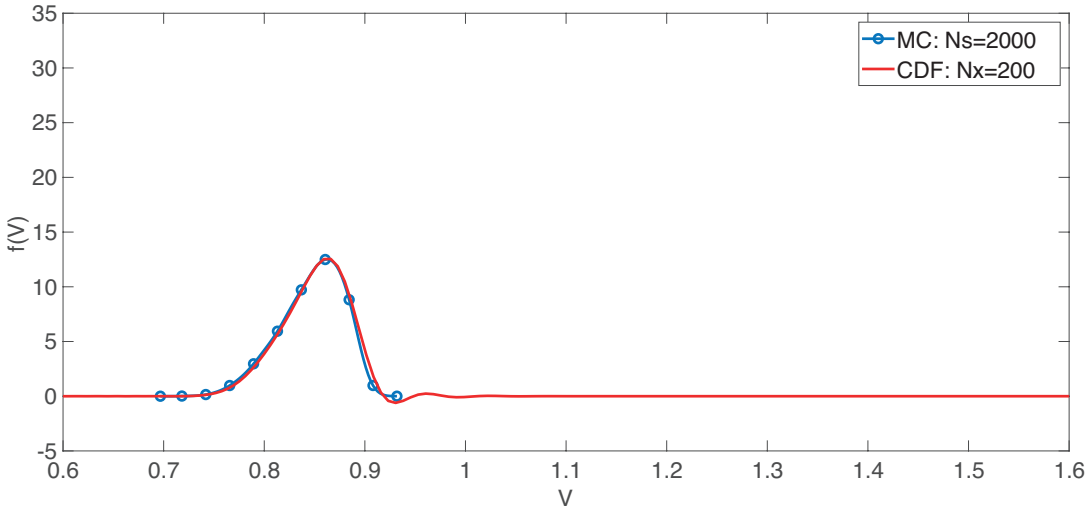


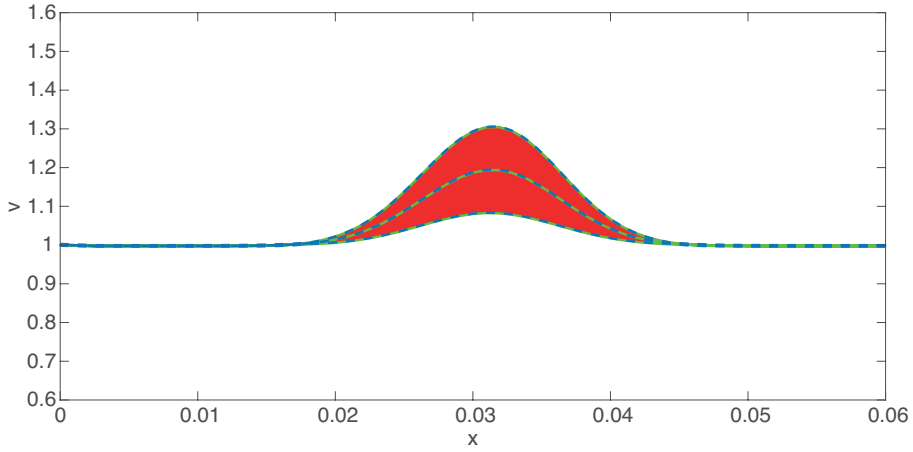
FIG. 12: PDF $f_v(V)$ at $t = 2.5 \times 10^{-3}$ and $x = 0.03$ obtained by solving the CDF equation (11) and MC simulation, assuming a negative source and $a \sim \mathcal{B}(2, 5)$

