

DATA-INFORMED METHOD OF DISTRIBUTIONS FOR HYPERBOLIC CONSERVATION LAWS*

FRANCESCA BOSO[†] AND DANIEL M. TARTAKOVSKY[†]

Abstract. Nonlinear hyperbolic balance laws with uncertain (random) initial data are ubiquitous in a plethora of transport phenomena that often exhibit shocks. We develop the method of distributions for such problems by adding a model error term to a deterministic equation for the cumulative distribution function (CDF) of the system states. We use two alternative strategies, Newtonian relaxation and neural networks, to infer this term from observations of the system dynamics. The former strategy is amenable to theoretical analysis of its convergence with respect to data sparsity, while the latter offers more flexibility. The CDF equation is exact for linear conservation laws and nonlinear conservation laws with a smooth solution, such that the CDF equation can be used to formulate predictions at times when observations cease to be available. Whenever shocks develop as a result of the nonlinearity, observations are used to detect the discrepancy that emerges as model error. Spatial data density is crucial for good interpolation accuracy, whereas long temporal sequences of observations improve future projections.

Key words. uncertainty quantification, data driven, traffic modeling

AMS subject classifications. 35R60, 35L65, 35L60

DOI. 10.1137/19M1260773

1. Introduction. Nonlinear hyperbolic conservation (or balance) laws are ubiquitous in science and engineering, where they are used to describe a plethora of transport phenomena across a multiplicity of scales, ranging from neurofilament transport in axons [8] to multiphase flow in petroleum reservoirs [33]. Coefficients and constitutive laws in these partial differential equations (PDEs) are typically uncertain because, among other reasons, they cannot be measured directly. Instead, they are either defined probabilistically or inferred from measurements of the state variable(s) whose dynamics are described by these PDEs. The former approach is the domain of uncertainty quantification, while the latter falls under the purview of data assimilation, which recently became known as “machine learning of PDEs.”

In both contexts, one needs a computationally efficient and robust method to solve hyperbolic conservation laws with random coefficients and/or initial conditions. Even for hyperbolic problems admitting a smooth solution, most existing (intrusive and nonintrusive) techniques often underperform brute-force Monte Carlo when the integration time or the problem’s stochastic dimension (i.e., the number of random variables needed to describe random inputs) is large. Furthermore, these techniques break down for problems with nonsmooth solutions (solutions with contact discontinuities and shocks) [17, Preface] and require adjustments.

Arguably, the sole exception to the rule is the method of distributions [27], which is tailor-made for first-order hyperbolic PDEs with random coefficients and does not suffer from the so-called curse of dimensionality. It often results in exact deterministic

*Submitted to the journal’s Methods and Algorithms for Scientific Computing section May 9, 2019; accepted for publication (in revised form) December 2, 2019; published electronically February 27, 2020.

<https://doi.org/10.1137/19M1260773>

Funding: This work was partially supported by the Defense Advanced Research Project Agency under award 101513612 (Lagrange program) and by the Air Force Office of Scientific Research under award FA9550-17-1-0417.

[†]Energy Resources Engineering, Stanford University, Stanford, CA 94305 (fboso@stanford.edu, tartakovsky@stanford.edu).

equations for either a probability density function (PDF) or a cumulative distribution function (CDF) of the state variable whose dynamics are governed by a hyperbolic PDE with uncorrelated (white noise) inputs. In that sense, it is complementary to its expansion-based (e.g., Karhunen–Loève transform) alternatives, whose efficacy vis-à-vis Monte Carlo deteriorates as the inputs’ correlation length decreases. Deterministic PDF and CDF equations have been derived to propagate parametric uncertainty in a wide range of (systems of) parabolic [6, 30] and hyperbolic [27, 32] nonlinear PDEs with smooth solutions. The method of distributions has also been used for uncertainty quantification in hyperbolic PDEs with discontinuous solutions, e.g., the Buckley–Leverett [33] and water hammer [2] equations, by partitioning their domain of definition into smooth-solution regions. The latter step requires the ability to track analytically the discontinuity locations, which is seldom possible. A more general approach exploits the conceptual similarity between the method of distributions and the kinetic formulation of deterministic hyperbolic conservation laws [22, 23]. This possibility was promulgated in [33] and, more recently, realized by combining the kinetic formulation with polynomial chaos expansions [10].

The kinetic formulation accounts for the presence of shocks by adding the so-called kinetic defect term to a linear transport equation. While the existence and uniqueness of this term have been proved for rather general conditions (see, e.g., [23] and the references therein), it is computable for a few particular cases such as step initial conditions. In all other instances, the kinetic defect has to be informed by, or “learned” from, data. Inference of the kinetic defect term in CDF equations is the major focus of our study. The techniques presented here can be employed to characterize structural (model) error.

In section 2 we develop the method of distributions for hyperbolic balance laws with shocks by adding a kinetic defect-like term to a deterministic equation for the CDF of the system states. We treat this term as a structural (model) error and infer it from the system-state measurements by using, alternatively, the Newtonian relaxation (section 3) and neural networks (section 4). The former strategy is amenable to theoretical analysis of its convergence with respect to data sparsity (Appendix B), while the latter offers more flexibility. We investigate the performance of both numerically by using a nonlinear traffic flow model as our computational testbed (section 5). The results of these numerical experiments are collated in section 6. Main conclusions drawn from this study are summarized in section 7.

2. Method of distributions for hyperbolic conservation laws. While the method presented below can handle, with some modification, PDEs with random coefficients [26, 27, 30], we focus on deterministic PDEs subject to uncertain (random) initial conditions. This problem continues to present challenges for various uncertainty quantification techniques [18], especially in the presence of discontinuous solutions (shocks) and poorly correlated inputs.

Consider a scalar state variable $u(\mathbf{x}, t) : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \Omega_U \subset \mathbb{R}$ whose evolution in d -dimensional space ($\mathbf{x} \in \mathbb{R}^d$) and time ($t \in \mathbb{R}^+$) is described by a hyperbolic PDE,

$$(2.1a) \quad \frac{\partial u}{\partial t} + \nabla \cdot \mathbf{q}(u) = 0, \quad \mathbf{x} \in \mathbb{R}^d, \quad t \in \mathbb{R}^+,$$

where the components of the d -dimensional flux $\mathbf{q}(u)$ are continuous functions of their argument. The spatial distribution of $u(\mathbf{x}, t)$ at time $t = 0$ is uncertain, so that this equation is subject to the initial condition

$$(2.1b) \quad u(\mathbf{x}, t = 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $u_0(\mathbf{x})$ is a random field with a prescribed CDF $F_0(U; \mathbf{x})$. Our goal is to derive an evolution equation for the CDF of u , $F(U; \mathbf{x}, t) \equiv \mathbb{P}[u(\mathbf{x}, t) \leq U] : \Omega_U \times \mathbb{R}^d \times \mathbb{R}^+ \rightarrow [0, 1]$, i.e., to compute the probability that the random value of u at space-time point (\mathbf{x}, t) does not exceed a deterministic threshold $U \in \Omega_U = [U_{\min}, U_{\max}]$.

2.1. CDF equation for smooth solutions. Derivation of the CDF equation comprises two steps. First, we derive a PDE for $\Pi = \mathcal{H}(u(\mathbf{x}, t) - U)$, where $\mathcal{H}(\cdot)$ is the Heaviside function. For (2.1) with a smooth solution $u(\mathbf{x}, t)$ or in the regions where this solution is smooth, $\Pi(U; \mathbf{x}, t)$ satisfies a linear PDE (see Appendix A for details),

$$(2.2) \quad \frac{\partial \Pi}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \Pi = 0, \quad \dot{\mathbf{q}}(U) \equiv \frac{d\mathbf{q}(U)}{dU},$$

which is subject to the random initial condition $\Pi(U; \mathbf{x}, t = 0) = \mathcal{H}(u_0(\mathbf{x}) - U)$. Second, since the ensemble average of Π over u is $\mathbb{E}[\Pi] = 1 - F(U; \mathbf{x}, t)$, the ensemble average of (2.2) yields the CDF equation

$$(2.3) \quad \frac{\partial F}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla F = 0,$$

subject to the initial condition $F(U; \mathbf{x}, t = 0) = F_0(U; \mathbf{x})$, where $F_0 = 1 - \mathbb{E}[\Pi(U; \mathbf{x}, t = 0)]$.

If one were interested only in the probability of exceedence of a specific threshold U^* , $\mathbb{P}[u(x, t) > U^*] = 1 - F(U^*; \mathbf{x}, t)$, then the CDF equation would have to be solved for that specific value of U , resulting in a considerable saving of computational time.

Solution of the linear d -dimensional PDE (2.3), for any threshold value U , is to be compared with its counterpart obtained via Monte Carlo simulations (MCS) of the nonlinear d -dimensional PDE (2.1). MCS approximate the local CDF by postprocessing multiple realizations of $u(x, t)$, which are computed by solving (2.1) subject to multiple realizations of the random initial conditions u_0 . The accuracy and computational cost of MCS increase (sublinearly and linearly, respectively) with the number of MC realizations.

2.2. CDF equation for discontinuous solutions. Equations (2.1) and (2.3) break down at space-time points wherein $u(\mathbf{x}, t)$ is discontinuous, i.e., at shock locations. The kinetic formulation of deterministic scalar conservation laws [22] replaces (2.1) with its weak formulation and then identifies the one solution that fulfills entropy constraints. Specifically, the weak solution of the deterministic version of (2.1) subject to appropriate entropy constraints is equivalent to [22] the deterministic solution $\Pi = \mathcal{H}(u(\mathbf{x}, t) - U)$ of

$$(2.4) \quad \frac{\partial \Pi}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \Pi = \mathcal{M}(U; \mathbf{x}, t).$$

The “kinetic defect” term \mathcal{M} is introduced to account for the possibility of a shock in linear high-dimensional equations, such as (2.2) or (2.4). While the functional form of \mathcal{M} is generally unknown and, hence, is not computable, its definition is unique due to a set of mathematical properties, which are summarized in Appendix A for the sake of completeness, together with the derivation of (2.4). Regardless of the functional form of \mathcal{M} , the ensemble averaging of (2.4) yields a linear CDF equation that accounts for the presence of a shock,

$$(2.5) \quad \frac{\partial F}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla F = -\langle \mathcal{M}(U, \mathbf{x}, t) \rangle.$$

Its solution $F(U; \mathbf{x}, t)$ has the form

$$(2.6) \quad F(U; \mathbf{x}, t) = \underbrace{\int_{\mathbb{R}^d} F_0(U; \boldsymbol{\xi}) \mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t) d\boldsymbol{\xi}}_{F_{ss}(U; \mathbf{x}, t)} - \underbrace{\int_0^t \int_{\mathbb{R}^d} \langle \mathcal{M}(U, \boldsymbol{\xi}, \tau) \rangle \mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t - \tau) d\boldsymbol{\xi} d\tau}_{-F_{kd}(U; \mathbf{x}, t)},$$

where $\mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t - \tau) = \delta(\mathbf{x} - \boldsymbol{\xi} - \dot{\mathbf{q}}(U)(t - \tau))$. The first term, F_{ss} , is the CDF of the smooth $u(\mathbf{x}, t)$, corresponding to (2.5) with $\langle \mathcal{M} \rangle = 0$. The second term, F_{kd} , accounts for the shocks. Under certain conditions, e.g., for the initial discontinuity specified in subsection 5.1, both \mathcal{M} and $\langle \mathcal{M} \rangle$ are computable analytically. Otherwise, their analogs can be “learned” from data (either Monte Carlo realizations or measurements), as we show below. Data models provide observational CDFs that account for noisy observations/stochastic simulations.

3. Newtonian relaxation for inference of CDF equations from data.

Let $F_{\text{obs}}(U; \mathbf{x}, t) = H[F(U; \mathbf{x}, t)]$ denote observations of the exact CDF F of the random state $u(\mathbf{x}, t)$, identified by the observation map $H[\cdot]$. The data are available in a subset of space-time locations (\mathbf{x}, t) of the computational domain. Our first strategy for assimilating $F_{\text{obs}}(U; \mathbf{x}, t)$ into the CDF equation is to use the Newtonian (or dynamic) relaxation [25], which is also known as nudging,

$$(3.1) \quad \frac{\partial \hat{F}}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \hat{F} = \lambda(F_{\text{obs}} - \hat{F}), \quad \hat{F}(U; \mathbf{x}, t = 0) = \hat{F}_0(U; \mathbf{x}).$$

Here $\hat{F}(U; \mathbf{x}, t)$ represents an approximation of the exact CDF $F(U; \mathbf{x}, t)$, and the nudging coefficient λ controls the rate of convergence of \hat{F} towards F via dynamic data assimilation.

Remark 3.1. The choice of λ and, more generally, the practical implementations of nudging, are characterized by a certain level of empiricism, typically relying on manual tuning [21]. A review of the strategies to select λ can be found in [19]; in particular, one can account for data sparsity by letting λ vary in space and time. If λ is expressed as a function of the Kalman gain matrix, for example, then nudging reduces to Kalman filtering in the case of linear dynamics, Gaussian error statistics, and perfectly described error distributions [16]. Under these conditions, data assimilation via nudging represents a true posterior in the Bayesian sense. Alternatively, observations can be interpolated in space and time to increase the coverage of the domain [31].

Comparison of (3.1) and (2.5) suggests the possibility of using the Newtonian relaxation to “learn” the generally uncomputable term $\langle \mathcal{M}(U, \mathbf{x}, t) \rangle$ in the CDF equation (2.5) from data. We prove in Appendix B that the convergence of \hat{F} to F relies on the convergence in the mean of their random counterparts,

$$(3.2) \quad \|\hat{F}(t) - F(t)\|_{L_1(\mathbb{R}^d \times \Omega_U)} \leq \mathbb{E}[\|\hat{\Pi}(t) - \Pi(t)\|_{L_1(\mathbb{R}^d \times \Omega_U)}],$$

where $\hat{\Pi}$ is a solution of

$$\frac{\partial \hat{\Pi}}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \hat{\Pi} = \lambda(H[\Pi] - \hat{\Pi}), \quad \hat{\Pi}(U; \mathbf{x}, t = 0) = \hat{\Pi}_0(U; \mathbf{x}).$$

The convergence is expressed in terms of the L_1 norm of the discrepancy in the space $\mathbb{R}^d \times \Omega_U$. We summarize in Appendix B the convergence rates for $\|\hat{\Pi}(t) -$

$\Pi(t)\|_{L_1(\mathbb{R}^d \times \Omega_U)}$, established in [7, 23] for different observation models $H[\cdot]$; together with the corresponding rates for \hat{F} . Assimilation of data into models based on hyperbolic conservation laws is challenging because a disturbance cannot be detected at a fixed location until the wave has reached it, i.e., measurements collected in the regions of the space-time domain not affected by the waves propagating the initial information do not reduce predictive uncertainty. For this reason, convergence bounds can be explicitly obtained as a function of temporal sparsity, whereas spatial sparsity bounds are affected by observability limitations pointed out in Appendix B.4.

The nudged dynamics in (3.1) is both physics-informed (via $\hat{\mathbf{q}}$ and the conservation law) and data-driven (because of F_{obs}), and provides a seamless way to combine both sources of information. Data assimilation via nudging dynamically relaxes the approximated solution \hat{F} towards F , reducing both (possible) initial error $\|\hat{F}_0 - F_0\|_{L_1(\mathbb{R}^d \times \Omega_U)}$ and possible error in the dynamic model (3.1) due to the undefined $\langle \mathcal{M} \rangle$ term. The initial error can always be reduced by repeated assimilation of the observations (e.g., with the Back and Forth Nudging (BFN) strategies described in Appendix C), whereas characterization of the model error holds only within the observation window. When observations cease, (3.1) relaxes towards the smooth-case solution.

As with any data assimilation procedure, as observations are assimilated there is no guarantee that the properties of the solution are maintained. In the case of the CDF equation (3.1), there is no intrinsic guarantee of \hat{F} being monotonic, especially when observations are sparse. In order to overcome this problem and in order to reduce the relaxation time while preserving the physics-informed and data-driven nature of dynamic relaxation, we introduce instantaneous relaxation via neural networks.

4. Neural networks for inference of CDF equations from data. We reformulate the nudging relaxation as a minimization problem defined at all N_{obs} observation locations $\{\mathbf{X}, \mathbf{T}\}$ in the form of

$$(4.1) \quad \mathcal{L} = \|\hat{F}(\mathbf{U}, \mathbf{X}, \mathbf{T}) - F_{\text{obs}}(\mathbf{U}, \mathbf{X}, \mathbf{T})\|_{L_p(\mathbb{R}^{N_{\text{obs}}} \times N_U)},$$

where $\|\cdot\|_{L_p(\mathbb{R}^{N_{\text{obs}}} \times N_U)}$ is a suitable norm to compare point values of CDFs observed in N_{obs} space-time locations at a finite set of N_U points U_n ($n = 1, \dots, N_U$) in Ω_U ; following common practice, we set $p = 2$ in the simulations reported below. The loss function (4.1) encapsulates the discrepancy between the predicted (approximated) CDF value, \hat{F} , and the observed CDF values, F_{obs} , at points $(\mathbf{U}, \mathbf{X}, \mathbf{T})$. To facilitate the minimization procedure, we take advantage of the linearity of the CDF dynamics in (2.5) by using (2.6) to express its solution as $\hat{F} = F_{\text{ss}} + \hat{F}_{\text{kd}}$.

The term F_{ss} corresponds to the smooth-solution dynamics. It is exact whenever the CDF of the initial state is known with certainty, $\hat{F}_0 = F_0$; otherwise, its estimate can be refined by employing the BFN described in Appendix C. The term \hat{F}_{kd} represents the (undefined) contribution of the kinetic defect to the solution. It is approximated with a Deep Neural Network (NN) with N_l layers,

$$\begin{aligned} \hat{F}_{\text{kd}}(U; \mathbf{x}, t) &= \varphi \circ \mathbf{A}^{(N_l)} \circ \dots \circ \varphi \circ \mathbf{A}^{(2)} \circ \varphi \circ \mathbf{A}^{(1)} \tilde{\mathbf{X}}, \\ \mathbf{A}^{(i)} \tilde{\mathbf{X}} &= \mathbf{W}^{(i)} \tilde{\mathbf{X}} + \mathbf{b}^{(i)}, \end{aligned}$$

whose parameters (weights \mathbf{W} and biases \mathbf{b}) serve as decision variables in the minimization problem (4.1). The activation function φ is applied recursively to all components (\circ) of the previous layer contribution. The input vector $\tilde{\mathbf{X}} = [\mathbf{U}, \mathbf{X}, \mathbf{T}]$ encompasses the observation points' coordinates. This choice of an approximator is

dictated by a NN's ability to represent nonperiodic, piecewise continuous functions exhibiting steep gradients [15], which is relevant because the exact counterpart of \hat{F}_{kd} is nonzero only in localized parts of the space-time domain in which shocks arise. Moreover, automatic differentiation of \hat{F} yields maps of the kinetic defect with no additional effort. If a training dataset is sufficiently large to span multiple scenarios, training of NNs would create a variety of responses, thus enabling prediction of $\langle \mathcal{M} \rangle$ when similar conditions arise.

Remark 4.1. Our numerical experiments did not necessitate imposition of regularization or additional constraints on the shape of the approximated solution \hat{F} . Certain desirable features of \hat{F} , such as monotonicity and boundary values $\hat{F}(U_{\min}) = 0$ and $\hat{F}(U_{\max}) = 1$, are introduced via observations (i.e., training set) and are mostly retained by the reconstructed neural network. If necessary, such constraints can be explicitly introduced in the formulation of the loss function (4.1) with little added complexity.

We are aware of no theoretical analysis of the convergence of NNs. Instead, we rely on the theoretical convergence results for Newtonian relaxation (Appendix B) to provide a robust indicator of convergence. This comes with two caveats: in this approach observations are simultaneously assimilated (as opposed to dynamically introduced), and relaxation is enforced in the form of the L_2 norm (rather than the L_1 norm).

5. Numerical experiments. Let $u(x, t) : \mathbb{R} \times \mathbb{R}^+ \rightarrow [0, 1]$ denote a state variable that satisfies a one-dimensional hyperbolic conservation law,

$$(5.1) \quad \frac{\partial u}{\partial t} + \frac{\partial q(u)}{\partial x} = 0, \quad q(u) = u(1 - u); \quad x \in \mathbb{R}, \quad t \in \mathbb{R}^+,$$

subject to the initial conditions $u(x, t = 0) = u_0(x)$ for $x \in \mathbb{R}$. Among other applications, (5.1) describes vehicular traffic far from congestion, with $u(x, t)$ representing the macroscopic car density $\rho(x, t)$ normalized with the maximum density ρ_{\max} . In this context, $q(u)$ is the Greenshields flux $q(\rho) = u_{\max}\rho(1 - \rho/\rho_{\max})$ [12], normalized with the maximum traffic velocity u_{\max} , and x and t are the space and time coordinates normalized with a characteristic length L and advective timescale L/u_{\max} , respectively.

Equation (5.1) is solved for both a randomized Riemann problem formulated in subsection 5.1 and a random initial condition,

$$(5.2) \quad u_0(x) = \begin{cases} \frac{m - l}{1 + \exp[-50(x - x_L)]} + l, & x \leq x_d, \\ \frac{m - r}{1 + \exp[50(x - x_R)]} + r, & x > x_d, \end{cases}$$

which describes a bottleneck with random intensities $m, l, r \in \mathbb{R}^+$, such that an event $\{m > l, r, x_L < x_d < x_R, x_L = x_d - (x_R - x_L), x_R = x_d + (x_R - x_L)\}$ corresponds to a bottleneck formed on the interval $x_L \leq x \leq x_R$. The former setting allows analytical treatment and serves to validate our method, while the latter demonstrates the method's ability to learn the model error and, hence, the CDF equation from data. Observations represent knowledge of the CDF that is provided by measurements of the physical process with respect to which we constrain the dynamic evolution of the CDF equation.