4. CONCLUSIONS AND FUTURE DIRECTIONS

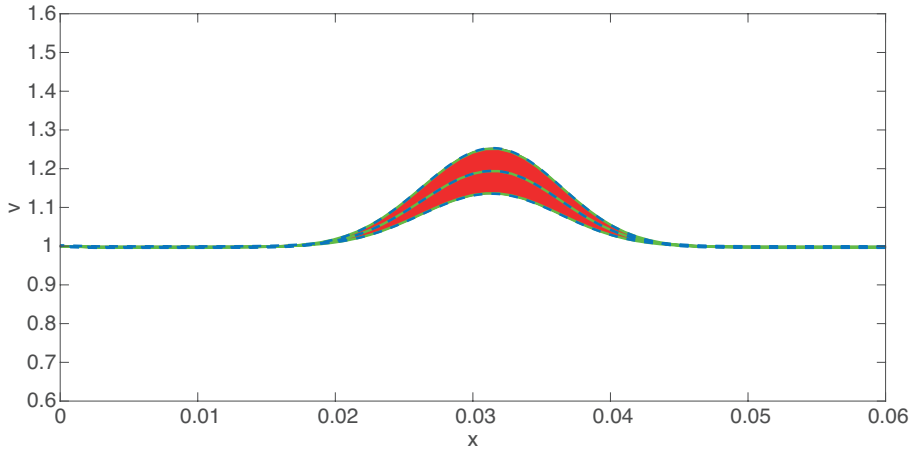
We developed a CDF method for hyperbolic systems with stochastic sources, which describe numerous multiphysics phenomena such particle-laden and chemically reacting flows. In particular, the Burgers equation with random source coefficients was considered. An equation for the joint CDF of the QoI and source coefficients was derived.

A major advantage of the CDF approach over other existing methods—MC, SFEM, method of moments, and PDF equations—is that it accounts for a full description of uncertain parameters, including their tail behavior, while being computationally efficient. Furthermore, the CDF method results in an unambiguous, closed system of equations. High dimensionality of the PDF/CDF equations would pose a computational challenge to the method's performance; a potential remedy is to deploy parallel tensor methods [29].

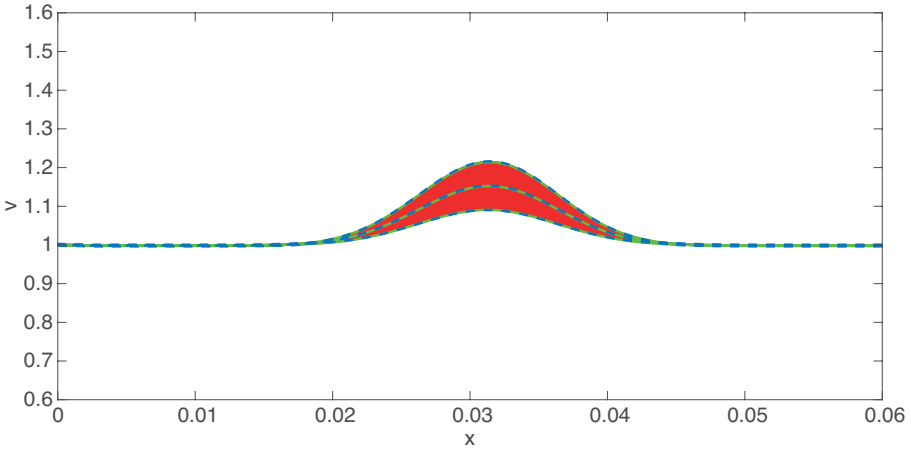
If the initial system state is known with certainty, the initial condition for the CDF equation takes the form of the Heaviside function. The latter needs to be regularized to ensure stable solutions, e.g., via the use of a Dirac-delta