5.1. Riemann problem. The Riemann problem (RP) consists of (5.1) subject to a piecewise constant initial condition,

$$(5.3) \quad u_0(x) = \begin{cases} u_L, & x \leq x_d \\ u_R, & x > x_d, \end{cases}$$

for the given (random) constants u_L and u_R , and a given position of the initial discontinuity x_d . The solution of this RP is either a rarefaction wave or a shock solution depending on whether $q(u)$ is concave or convex. For the concave flux $q(u) = u(1 - u)$, this solution takes the form

$$(5.4a) \quad u(x/t) = \begin{cases} u_L, & x/t \leq \dot{q}(u_L), \\ u^*(x/t), & \dot{q}(u_L) < x/t < \dot{q}(u_R), \\ u_R, & x/t \geq \dot{q}(u_R) \end{cases} \quad \text{if } \dot{q}(u_L) < \dot{q}(u_R),$$

and

$$(5.4b) \quad u(x/t) = \begin{cases} u_L, & x/t < v_s, \\ u_R, & x/t > v_s \end{cases} \quad \text{if } \dot{q}(u_L) \geq \dot{q}(u_R).$$

The shock speed $v_s = [q(u_R) - q(u_L)] / (u_R - u_L)$ is defined via the Rankine–Hugoniot condition, which introduces the entropy constraint. The solution in the rarefaction fan $u^*(x/t)$ is obtained by inverting $\dot{q}(u^*) = x/t$.

The RP is the basis of many numerical studies of nonlinear hyperbolic conservation laws. It represents the building block of numerical methods solving nonlinear PDEs in general, and the cell transmission model of traffic [9] in particular. This configuration also has practical significance: discontinuities along a road might originate from traffic lights, accidents, or changes in the number of lanes.

5.2. CDF equations and data assimilation. The analytical solution (5.4) enables one to compute $F(U; x, t)$, the CDF of $u(x, t)$, in terms of the PDF of u_L and u_R , $f_{u_L}(U_L)$ and $f_{u_R}(U_R)$, respectively. We use this exact CDF to analyze the accuracy of the one-dimensional version of the CDF equation (2.5),

$$(5.5) \quad \frac{\partial F}{\partial t} + \dot{q}(U) \frac{\partial F}{\partial x} = -\langle \mathcal{M}(U, x, t) \rangle, \quad \dot{q}(U) = 1 - 2U,$$

subject to the initial condition $F(U; x, t = 0) = F_0(U; x)$; we use it as well as to assess the performance of the nudging assimilation equation (3.1),

$$(5.6) \quad \frac{\partial \hat{F}}{\partial t} + \dot{q}(U) \frac{\partial \hat{F}}{\partial x} = \lambda(F_{\text{obs}} - \hat{F}), \quad \hat{F}(U; x, t = 0) = \hat{F}_0(U; x),$$

and the accuracy of the NN reconstruction of the solution.

The latter tasks are facilitated by the fact that, for any realization of u_L and u_R , the kinetic defect \mathcal{M} for the RP (5.1) and (5.3) is given by [23, sect. 3.3]

$$(5.7) \quad \mathcal{M} = [\dot{q}(U) - v_s] \delta(x - v_s t) [\mathcal{H}(u_L - U) - \mathcal{H}(u_R - U)] \mathcal{H}(u_R - u_L),$$

and its ensemble mean by

$$(5.8) \quad \langle \mathcal{M} \rangle = \int \int \mathcal{M}(U_L, U_R, \cdot) f_{u_L}(U_L) f_{u_R}(U_R) dU_L dU_R.$$

We use explicit solutions of (5.5)–(5.8),

$$(5.9) \quad F(U; x, t) = \underbrace{F_0(U; x - \dot{q}(U)t)}_{F_{\text{ss}}(U; \mathbf{x}, t)} - \underbrace{\int_0^t \langle \mathcal{M}(U; x - \dot{q}(U)(t - t'), t') \rangle dt'}_{-F_{\text{kd}}(U; \mathbf{x}, t)},$$

to compute $F_{\text{obs}} = H[F]$. The measurement functional $H[\cdot]$ identifies data availability scenarios, $F_{\text{obs}}(U_i; \nu^{(x)} x_n, \nu^{(t)} t_m)$, for the space-time domain discretized with a mesh of size $\{\Delta x, \Delta U, \Delta t\}$. Here $U_i = i\Delta U$ with $i = 0, 1, \dots, N_U$, $x_n = n\Delta x$ with $n = 0, 1, \dots, N_X$, and $t_m = m\Delta t$ with $m = 0, 1, \dots, N_T$. The parameters $\nu^{(x)}, \nu^{(t)} \in \mathbb{Z}^+$ control the spatial and temporal sampling frequency, such that $\nu^{(x)} = \nu^{(t)} = 1$ represents “complete data” and increasing $\nu^{(x)}$ and/or $\nu^{(t)}$, correspond to the reduced frequency of sampling in space and in time, respectively.

For arbitrary initial conditions u_0 , a general form of \mathcal{M} is not available. Instead, in section 6, we use MCS to estimate F from solutions of (5.1) with multiple realizations of the random initial conditions (5.2). Alternatively, in traffic applications, macroscopic density estimates could come from camera recordings or satellite images (Eulerian observations) or from single-vehicle tracking (Lagrangian observations), which are mapped onto the macroscopic scale [14]. A data assimilation set-up acting at the level of CDFs, such as the one proposed in the present paper, would facilitate incorporating differential privacy constraints in case of sensitive data [20].

6. Simulation results. We start by investigating the performance and accuracy of the CDF method on the randomized RP; this setting is important in its own right since it poses challenges for many uncertainty quantification techniques and remains the subject of active research. Next, we learn the kinetic defect term through calibration of NNs via assimilation of sparse observations with the bottleneck initial condition and compare its performance with MCS.

6.1. Exact CDF solution of Riemann problem. Temporal snapshots (at $t = 1$) of the analytical solution $F(U; x, \cdot)$ of the CDF equation (5.5), with (5.8) for the RP (5.1) and (5.3), are shown in Figure 1 for the rarefaction (left) and shock (right) scenarios. The initial condition (5.3) is parametrized by two independent random variables u_L and u_R distributed uniformly on the intervals $[u_{L,\min}, u_{L,\max}]$ and $[u_{R,\min}, u_{R,\max}]$, respectively. The rarefaction and shock are achieved by setting $u_{L,\min} \geq u_{R,\max}$ and $u_{L,\max} \leq u_{R,\min}$, respectively. The solution $F(U; x, t)$ exhibits contact discontinuities but no shocks because the CDF equation (5.5) is linear.

Figure 2 elucidates further the behavior of the CDF, $F(U; \cdot)$, and PDF, $f(U; \cdot) = dF/dU$, of the solution $u(x, t)$ to the RP with shock. It shows these functions evaluated at $(x, t) = (-0.05, 1.0)$, together with the mean μ_u of u (dashed lines) and corresponding confidence interval $\mu_u \pm \sigma_u$ (grey region), both of which are computed analytically as

$$\mu_u(x, t) = \int_0^1 U f(U; x, t) dU, \quad \sigma_u^2(x, t) = \int_0^1 U^2 f(U; x, t) dU - \mu_u^2(x, t).$$

The shock configuration is characterized by bimodality in the regions of the domain close to the (random) position of the shock. This bimodal behavior underscores the limitations of the Kalman filter-based data assimilation, which is often used in traffic modeling to estimate the mean μ_u and variance σ_u^2 of the distribution $F(U; \cdot)$; these statistics are not particularly meaningful, given the bimodality of F .

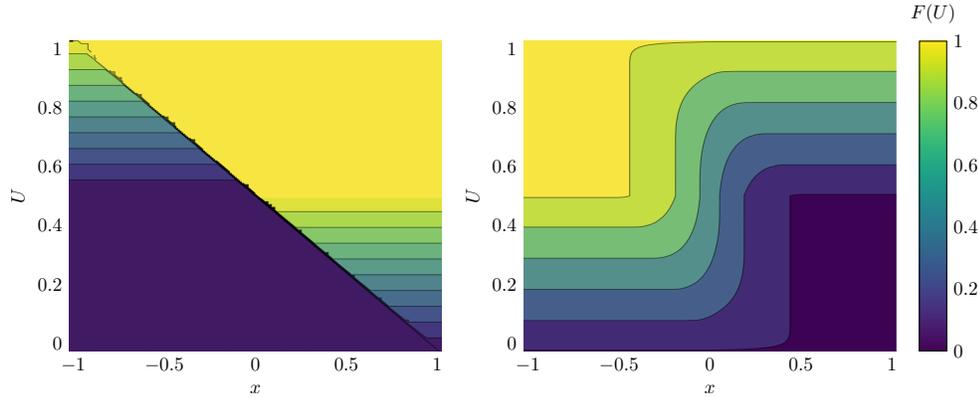


FIG. 1. CDF $F(U; x, t = 1.0)$ for the RP (5.1) and (5.3) given by the analytical solution of the CDF equation (5.5) with (5.8), for $x_d = 0$ and uniformly distributed independent random variables u_L and u_R . In the rarefaction case (left), u_L and u_R are defined on $[0.5, 1.0]$ and $[0.0, 0.5]$, respectively. In the shock case (right), u_L and u_R are defined on $[0.0, 0.5]$ and $[0.5, 1.0]$, respectively.

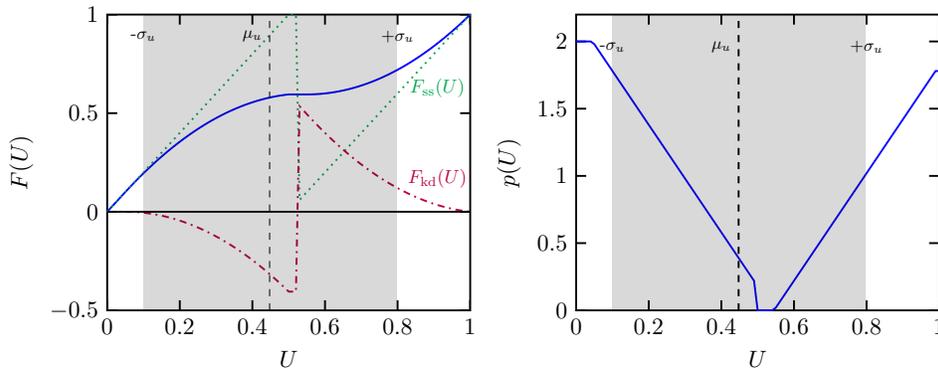


FIG. 2. CDF $F(U; x = -0.05, t = 1.0)$ (left) and PDF $f(U; x = -0.05, t = 1.0)$ (right) for the RP (5.1) and (5.3). These are computed by solving analytically the CDF equation (5.5) with (5.8), for $x_d = 0$ and independent random variables u_L and u_R distributed uniformly on $[0.0, 0.5]$ and $[0.5, 1.0]$, respectively. The exact CDF solution results from the summation of the smooth-solution component, F_{ss} , and the kinetic-defect component, F_{kd} , both shown in the left panel.

6.2. Data-driven inference of CDF equation. We use synthetic data to learn the kinetic defect (i.e., the model error) term $\langle \mathcal{M} \rangle$ in (5.5) and, hence, to render the approximate solution (2.6) fully computable and predictive. The statistical inference of $\langle \mathcal{M} \rangle$ is done, alternatively, with the Newtonian relaxation (section 3) and a NN (section 4) for the randomized RP (subsection 5.1), and with the NN for the bottleneck problem (5.1) and (5.2).