(a) $a \sim \mathcal{U}([0.5, 1.5])$

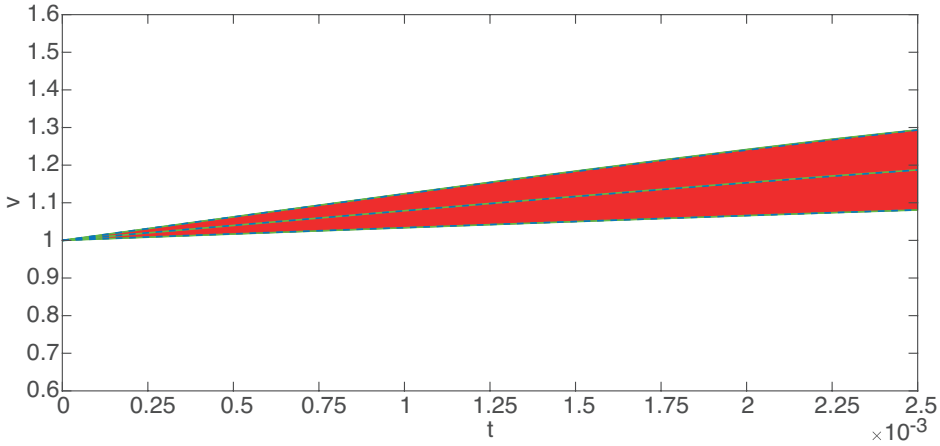


(b) $a \sim \mathcal{N}(1, 0.15)$

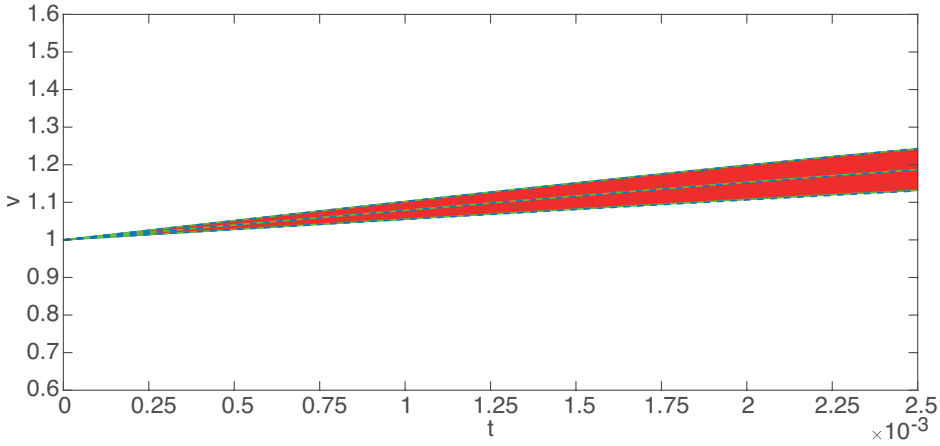


(c) $a \sim \mathcal{B}(2, 5)$

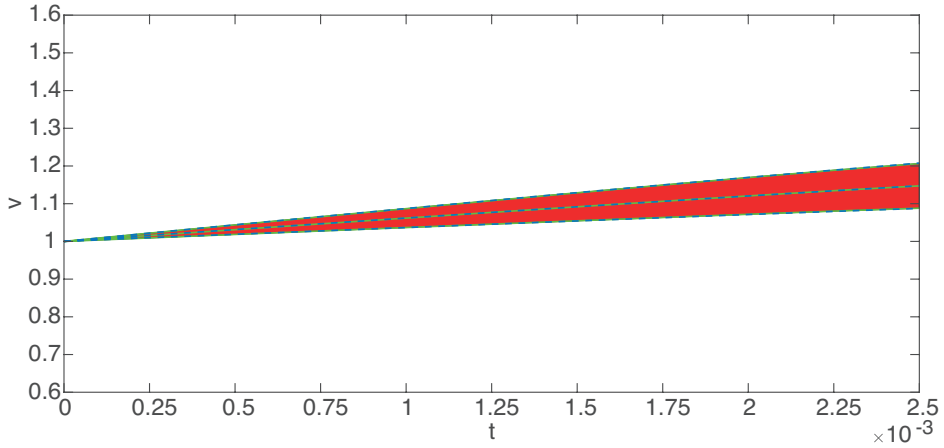
FIG. 13: Prediction of the flow velocity $v(x, t = 2.5 \times 10^{-3})$ and the corresponding two standard deviations uncertainty bound, for (a) uniform, (b) Gaussian, and (c) beta distributions of a. Dotted lines: MC. Blue dotted lines: CDF. Centerline is the mean. Color version online.



(a) $a \sim \mathcal{U}([0.5, 1.5])$



(b) $a \sim \mathcal{N}(1, 0.15)$



(c) $a \sim \mathcal{B}(2, 5)$

FIG. 14: Prediction of the flow velocity $v(x = 0.03, t)$ and the corresponding two standard deviations' uncertainty bound, for (a) uniform, (b) Gaussian, and (c) beta distributions of a. Dotted lines: MC. Blue dotted lines: CDF. Centerline is the mean. Color version online.

polynomial kernel [30]. Our numerical experiments showed that a (purely numerical) boundary singularity appears in the CDF. An exponential filter [33] was applied to the solution after every time step to suppress this behavior.

We compared solutions of a simplified CDF equation with a positive source and a single random coefficient, with MC simulation. Chebyshev collocation was used for the spatial discretization of the PDE, and a suitable higher-order Runge-Kutta method was employed for time marching.

The CDF method accurately predicts the mean and standard deviation of the QoI, and is able to approximate the PDF of the QoI outside of the regularization zone, preserving the main characteristics of the PDF. However, under/overshoots generated by the regularization process make a thorough error analysis difficult.

For negative sources, an adjusted version of the method is needed to avoid severe instabilities within a few time steps. Instead of the joint CDF F_{av}, one can consider the variable $G_{av} = F_a - F_{av}$, which is not a CDF, but does satisfy the CDF equation with adjusted initial and boundary conditions. Results are again in good agreement with MC simulation.

Future work will focus on extending the approach to more complex and general systems, such as coupled gas-particle models, and to implementing a data-driven learning approach to decrease the uncertainty in the system.

ACKNOWLEDGMENTS

We thank Eindhoven University of Technology, and in particular Dr. Ten Thije Boonkkamp for arranging R.J.L. Rutjens' internship at SDSU. We appreciate the useful feedback given by two anonymous reviewers. This research was supported by the Computational Mathematics program of AFOSR (Grants Nos. FA9550-19-1-0387 and FA9550-18-1-0474).

REFERENCES

1. Najm, H.N., Debusschere, B.J., Marzouk, Y.M., Widmer, S., and Le Maître, O.P., Uncertainty Quantification in Chemical Systems, *Int. J. Numer. Methods Eng.*, **80**(6-7):789–814, 2009.
2. Cai, L., Pitsch, H., Mohamed, S.Y., Raman, V., Bugler, J., Curran, H., and Sarathy, S.M., Optimized Reaction Mechanism Rate Rules for Ignition of Normal Alkanes, *Combust. Flame*, **173**:468–482, 2016.
3. Reagan, M.T., Najm, H.N., Pébay, P.P., Knio, O.M., and Ghanem, R.G., Quantifying Uncertainty in Chemical Systems Modeling, *Int. J. Chem. Kinet.*, **37**(6):368–382, 2005.
4. Boso, F., Marzadri, A., and Tartakovsky, D.M., Probabilistic Forecasting of Nitrogen Dynamics in Hyporheic Zone, *Water Resour. Res.*, **54**(7):4417–4431, 2018.
5. Jacobs, G.B. and Udaykumar, H.S., Uncertainty Quantification in Eulerian-Lagrangian Simulations of (Point-)Particle-Laden Flows with Data-Driven and Empirical Forcing Models, *Int. J. Multiphase Flow*, **121**:103–114, 2019.
6. Shotorban, B., Jacobs, G.B., Ortiz, O., and Truong, Q., An Eulerian Model for Particles Nonisothermally Carried by a Compressible Fluid, *Int. J. Heat Mass Transf.*, **65**:845–854, 2013.
7. Guerra, G.M., Zio, S., Camata, J.J., Dias, J., Elias, R.N., Mattoso, M., Paulo, P.L., Alvaro, A.L., and Rochinha, F.A., Uncertainty Quantification in Numerical Simulation of Particle-Laden Flows, *Comput. Geosci.*, **20**(1):265–281, 2016.
8. Giles, M.B., Multilevel Monte Carlo Path Simulation, *Oper. Res.*, **56**(3):607–617, 2008.
9. Iman, R.L. and Conover, W.J., Small Sample Sensitivity Analysis Techniques for Computer Models with an Application to Risk Assessment, *Commun. Stat. Theory Methods*, **9**(17):1749–1842, 1980.
10. Owen, A.B., *Monte Carlo Theory, Methods and Examples*, accessed from https://statweb.stanford.edu/∼owen/mc/Ch-var-adv.pdf, 2013.
11. Wiener, N., The Homogeneous Chaos, *Am. J. Math.*, **60**(4):897–936, 1938.
12. Stefanou, G., The Stochastic Finite Element Method: Past, Present and Future, *Comput. Methods Appl. Mech. Eng.*, **198**(9-12):1031–1051, 2009.
13. Wan, X. and Karniadakis, G.E., An Adaptive Multielement Generalized Polynomial Chaos Method for Stochastic Differential Equations, *J. Comput. Phys.*, **209**(2):617–642, 2005.
14. Wan, X. and Karniadakis, G.E., Multi-Element Generalized Polynomial Chaos for Arbitrary Probability Measures, *SIAM J. Sci. Comput.*, **28**(3):901–928, 2006.