For both scenarios, we obtain observations up to $t = 1.44$ (this and other physical quantities are expressed in consistent units) via 10000 Monte Carlo realizations of (2.1) with random initial conditions on a spatial grid with $\Delta x = 0.02$. MCS are obtained by employing a shock-capturing Godunov scheme [29, Chap. 2], with time stepping guided by the Courant–Friedrichs–Lewy (CFL) condition. For $CFL = 0.8$ and the

maximum velocity throughout the domain, this yields $\Delta t = 0.16$. We use MCS, rather than the analytical solution of (2.5) available for one of the scenarios (RP), in order to have a consistent discretization error. The nudged CDF equation (3.1) is solved with the FiPy implementation [13] of the finite volume method. The computational domain $\mathcal{D} = \{(U, x) : 0 \leq U \leq 1, |x| \leq L; L = 1\}$ is discretized with a rectangular mesh of grid size $\Delta x = 0.02$ and $\Delta U = 0.015$ and $\Delta t = 0.016$, matching the discretization used for the NN strategy.

The NN minimization is performed via L-BFGS, a gradient-based minimization technique available in Tensorflow [1]. We approximate \hat{F}_{kd} by a NN with fixed architecture (9-layers, 20 neurons per hidden layer and the hyperbolic tangent activation function φ), initialized randomly but uniformly throughout the numerical experiments. Given the current lack of universal rules for choosing the NN architecture [24], empirical tuning of the NN hyperparameters is required. The chosen values appear to yield a sufficiently wide and deep NN capable of capturing the features of the function to be approximated.

Observations are assimilated up to $t = 0.96$ (interpolation or training window), whereas predictions are obtained up to $t = 1.44$ (extrapolation window). Spatial sampling frequency is indicated by the spatial lag $\nu^{(x)} \in \mathbb{Z}^+$ (i.e., every $\nu^{(x)}$ th spatial observation is assimilated), whereas $\nu^{(t)} \in \mathbb{Z}^+$ represent temporal sampling lag. A choice of $\nu^{(x)}$ and $\nu^{(t)}$ controls the size of the training set, with $\nu^{(x)} = \nu^{(t)} = 1$ corresponding to $[N_X = 101, N_T = 60]$ observation points. When these data are incorporated into the nudging equation (5.6), the nudging coefficient is assigned as $\lambda = \nu^{(t)}\lambda_0$, where $\lambda_0 = 100$; λ increases (as $\nu^{(t)}$ does) to compensate for the loss of accuracy due to the sparsity of temporal data.

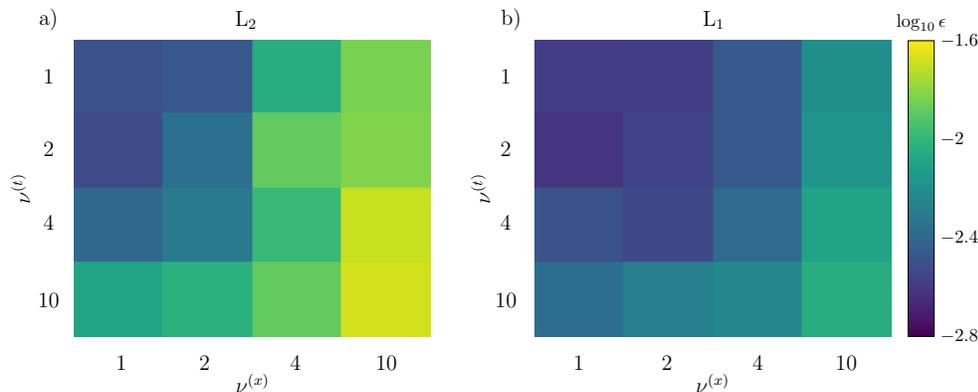


FIG. 3. Normalized overall errors $\epsilon = \|\hat{F} - F\|_{L_p(\mathbb{R}^{N_X \times N_U \times N_T})}$ as functions of spatial (ν^x) and temporal (ν^t) sampling frequencies, defined in terms of (a) L_2 and (b) L_1 norms (i.e., $p = 2$ and 1, respectively). The observation time horizon is set to $T = 0.96$, and the approximation \hat{F}_{kd} of F_{kd} is obtained with a NN. The colormap is in logarithmic scale.

6.2.1. Riemann problem with shocks. Figure 3 provides a comparison of the overall discrepancy between the exact CDF F and its NN-inferred approximation \hat{F} . The discrepancy is computed as an L_p ($p = 1$ and 2) norm in both space and time, with $\epsilon = \|\hat{F}(U; x, t) - F(U; x, t)\|_{L_p(\mathbb{R}^{N_X \times N_U \times N_T})}$ for CDF values in all locations $N_X \times N_U \times N_T$ up to the final interpolation time $T = 0.96$. The L_1 - and L_2 -norm

errors are qualitatively similar, an observation that holds for all of the experiments reported below. This is reassuring because we have convergence guarantees (in certain conditions) only for the L_1 norm. Figure 3 also reveals that, within the interpolation window, the NN approximation \hat{F} is more robust with respect to data sparsity in time (larger $\nu^{(t)}$) rather than in space (larger $\nu^{(x)}$). That is because more spatial data allow for a better reproduction of the spatial features, resulting in a smaller overall error.

Figure 4 exhibits the exact and NN-approximated CDFs at the end of the interpolation and extrapolation time intervals. The approximate CDF $\hat{F} = F_{ss} + \hat{F}_{kd}$ in (5.9) is obtained from the exact smooth-solution component F_{ss} and the NN-approximated component \hat{F}_{kd} . The two approximate solutions in the middle and right columns correspond to different sampling patterns: sparse-in-space/dense-in-time and sparse-in-time/dense-in-space, respectively. The latter sampling strategy is seen to lead to the more accurate estimators of the CDF $F(U; x, t)$ in interpolation mode, whereas the opposite is true in extrapolation mode (see below). As expected, the discrepancy between the exact and approximate solutions is higher in the extrapolation mode than in the interpolation one.

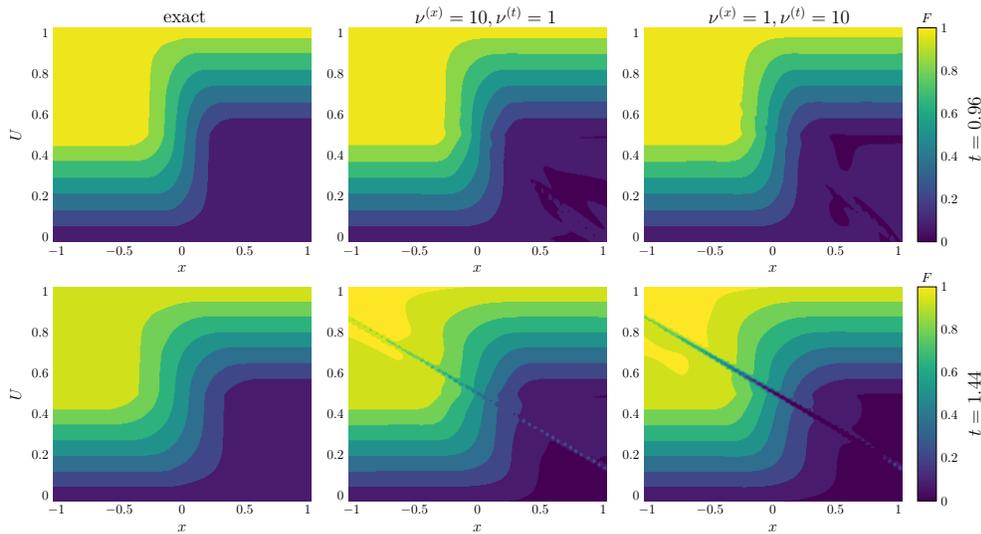


FIG. 4. RP with shock. The exact CDF $F(U; x, \cdot)$ (left column) and its NN approximations \hat{F} (middle and right columns) at the final interpolation time $t = 0.96$ (top row) and final extrapolation time $t = 1.44$ (bottom row). The middle and right columns present the NN estimators \hat{F} obtained with the sparse-in-space/dense-in-time and sparse-in-time/dense-in-space sampling strategies, respectively.

To quantify this approximation error, we plot in Figure 5 the absolute difference between the exact CDF and its approximations computed with the two sampling strategies, $|\hat{F}(U; x, \cdot) - F(U; x, \cdot)|$, at the end of the extrapolation time $t = 1.44$. The error is localized in the region with the largest gradients of F_{ss} and F_{kd} along the characteristic line that carries the initial discontinuity. Since F_{ss} is exact, the error is due to a mismatch between F_{kd} and \hat{F}_{kd} . Although not shown here, the error in the interpolation regime has the same spatial distribution but is one order of magnitude smaller. Extrapolation accuracy benefits from dense temporal observations within the interpolation window, such that dense-in-time/sparse-in-space observational maps

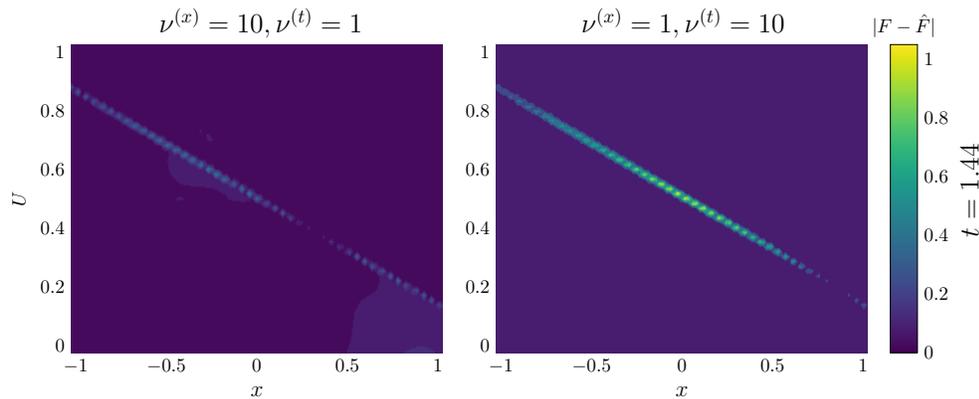


FIG. 5. *RP with shock.* The error $|\hat{F}(U; x, \cdot) - F(U; x, \cdot)|$ between the exact CDF F and its NN approximation \hat{F} obtained with the sparse-in-space/dense-in-time (left) and sparse-in-time/dense-in-space (right) sampling strategies at the final extrapolation time $t = 1.44$.

result in smaller errors than the opposite sampling strategy. This is to be expected, especially considering that with the current set up, $\nu^{(t)} = 10$ corresponds to only six temporal observations in all spatial locations as opposed to sixty when $\nu^{(t)} = 1$.

Figure 6 further elucidates the error's spatial structure by exhibiting the cross-section of the CDF maps in Figure 4 at $U = 0.75$. The exact solution $F(x, U^*, t = \{0.96, 1.44\})$ and its components F_{kld} and F_{ss} are represented with unmarked lines, while the marked lines refer to the NN-approximated \hat{F} . The errors emerge as spurious oscillations in the region where F_{kld} displays the largest gradients and grow with time within the extrapolation window. However, the predictive reconstruction holds far from the discontinuity.