15. Debusschere, B., Intrusive Polynomial Chaos Methods for Forward Uncertainty Propagation, in *Handbook of Uncertainty Quantification*, R. Ghanem, H. Owhadi, and D. Higdon, Eds., Switzerland: Springer, pp. 617–636, 2017.
16. Xiu, D., Stochastic Collocation Methods: A Survey, in *Handbook of Uncertainty Quantification*, R. Ghanem, H. Owhadi, and D. Higdon, Eds., Switzerland: Springer, pp. 617–636, 2017.
17. Ma, X. and Zabaras, N., An Adaptive Hierarchical Sparse Grid Collocation Algorithm for the Solution of Stochastic Differential Equations, *J. Comput. Phys.*, **228**(8):3084–3113, 2009.
18. Barajas-Solano, D. and Tartakovsky, D.M., Stochastic Collocation Methods for Nonlinear Parabolic Equations with Random Coefficients, *SIAM/ASA J. Uncert. Quantif.*, **4**(1):475–494, 2016.
19. Pope, S.B., *Turbulent Flows*, Cambridge, UK: Cambridge University Press, 2000.
20. Lichtner, P.C. and Tartakovsky, D.M., Stochastic Analysis of Effective Rate Constant for Heterogeneous Reactions, *Stoch. Environ. Res. Risk Assess.*, **17**(6):419–429, 2003.
21. Tartakovsky, D.M., Dentz, M., and Lichtner, P.C., Probability Density Functions for Advective-Reactive Transport in Porous Media with Uncertain Reaction Rates, *Water Resour. Res.*, **45**:W07414, 2009.
22. Broyda, S., Dentz, M., and Tartakovsky, D.M., Probability Density Functions for Advective-Reactive Transport in Radial Flow, *Stochastic Environ. Res. Risk Assess.*, **24**(7):985–992, 2010.
23. Dentz, M. and Tartakovsky, D.M., Probability Density Functions for Passive Scalars Dispersed in Random Velocity Fields, *Geophys. Res. Lett.*, **37**:L24406, 2010.
24. Wang, P., Tartakovsky, A.M., and Tartakovsky, D.M., Probability Density Function Method for Langevin Equations with Colored Noise, *Phys. Rev. Lett.*, **110**(14):140602, 2013.
25. Zhao, X., Kolla, H., Zhang, P., Wu, B., Calello, S., and Najm, H.N., A Transported Probability Density Function Method to Propagate Chemistry Uncertainty in Reacting Flow CFD, in *Proc. of AIAA Scitech 2019 Forum*, AIAA, San Diego, CA, Jan. 7–11, 2019.
26. Venturi, D. and Karniadakis, G.E., New Evolution Equations for the Joint Response-Excitation Probability Density Function of Stochastic Solutions to First-Order Nonlinear PDEs, *J. Comput. Phys.*, **231**(21):7450–7474, 2012.
27. Malakhov, A.N. and Saichev, A.I., Kinetic Equations in the Theory of Random Waves, *Radiophys. Quantum Electron.*, **17**(5):526–534, 1974.
28. Cho, H., Venturi, D., and Karniadakis, G.E., Statistical Analysis and Simulation of Random Shocks in Stochastic Burgers Equation, *Proc. R. Soc. A*, **470**:20140080, 2014.
29. Boelens, A.M.P., Venturi, D., and Tartakovsky, D.M., Parallel Tensor Methods for High-Dimensional Linear PDEs, *J. Comput. Phys.*, **375**(12):519–539, 2018.
30. Suarez, J., Jacobs, G.B., and Don, W.S., A High-Order DIRAC-Delta Regularization with Optimal Scaling in the Spectral Solution of One-Dimensional Singular Hyperbolic Conservation Laws, *J. Sci. Comput.*, **36**(4):A1831–A1849, 2014.
31. Wissink, B.W., Jacobs, G.B., Ryan, J.K., Don, W.S., and van der Weide, E.T.A., Shock Regularization with Smoothness-Increasing Accuracy-Conserving Dirac-Delta Polynomial Kernels, *J. Sci. Comput.*, **77**(1):579–596, 2018.
32. Gottlieb, D. and Hesthaven, J.S., Spectral Methods for Hyperbolic Problems, *J. Comput. Appl. Math.*, **128**:83–131, 2001.
33. Hesthaven, J.S., Gottlieb, S., and Gottlieb, D., *Spectral Methods for Time-Dependent Problems, Cambridge Monographs on Applied and Computational Mathematics*, Cambridge, UK: Cambridge University Press, 2007.
34. Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., Numerical Recipes for Fortran 77, in *The Art of Scientific Computing*, Vol. 1, 2nd ed., Cambridge, UK: Cambridge University Press, 1992.
35. Gottlieb, S. and Shu, C., Total Variation Diminishing Runge-Kutta Schemes, *Math. Comput.*, **67**(221):73–85, 1998.
36. Vandeven, H., Family of Spectral Filters for Discontinuous Problems, *J. Sci. Comput.*, **6**(2):159–192, 1991.
37. MathWorks, Inc., Matlab Statistics and Machine Learning Toolbox User's Guide R2019a, pp. 3603–3629, 2019.
38. Silverman, B.W., *Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability*, London, UK: Chapman and Hall, 1986.
39. Tornberg, A.K., Multi-Dimensional Quadrature of Singular and Discontinuous Functions, *BIT Numer. Math.*, **42**:644–669, 2002.
40. Suarez, J. and Jacobs, G.B., Regularization of Singularities in the Weighted Summation of DIRAC-Delta Functions for the Spectral Solution of Hyperbolic Conservation Laws, *J. Sci. Comput.*, **72**(3):1080–1092, 2017.