Remark 6.1. One might apply the NN approximation directly to F rather than F_{kld} . Although not shown here, this strategy results in better overall accuracy at the end of the interpolation time in the event of spatially dense observations by virtue of the smoothness of the CDF F . However, this gain quickly disappears as observation sparsity increases, yielding distorted estimates of the CDF. Taking advantage of the dynamics-dependent deterministic component, F_{ss} , as we propose in this study, improves the overall behavior of the final solution. It also delineates the region of the spatio-temporal domain wherein the NN-approximated correction \hat{F}_{kld} is present, i.e., the region over which the solution is nonsmooth.

We quantify the degradation in the predictive power of our “learned” model (2.6) in terms of the relative L_2 norm of the discrepancy between the exact CDF F and its NN approximation \hat{F} , i.e., in terms of the time-dependent error,

$$\mathcal{E}(t) = \frac{\|\hat{F}(t) - F(t)\|_{L_2(\mathbb{R}^{N_x \times N_U})}}{\|F(t)\|_{L_2(\mathbb{R}^{N_x \times N_U})}}.$$

This is plotted in Figure 7, with the training and prediction windows represented by the white and grey backgrounds, respectively. As expected, the error is low (compatible with the convergence conditions for minimization) within the interpolation window, especially for small $\nu^{(x)}$, and grows considerably for $t > 0.96$, reaching its maximum values of around 8% at $t = 1.44$. The interpolation error is larger at early

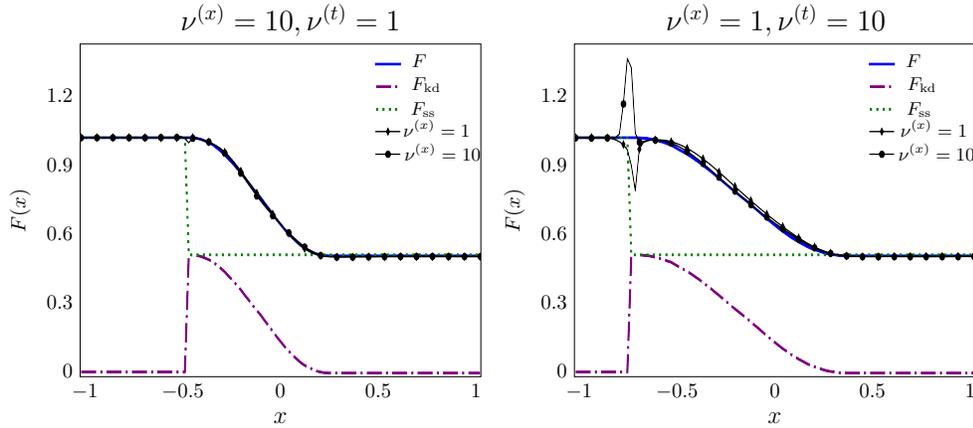


FIG. 6. RP with shock. Spatial profiles $F(\cdot; x, \cdot)$ of the exact CDF $F(U; x, t)$ and its components F_{ss} and F_{kd} , and of its approximation $\hat{F} = F_{ss} + \hat{F}_{kd}$. The profiles are reported for $U = 0.75$ at the final interpolation time $t = 0.96$ (left) and final extrapolation time $t = 1.44$ (right). The lines with/without symbols represent the approximate/exact results.

times, when F_{kd} displays sharp gradients and observations are spatially sparse. The error growth during extrapolation is stronger for scarce spatial and—most of all—temporal observations. Values of these errors are expected to change slightly, while maintaining the same qualitative behavior, for different configurations or initializations of the NN.

The predictive power of the NN approximation of \hat{F}_{kd} is highlighted in Figure 8, which provides a comparison between the exact contribution of the model error, F_{kd} , and its NN reconstruction, \hat{F}_{kd} . The analytical F_{kd} component is represented in Figure 8 (left column) at $t = 0.96$ and $t = 1.44$. As expected, the reconstruction is accurate at $t = 0.96$ and more distorted at the final extrapolation time $t = 1.44$. Even outside its training region, the NN captures the rotation of the discontinuity from $t = 0.96$ to $t = 1.44$ due to a different spatial reach of the characteristic line that propagates the initial discontinuity, as well as the overall diffusion. Across this discontinuity, F_{kd} changes sign to compensate for the discrepancy between the exact F and its smooth-solution component F_{ss} . The kinetic defect fills in the solution on both sides of the discontinuity.

Next, we compare the approximations of the model error F_{kd} estimated by a NN and Newtonian relaxation (NR) from the same observations. Figures 9 and 10 depict the nudging counterparts of Figures 4 and 7 for $\nu^{(x)} = 1$ and variable $\nu^{(t)}$. The NR-approximated CDF solution is relatively accurate at the end of the interpolation window, although it requires the data to be space-time dense for it to be of the same order of accuracy as the NN reconstruction ($\nu^{(x)} = \nu^{(t)} = 1$). The NR approximation is strongly affected by data sparsity even within the interpolation window ($\nu^{(t)} = 10$), and its accuracy deteriorates even further during extrapolation (Figure 9). That is because, in the absence of observations, the nudged CDF dynamics (3.1) coincides with the smooth-solution dynamics, resulting in distorted CDFs. In particular, the approximate CDFs completely lose salient features of CDFs, such as monotonicity in the U direction.

The corresponding normalized error is represented in Figure 10 for both interpola-

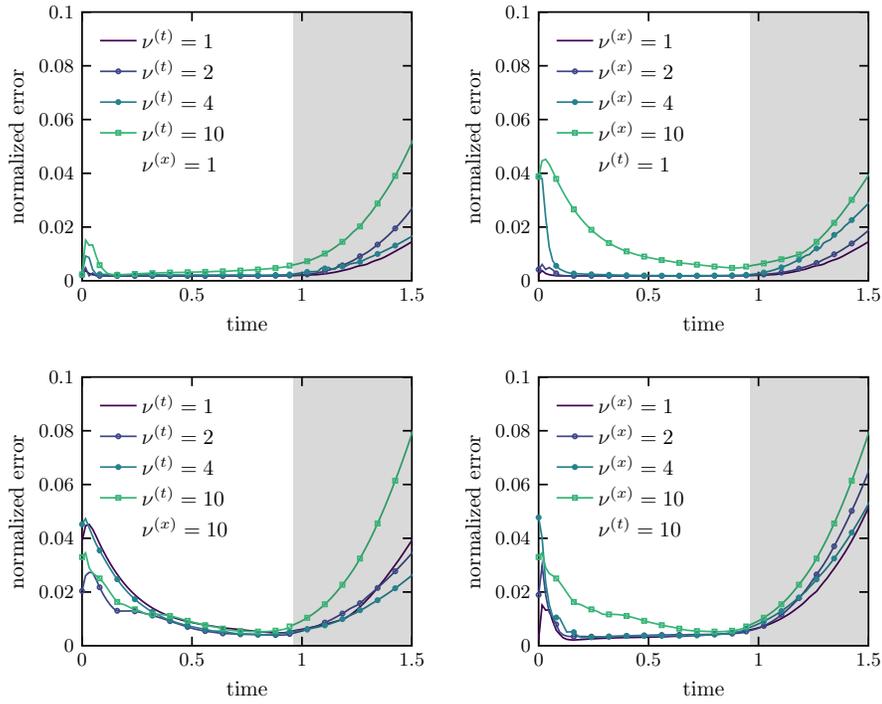


FIG. 7. RP with shock. Normalized error $\mathcal{E}(t)$ for different observational strategies. Sampling lag in space and in time is represented by values of $\nu(x)$ and $\nu(t)$, respectively.

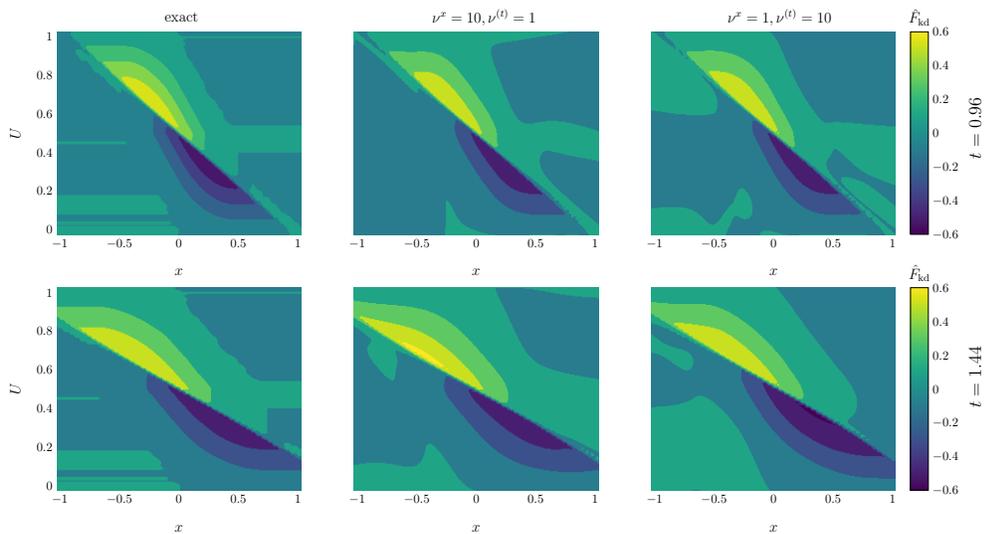


FIG. 8. RP with shock. Maps of the kinetic defect component of the CDF solution F_{kd} (left column) and its NN approximation \hat{F}_{kd} (middle and right columns) at the final interpolation time $t = 0.96$ (top row) and at the final extrapolation time $t = 2$ (bottom row). The middle and right columns present results for the sparse in space/dense in time and sparse in time/dense in space sampling patterns, respectively.

tion and extrapolation windows. Within the interpolation range, the error is bounded (as predicted by the theoretical convergence proofs in the L_1 norm), although not monotonically decreasing. The error increases as observations cease. Throughout the whole range, the NR errors are comparatively larger than those of a NN (the errors displayed in Figure 10 are to be compared with the top right panel in Figure 7). Moreover, the NR approximation introduces a saw-like behavior within the interpolation window. This is due to the fact that between observations, the approximated CDF obeys a smooth-solution dynamics with no kinetic correction. This behavior could be smoothed by introducing space/time variability of the nudging coefficients [31]. These findings suggest the superiority of the NN approximation of F_{kd} . Hence, we do not report the NR results for the next test case.