41. Alawadhi, A.A., Boso, F., and Tartakovsky, D.M., Method of Distributions for Water-Hammer Equations with Uncertain Parameters, *Water Resour. Res.*, **54**(11):9398–9411, 2018.
42. Boso, F., Broyda, S.V., and Tartakovsky, D.M., Cumulative Distribution Function Solutions of Advection-Reaction Equations with Uncertain Parameters, *Proc. R. Soc. A*, **470**(2166):20140189, 2014.
43. Yang, H.J., Boso, F., Tchelepi, H.A., and Tartakovsky, D.M., Probabilistic Forecast of Single-Phase Flow in Porous Media with Uncertain Properties, *Water Resour. Res.*, **55**(11):8631–8645, 2019.
44. Giles, M.B., Nagapetyan, T., and Ritter, K., Multilevel Monte Carlo Approximation of Distribution Functions and Densities, *SIAM/ASA J. Uncert. Quantif.*, **3**:267–295, 2015.
45. Taverniers, S. and Tartakovsky, D.M., Estimation of Distributions via Multilevel Monte Carlo with Stratified Sampling, *J. Comput. Phys.*, vol. **419**, p. 109572, 2020.

APPENDIX A. DERIVATION OF THE JOINT CDF EQUATION (POSITIVE SOURCE)

In addition to the two random functions $\mathbf{a}$ and $v(x,t)$, we consider a fine-grained CDF,

$$\Pi(\mathbf{A},\mathbf{a};V,v)\equiv\mathcal{H}(\mathbf{A}-\mathbf{a})\mathcal{H}(V-v(x,t)), \tag{A.1}$$

where $\mathbf{A}$ and V are deterministic variables, and $\mathcal{H}(\cdot)$ is the Heaviside function. The ensemble mean of any integrable function $g(\mathbf{a},v)$ of random variables $\mathbf{a}\in\mathbb{R}^{N_g}$ and $v\in\mathbb{R}$ with the joint PDF $f_{\mathbf{a}v}(\mathbf{A}',V'):\mathbb{R}^{N_g}\times\mathbb{R}\to\mathbb{R}^+$ is

$$\mathbb{E}[g(\mathbf{a},v)]=\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}} g(\mathbf{A}',V')f_{\mathbf{a}v}(\mathbf{A}',V')\mathrm{d}\mathbf{A}'\mathrm{d}V'. \tag{A.2}$$

In particular, at any space-time point (x,t), the ensemble mean of $\Pi(\mathbf{A},\mathbf{a};V,v)$ over random realizations of the random variables $\mathbf{a}$ and v is

$$\begin{aligned}\mathbb{E}[\Pi]&=\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}}\Pi(\mathbf{A},\mathbf{A}';V,V')f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'\\&=\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}}\mathcal{H}[\mathbf{A}-\mathbf{A}']\mathcal{H}[V-V']f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'\\&=\int_{-\infty}^{V}\int_{-\infty}^{A_1}\cdots\int_{-\infty}^{A_{N_g}}f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'=F_{\mathbf{a}v}(\mathbf{A},V;x,t),\end{aligned} \tag{A.3}$$

where $F_{\mathbf{a}v}$ is the joint cumulative distribution function (CDF) for $\mathbf{a}=\{a_0,\ldots,a_{N_g}\}$ and v at any space-time point (x,t). This property suggests a two-step procedure for the derivation of a PDE for $F_{\mathbf{a}v}$. First, we derive an equation for Π. Then, we average this equation. It follows from (A.1) and the sifting property of the Dirac-delta function that

$$\frac{\partial\Pi}{\partial t}=-\frac{\partial\Pi}{\partial V}\frac{\partial v}{\partial t},\qquad\frac{\partial\Pi}{\partial x}=-\frac{\partial\Pi}{\partial V}\frac{\partial v}{\partial x},\qquad\text{and}\qquad g(v)\frac{\partial\Pi}{\partial V}=g(V)\frac{\partial\Pi}{\partial V}, \tag{A.4}$$

for any test function $g(v)$. Hence, multiplication of the stochastic Burgers equation,

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}=(u-v)\sum_{i=0}^{N_g}a_i\mathrm{T_i}(u-v), \tag{A.5}$$

with $-\partial\Pi/\partial V$ yields

$$\frac{\partial\Pi}{\partial t}+V\frac{\partial\Pi}{\partial x}=-\frac{\partial\Pi}{\partial V}(u-V)\sum_{i=0}^{N_g}a_iT_i(u-V). \tag{A.6}$$

This is a linear PDE with the random (constant) coefficients a_0, ..., a_{N_g}. (Recall that both u and $V$are deterministic). By virtue of Eqs. (A.2) and (A.3), the ensemble average of this PDE is

$$\begin{aligned}\frac{\partial F_{\mathbf{a}v}}{\partial t}+V\frac{\partial F_{\mathbf{a}v}}{\partial x} &= -\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}}\frac{\partial \Pi}{\partial V}(u-V)\sum_{i=0}^{N_g}A_i' T_i(u-V) f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V' \\ &= -(u-V)\sum_{i=0}^{N_g}\left[\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}}\frac{\partial \Pi}{\partial V}A_i' f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'\right]T_i(u-V) \\ &= -(u-V)\sum_{i=0}^{N_g}\left[\frac{\partial}{\partial V}\int_{\mathbb{R}}\int_{\mathbb{R}^{N_g}}\Pi A_i' f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'\right]T_i(u-V). \end{aligned} \tag{A.7}$$

For the left-hand side, we used the following theorem and lemma. Their proofs are omitted.

Theorem 1 (Dominated convergence)**.** *Let X_n be a sequence of integrable random variables and let the limit $\lim_{n\to\infty}X_n(\omega)=X(\omega)$ exist for all $\omega\in\Omega$. If there is a nonnegative random variable Y such that $|X_n(\omega)|\leq Y(\omega)$ for all $\omega\in\Omega$ and all n, then X is integrable and $\lim_{n\to\infty}\mathbb{E}[X_n]=\mathbb{E}[X]$.*

Lemma 1. *Let $X\in\mathcal{X}$ be a random variable and $g:\mathbb{R}\times\mathcal{X}\to\mathbb{R}$ a function such that $g(X,t)$ is integrable for all t, and g is differentiable with respect to t. Assume that there is a random variable Y such that $|\partial_t g(X,t)|\leq Y$ a.s. for all t, and $\mathbb{E}[Y]<\infty$. Then $\partial_t\mathbb{E}[g(X,t)]=\mathbb{E}[\partial_t g(X,t)]$.*

Applied to our case, it follows that

$$\mathbb{E}\left[\frac{\partial \Pi}{\partial t}+V\frac{\partial \Pi}{\partial x}\right]=\frac{\partial \mathbb{E}(\Pi)}{\partial t}+V\frac{\partial \mathbb{E}(\Pi)}{\partial x}=\frac{\partial F_{\mathbf{a}v}}{\partial t}+V\frac{\partial F_{\mathbf{a}v}}{\partial x}. \tag{A.8}$$

For the right-hand side of Eq. (A.7), we interchange summation and integration and apply Leibniz's integral rule to obtain the final result.

The definition of Π in terms of the Heaviside function implies that the integral in the last line on the right-hand side of Eq. (A.7) reduces to

$$\mathcal{I}=\int_{-\infty}^{V}\int_{-\infty}^{A_1}\cdots\int_{-\infty}^{A_{N_g}}A_i' f_{\mathbf{a}v}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V'. \tag{A.9}$$

Recalling the relationship between PDF and CDF, this yields

$$\begin{aligned}\mathcal{I} &= \int_{-\infty}^{V}\int_{-\infty}^{A_1}\cdots\int_{-\infty}^{A_{N_g}}A_i'\frac{\partial^{N_g+1}F_{\mathbf{a}v}}{\partial A_1'\cdots\partial A_{N_g}'\partial V'}(\mathbf{A}',V';x,t)\mathrm{d}\mathbf{A}'\mathrm{d}V' \\ &= \int_{-\infty}^{A_i}A_i'\frac{\partial F_{\mathbf{a}v}}{\partial A_i'}(\mathbf{A}\backslash A_i,A_i',V;x,t)\mathrm{d}A_i' \\ &= A_i F_{\mathbf{a}v}(\mathbf{A},V;x,t)-\int_{-\infty}^{A_i}F_{\mathbf{a}v}(\mathbf{A}\backslash A_i,A_i',V;x,t)\mathrm{d}A_i', \end{aligned} \tag{A.10}$$

where $\mathbf{A}\backslash A_i=\left(A_1,\ldots,A_{i-1},A_{i+1},\ldots,A_{N_g}\right)$. Combining the terms leads to

$$\frac{\partial F_{\mathbf{a}v}}{\partial t}+V\frac{\partial F_{\mathbf{a}v}}{\partial x}=-(u-V)\sum_{i=0}^{N_g}T_i(u-V)\frac{\partial}{\partial V}\left[A_i F_{\mathbf{a}v}-\int_{-\infty}^{A_i}F_{\mathbf{a}v}(\mathbf{A}\backslash A_i,A_i',V;x,t)\mathrm{d}A_i'\right]. \tag{A.11}$$