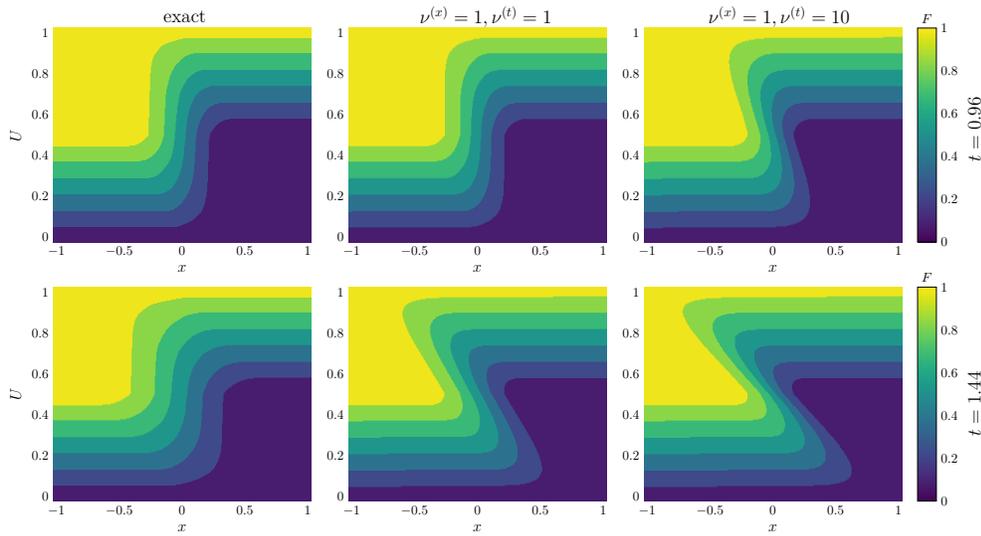


FIG. 9. *RP with shock.* The exact CDF $F(U; x, \cdot)$ (left column) and its Newtonian relation approximations \hat{F} (middle and right columns) at the final interpolation time $t = 0.96$ (top row) and final extrapolation time $t = 1.44$ (bottom row). The middle and right columns present the estimators \hat{F} obtained with the dense-in-space/dense-in-time and sparse-in-time/dense-in-space sampling strategies, respectively.

6.2.2. Bottleneck problem. The results reported in this section correspond to the initial condition (5.2) with $x_L = -0.1$ and $x_R = 0.1$, and the random constants l , m , and r distributed uniformly on $[0.4, 0.6]$, $[0.8, 1.0]$, and $[0.0, 0.6]$, respectively. A major goal of this test is to demonstrate our method’s ability to handle random initial data for which an explicit expression for the model error term $\langle \mathcal{M} \rangle$ is not available.

Figures 11–15 are direct analogs of Figures 4–8 for the bottleneck problem. The method’s performance on the two problems is qualitatively similar, suggesting a certain degree of robustness of the proposed approach. The largest discrepancy between the exact CDF F and the approximated CDF \hat{F} is localized along the characteristic lines that originate at the physical boundaries of the bottleneck, which correspond to the areas where the kinetic defect introduces the most significant correction (Figures 11 and 12).

The exact and NN-approximated CDF profiles (at $U = 0.75$) in Figure 13 provide

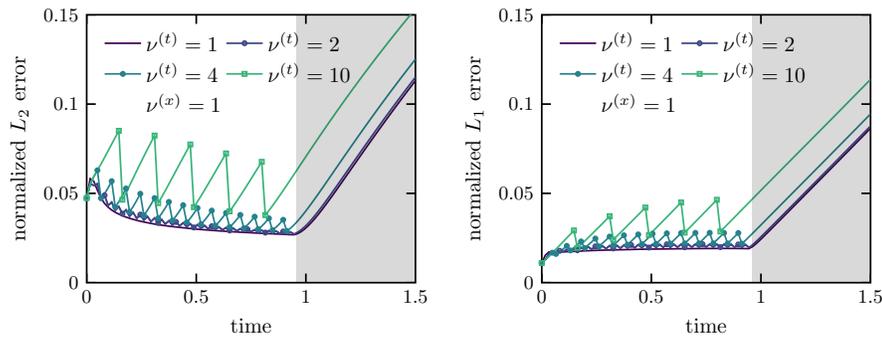


FIG. 10. *RP with shock. Normalized L_2 and L_1 error $\mathcal{E}(t)$ (left and right panel, respectively) for different observational strategies and the nudging assimilation procedure. Sampling lag in time is represented by values of $\nu^{(t)}$.*

a more quantitative understanding of the challenges of approximating F_{kd} with a NN given its rather irregular behavior. The discrepancy is large at the discontinuity, but the CDF solution is well captured in all other parts of the domain where the correction introduced by F_{kd} is large but constant. Large errors are thus mostly due to the spurious oscillations that arise at the discontinuity, but do not affect the accuracy of the solution far from it.

The bottleneck scenario is more sensitive to the spatial sparsity of the data, even within the interpolation window (Figure 14), at least in extreme conditions (large $\nu^{(x)}$). This is due to the fact that the bottleneck scenario is characterized by localized spatial variability of the CDF, which requires denser data to be properly characterized and reconstructed.

Figure 15 zeros in on a NN's ability to learn the model error F_{kd} . The benchmark solution for F_{kd} (Figure 15, left column) is obtained by subtracting the exact smooth-solution component F_{ss} from the complete set of observations of F . The benchmark and NN-approximated predictions of F_{kd} are comparable. A better agreement between the two could be achieved by enforcing smoothness of the approximated solution, in order to smooth out numerical inaccuracies along the discontinuity lines where the two contributions (F_{ss} and F_{kd}) are supposed to compensate each other.

7. Conclusions. Uncertainty quantification and data assimilation techniques for hyperbolic balance laws must be able to deal with the development of shocks. Classic UQ and filtering techniques are ill-suited for such problems because their solutions can be highly non-Gaussian. For example, PDFs of traffic density often develop bimodalities [5], which render estimates of its mean and covariance not informative, and assimilation of data provided by traffic cameras and GPS trackers challenging.

To overcome this limitation, we developed an evolution equation for the CDF of solutions to hyperbolic balance laws with random initial data. It is fully computable and, hence, usable as a predictive tool for linear balance laws, and nonlinear balance laws with smooth solutions and/or random step initial conditions. Other scenarios are accommodated by adding to the CDF equation a model error (kinetic defect) term, which has to be inferred from observations.

We explored two alternative strategies to accomplish this task: Newtonian relaxation (NR, or nudging) and Deep Neural Networks (NNs). The former strategy

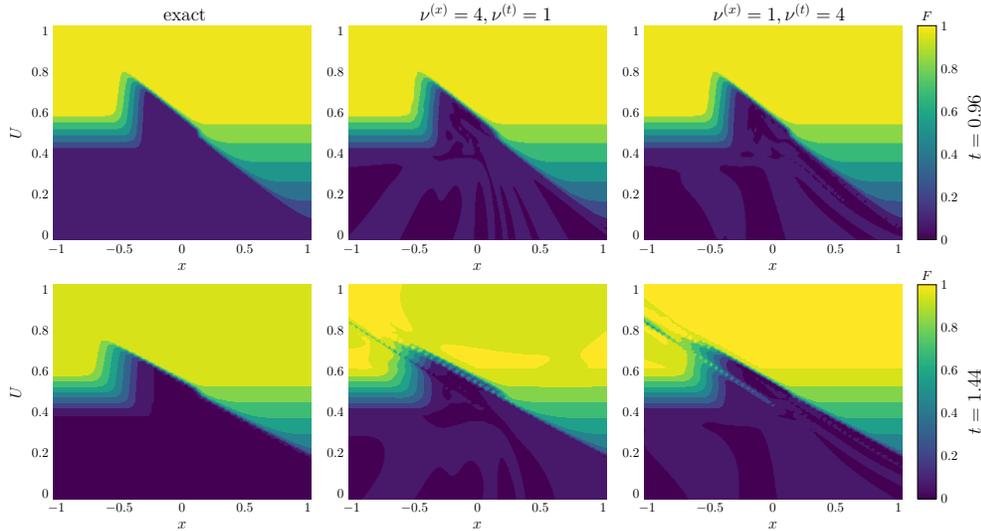


FIG. 11. Bottleneck problem. The exact CDF $F(U; x, \cdot)$ (left column) and its NN approximations \hat{F} (middle and right columns) at the final interpolation time $t = 0.96$ (top row) and final extrapolation time $t = 1.44$ (bottom row). The middle and right columns present the estimators \hat{F} obtained with the sparse-in-space/dense-in-time and sparse-in-time/dense-in-space sampling strategies, respectively.

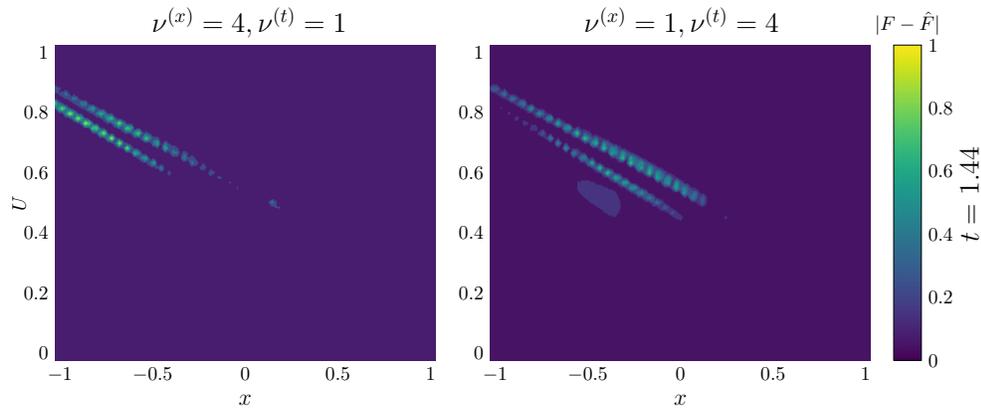


FIG. 12. Bottleneck problem. The error $|\hat{F}(U; x, \cdot) - F(U; x, \cdot)|$ between the exact CDF F and its NN approximation \hat{F} obtained with the sparse-in-space/dense-in-time (left) and sparse-in-time/dense-in-space (right) sampling strategies at the final extrapolation time $t = 1.44$.

lends itself to theoretical analysis, while the latter remains largely empirical. The performance of these two methods was investigated numerically on two problems that admit shocks. Our analysis leads to the following major conclusions.

1. For smooth solutions and for the class of nonsmooth solutions described above, our CDF method is exact and leads to significant computational savings over Monte Carlo simulations.

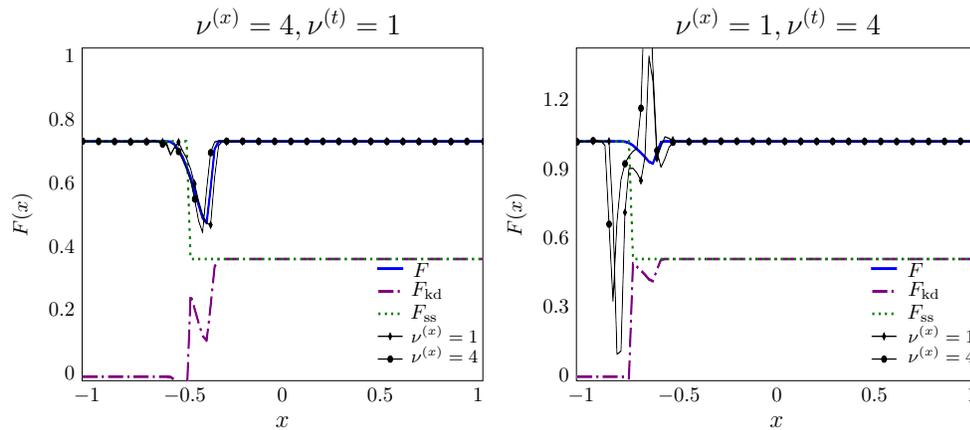


FIG. 13. *Bottleneck problem. Spatial profiles $F(\cdot; x, \cdot)$ of the exact CDF $F(U; x, t)$ and its components F_{ss} and F_{kd} , and its approximation \hat{F} . The profiles are reported for $U = 0.75$ at the final interpolation time $t = 0.96$ (left) and final extrapolation time $t = 1.44$ (right). The lines with/without symbols represent the approximate/exact results.*

2. For general hyperbolic laws with shocks, the CDF equation contains the model error (kinetic defect) term that is unique but unknown.
3. We proved theoretical convergence of the NR-approximated CDF equation towards its exact counterpart as data are dynamically assimilated into the CDF equation.
4. In comparison to the NR approximation, the NN-inferred CDF equation is more robust towards both data scarcity and future predictions.
5. Spatial data density is crucial for good interpolation accuracy, whereas long temporal sequences of observations improve future projections.

Appendix A. Kinetic formulation of hyperbolic conservation laws. For smooth solutions $u(\mathbf{x}, t)$ of (2.1), multiplying (2.1) by $-\partial\Pi/\partial U$ and accounting for

$$\frac{\partial\Pi}{\partial t} = -\frac{\partial\Pi}{\partial U} \frac{\partial u}{\partial t} \quad \text{and} \quad \nabla\Pi = -\frac{\partial\Pi}{\partial U} \nabla u$$

yields an exact equation (2.2) for Π . When discontinuities develop, (2.1) is solved in the sense of distributions. The entropic constraint

$$(A.1) \quad \frac{\partial S}{\partial t} + \nabla \cdot \boldsymbol{\eta} \leq 0, \quad \boldsymbol{\eta}(u) = \int_0^u \frac{\partial S(u')}{\partial u'} \frac{\partial \mathbf{q}(u')}{\partial u'} du'$$

is introduced to guarantee the uniqueness of $u(\mathbf{x}, t)$ as a solution of the weak counterpart of (2.1). Equation (A.1) is fulfilled for any entropy pair $(S, \boldsymbol{\eta})$ of a convex function $S(\cdot)$ and the corresponding flux $\boldsymbol{\eta}(u)$.

The pair (2.1) and (A.1) is equivalent to a kinetic formulation [22],

$$(A.2) \quad \frac{\partial\Pi(U, \mathbf{x}, t)}{\partial t} + \dot{\mathbf{q}}(U; \mathbf{x}, t) \cdot \nabla\Pi(U, \mathbf{x}, t) = \frac{\partial m(U, \mathbf{x}, t)}{\partial U},$$

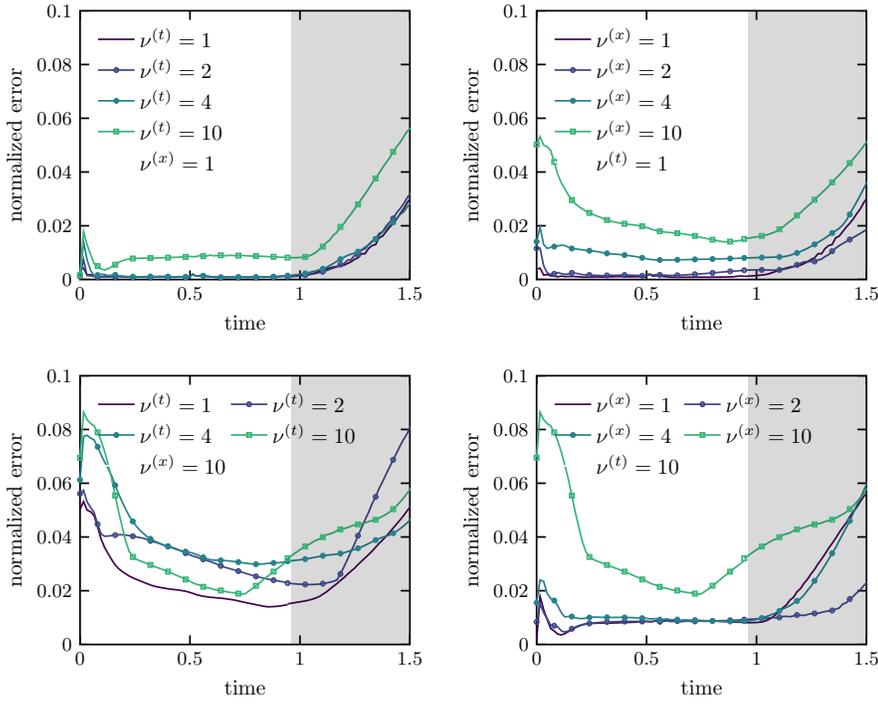


FIG. 14. Bottleneck problem. Normalized error $\mathcal{E}(t)$ for different observational maps. Sampling lag in space and in time is represented by values of $\nu(x)$ and $\nu(t)$, respectively.

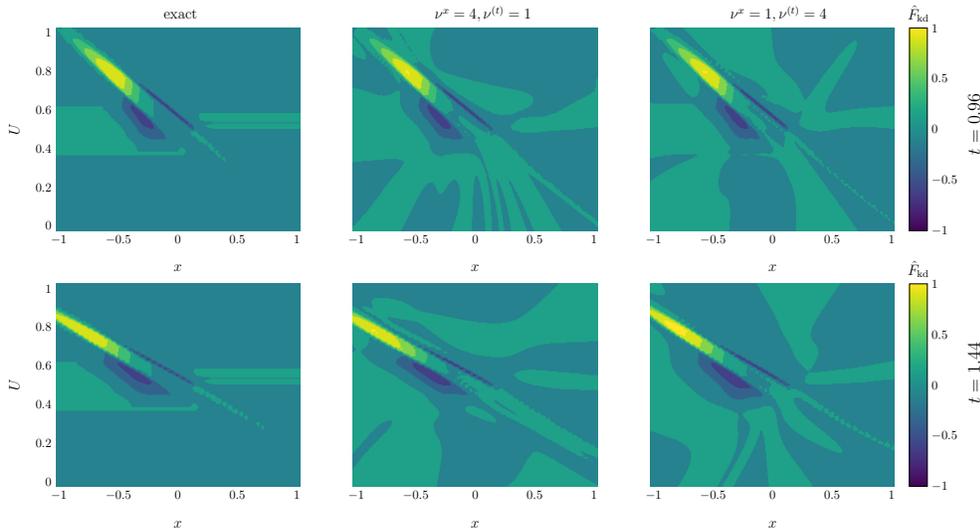


FIG. 15. Bottleneck problem. Maps of the kinetic defect component of the CDF solution F_{kd} (left column) and its approximation \hat{F}_{kd} (middle and right columns) at the final interpolation time $t = 0.96$ (top row) and at the final extrapolation time $t = 1.44$ (bottom row). The middle and right columns present results for the sparse in space/dense in time and sparse in time/dense in space sampling patterns, respectively.

whose solution $\Pi(U, \mathbf{x}, t) = \chi_{u(\mathbf{x}, t)}(U)$ is defined by

$$\chi_u(U) = \begin{cases} 1 & \text{if } 0 \leq U \leq u, \\ -1 & \text{if } u < U < 0, \\ 0 & \text{otherwise,} \end{cases}$$

subject to the initial condition $\Pi(U, \mathbf{x}, t = 0) = \chi_{u_0(\mathbf{x})}(U)$. The solution of the original problem, $u(\mathbf{x}, t)$, is recovered by integration,

$$u(\mathbf{x}, t) = \int_{\Omega_U} \Pi(U, \mathbf{x}, t) dU.$$

Equation (A.2) is linear and defined on an augmented space $\tilde{\Omega} = \mathbb{R}^d \cup \Omega_U$, where the additional direction $U \in \Omega_U$ spans the range of variation of the solution $u(\mathbf{x}, t)$. The coordinate U acts in (A.2) as a parameter. For consistency with the physical systems under consideration, we limit the discussion to the case of $u(\mathbf{x}, t) > 0$ [10]; a full treatment can be found in [22, 23].

The kinetic formulation has the advantage of being linear, but it is a priori undefined because of the introduction of $m(\mathbf{x}, t)$. The latter is generally unknown but is proved to be unique and equal to zero whenever the solution $u(\mathbf{x}, t)$ is smooth. This accounts for the possible lack of regularity of $u(\mathbf{x}, t)$. Its role is to constrain all possible solutions to those of the form of χ . A functional form of m can be derived analytically under certain conditions, e.g., when u_0 is a step initial condition and the concavity of \mathbf{q} has a definite sign.

Because (A.2) is linear, its solution can be written as

$$(A.3) \quad \Pi(U, \mathbf{x}, t) = \int_{\Omega} \Pi_0(U, \boldsymbol{\xi}) \mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t) d\boldsymbol{\xi} + \int_0^t \int_{\Omega} \mathcal{M}(U, \boldsymbol{\xi}, \tau) \mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t - \tau) d\boldsymbol{\xi} d\tau,$$

where $\mathcal{M}(U, \mathbf{x}, t) = \partial m(U, \mathbf{x}, t) / \partial U$, and $\mathcal{G}(\mathbf{x}, \boldsymbol{\xi}, t - \tau)$ is the Green's function for a homogeneous problem

$$\frac{\partial \mathcal{G}}{\partial \tau} = -\dot{\mathbf{q}}(U) \cdot \nabla_{\boldsymbol{\xi}} \mathcal{G} - \delta(\mathbf{x} - \boldsymbol{\xi}) \delta(t - \tau).$$

In this work we are not interested in recovering the solution of the deterministic conservation law, but in understanding the behavior of the CDF of its stochastic counterpart, which is defined as $F(U; \mathbf{x}, t) = 1 - \langle \Pi(U; \mathbf{x}, t) \rangle$, where

$$(A.4) \quad \langle \Pi(U; \mathbf{x}, t) \rangle = \int_{\Omega_U} \Pi(U; \mathcal{U}) f(\mathcal{U}; \mathbf{x}, t) d\mathcal{U},$$

and $f = dF/dU$ is the PDF of the solution $u(\mathbf{x}, t)$ at space-time point (\mathbf{x}, t) .

Appendix B. Convergence of CDF equation to observations. Recall that an approximate CDF \hat{F} is a solution of the nudged dynamics equation (5.6), and F is the exact CDF defined as a solution of (5.5). Let $F_{\text{obs}} = H(F)$ represent a finite set of observations of the exact CDF F through the observation map $H(\cdot)$. Then convergence of \hat{F} to F via assimilation of observations F_{obs} is proven by finding a bound to the time evolution of the L_1 error norm $\|\hat{F} - F\|_{L_1(\tilde{\Omega})}$ computed in the Lebesgue space $L_1(\tilde{\Omega})$. By definition, $\hat{F} = 1 - \langle \hat{\Pi} \rangle$ and $F = 1 - \langle \Pi \rangle$, where $\hat{\Pi}$ and Π obey

$$(B.1) \quad \frac{\partial \hat{\Pi}}{\partial t} + \dot{q}(U) \frac{\partial \hat{\Pi}}{\partial x} = \lambda(\Pi_{\text{obs}} - \hat{\Pi})$$

and (A.2), respectively. The dynamically assimilated data $\Pi_{\text{obs}} = H(\Pi)$ represent the observations of Π .

By virtue of the triangle inequality, considering the ensemble averaging $\langle \cdot \rangle$ as a summation of infinite terms,

$$(B.2) \quad \|\hat{F} - F\|_{L_1(\tilde{\Omega})} \leq \langle \|\hat{\Pi} - \Pi\|_{L_1(\tilde{\Omega})} \rangle.$$

Hence, the convergence of \hat{F} to F relies on the convergence of $\hat{\Pi}$ to Π , making (B.2) a corollary of the theorems for the convergence of $\hat{\Pi}$.

Convergence bounds for $\|\Pi - \Pi\|_{L_1(\tilde{\Omega})}$ have been proved in [23, 7] for certain combinations of solution types (i.e., for smooth and nonsmooth u) and observation maps $H(\cdot)$. We summarize their findings below, along with the corresponding bound for $\|\hat{F} - F\|$.

B.1. Smooth u and complete observations. For smooth solutions (i.e., $\mathcal{M} = 0$) and complete observations (one-to-one observation map, i.e., $\nu^{(x)} = \nu^{(t)} = 1$), Boulanger et al. [7, Prop. 2.3, p. 593] prove that

$$(B.3) \quad \lim_{t \rightarrow 0} \|\hat{\Pi} - \Pi\|_{L_1(\tilde{\Omega})} = 0.$$

The convergence rate is exponential and the nudging coefficient λ governs the rate of exponential decline of the error. The convergence rate holds also, in the ensemble sense, for \hat{F} .

B.2. Nonsmooth u and complete observations. For nonsmooth solutions (i.e., $\mathcal{M} \neq 0$), complete observations and bounded initial conditions, Boulanger et al. [7, Thm. 2.4, p. 594] proved that, for a given time $T > 0$,

$$(B.4) \quad \|\hat{\Pi}(T) - \Pi(T)\|_{L_1(\tilde{\Omega})} \leq C_\lambda e^{-\lambda T} + C(d, U, \|u_0\|_{BV}) \sqrt{\frac{T}{\lambda}},$$

where C_λ is a coefficient dependent on the initial condition, and $C(\cdot)$ is a time-independent constant. Convergence is accelerated by large λ because of the exponential decay of the initial error $C_0 = \|\Pi(t = 0) - \Pi_\lambda(t = 0)\|_{L_1(\tilde{\Omega})}$, and the $\sqrt{T/\lambda}$ decay of the problem-dependent component of the error bound. Moreover, the bound is timeframe dependent. Since the discrepancy tends to increase with time, λ needs to be chosen large enough to compensate for the growth of the error bound with \sqrt{T} . The corresponding bound for the CDF convergence is obtained by ensemble averaging the right-hand side of (B.4).

B.3. Nonsmooth u and discrete temporal observations. For complete spatial observations that are assimilated at discrete times t_k ($k \in \mathbb{I}$) and interpolated over finite time windows with size t_w , Boulanger et al. [7, Thm. 2.14, p. 602] propose smoothed nudging in the form

$$\lambda \left\{ \sum_{k \in \mathbb{I}} \varphi_{t_w}(t - t_k) [\Pi_{\text{obs}}(t_k, x, U) - \Pi(t_k, x, U)] \right\}.$$

This results in a bounded convergence for Π ,

$$\|\hat{\Pi}(T) - \Pi(T)\|_{L_1(\tilde{\Omega})} \leq C_0 e^{-\lambda K} + t_w \mathcal{I}(T) + \frac{1}{\lambda} \sup_{0 < t \leq T} \|\mathcal{M}(t)\|_{L_1(\tilde{\Omega})},$$

where K is the number of time steps in $[0, T - t_w]$, C_0 is an initial condition dependent constant, and $\mathcal{I}(T)$ is a convergence-rate dependent quantity (for details, see [7]). The error bound depends on (i) the initial error, whose effect decreases with time and increasing λ , (ii) a nonnegligible, t_w -dependent component (that does not increase with λ), and (iii) a kinetic-defect related term that decreases with λ . The ensemble average of this term provides a bound for the convergence of the CDF, with the same λ -dependent dynamics.

B.4. Smooth u and limited spatial observations. Boulanger et al. [7] provide proof of convergence for simplified one-dimensional problem periodic boundary conditions, whose smooth solution is observed on a limited portion of the domain $[s_l, s_u] \in [0, 1]$. The problem of data-scarcity in space is particularly difficult for hyperbolic conservation laws that propagate a wave with finite speed. The periodic set-up is particularly suitable for traffic applications, since it mimics the conditions of a ring road. Under these conditions [7, Prop. 2.18, p. 604],

$$\|\Pi(t) - \Pi_{\text{obs}}(t)\|_{L_2} \leq \|\Pi(t=0) - \Pi_{\text{obs}}(t=0)\|_{L_2} \exp(-\lambda X_t^{\text{inf}}).$$

This exponential convergence is ensured for $\lambda \geq t/X_t^{\text{inf}}$, with $X_t^{\text{inf}} = \min_x X_t > 0$, where

$$X_t(x) = \int_0^t \mathcal{H}(x + \dot{q}(U)t' - s_l) \mathcal{H}(s_u - x - \dot{q}(U)t') dt'$$

is the travel time that is necessary for a particle within the observation window that travels with the characteristic speed to reach all estimate locations.

Appendix C. Nudging: Density-enhanced data assimilation. For the traffic problem considered in subsection 5.1, we verify the convergence behavior of the nudged dynamics (5.5) for different temporal sampling patterns. Spatial sparsity is not considered because it distorts the behavior of the nudged CDF. The data, in the form of CDF measurements $F_{\text{obs}}(U; x, t)$, are generated by solving both the CDF equation (5.5) and (5.8) with the FiPy implementation [13] of the finite-volumes method. (We use the numerical solution, rather than its analytical counterpart used in subsection 6.1, in order to have a consistent discretization error between this solution and the solution of the nudging equation (5.6).) The numerical discretization and configuration are described in subsection 6.2. As before, we consider several data availability scenarios by varying $\nu^{(t)}$.

We also account for uncertainty in the initial data by letting the random constants u_L and u_R in the initial condition (5.6) be characterized by a given distribution \hat{F}_0 . In one case, we let u_L and u_R be uniformly distributed on $[u_{L,\text{min}} = 0.0, u_{L,\text{max}} = 0.5]$ and $[u_{R,\text{min}} = 0.5, u_{R,\text{max}} = 1.0]$, respectively; in the other case, we assign truncated Gaussian distributions on the same intervals. (Recall that the data $F_{\text{obs}}(U_i; x_n, \nu t_m)$ were generated for the uniformly distributed random constants u_L and u_R .)

We show in Appendix B.2 that in the case of discontinuous solutions and when data are complete (i.e., $\nu^{(x)} = \nu^{(t)} = 1$), the approximate solution \hat{F} of (5.6) with large λ converges exponentially to the exact solution F of (5.5) as time $t \rightarrow \infty$, reducing the initial discrepancy whenever present. For incomplete data sets, a bound on the L_1 -norm error is obtained depending on the sparsity of observations (Appendix B.3). Figure 16 confirms these theoretical results by showing the normalized L_1 norm of the error as a function of time t . To focus on the method's convergence rate, we do not introduce any interpolation between observations at subsequent observation times, as is often done in practice [25]. This would further improve the performance of data assimilation and reduce the error.

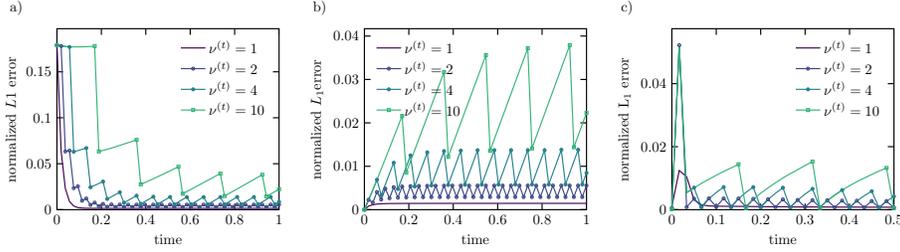


FIG. 16. *RP with shock (panels (a) and (b)). Normalized L_1 norm, $\mathcal{E}(t) = \|F(\cdot, t) - \hat{F}(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})} / \|F(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})}$, of the difference between the exact CDF F computed with (5.5), (5.8) and the data-informed CDF $\hat{F}(U)$ computed with (5.6) and data F_{obs} with $\nu^{(x)} = 1$ and varying $\nu^{(t)}$, and with inaccurate (a) and exact (b) initial conditions. Slowdown scenario (panel c): Normalized L_1 norm, $\mathcal{E}(t) = \|F(\cdot, t) - \hat{F}(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})} / \|F(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})}$, of the difference between the exact CDF F computed via MC simulations and the nudged $\hat{F}(U)$ computed with (5.6), and observations F_{obs} with $\nu^{(x)} = 1$ and varying $\nu^{(t)}$.*

In the case of inaccurate initial conditions, i.e., $\hat{F}_0 \neq F_0$, Figure 16a presents the normalized error in the form $\mathcal{E}(t) = \|F(\cdot, t) - \hat{F}(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})} / \|F(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})}$. The complete data case ($\nu^{(x)} = \nu^{(t)} = 1$) displays the theoretical exponential decrease of the error from its initial nonzero value. When data are incomplete ($\nu^{(t)} > 1$), the error exhibits an exponentially decreasing trend albeit with a caveat: It increases locally in the time intervals when the data are not available, and then drops at the assimilation step. In the case of exact initial conditions, i.e., $\hat{F}_0 \equiv F_0$, the normalized error $\|F(\cdot, t) - \hat{F}(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})} / \|F(\cdot, t)\|_{L_1(\mathbb{R}^{N_X \times N_U})}$ increases initially and later reaches values similar to the asymptotic ones displayed in Figure 16b. In both cases, the overall discrepancy, while bounded, increases slightly over time with sparse time measurements and needs to be compensated by the choice of larger λ .

For general initial conditions, the true solution of (5.5) is not available. We extract F_{obs} from the generation of a training set of $N_r = 200$ MCS of (5.1); the numerical solver employs a Godunov scheme [11, 28] with the following integrable and bounded initial condition: $u_0(x) = A \exp(-100x^2)$, where the random variable A is distributed normally with mean 0.5 and standard deviation 0.1. The size and resolution of the spatial domain is the same as in the previous case, with $\Delta t = 0.016$ as a result of CFL = 0.8. The set of MCS yields an empirical CDF $F_{MC}(v, x, t)$, which is to be dynamically assimilated by posing $F_{obs} = H[F_{MC}]$ with a sampling frequency controlled by $\nu^{(x)}$ (here constant and equal to 1) and $\nu^{(t)}$. The initial condition is assumed to be known, $\hat{F}_0 = F_0$. The normalized L_1 norm of the error $\mathcal{E}(t)$ (Figure 16c) confirms convergence of the assimilation procedure in time for different sampling frequencies $\nu^{(t)}$. The error decreases in time as more data are assimilated; the error increases locally between measurement times but remains bounded, as demonstrated in Appendix B.

Nudging can be used in its BFN variant to improve the accuracy in the assignment of the initial condition \hat{F}_0 in case that step is affected by error. This is further explored in Appendix C.1. Once a reliable estimate of \hat{F}_0 is obtained (either directly or via BFN), (2.5) with $\langle \mathcal{M} \rangle = 0$ can be used in forward mode to predict the smooth solution CDF, F_{ss} . Discrepancy between F_{ss} and observations signals the development of shocks in the physical solution of the problem, and allows us to map $|H(F_{kd})| = F_{obs} - H(F_{ss})$ in the observation points.

C.1. BFN for identification of initial conditions. To refine an estimate of the initial state of the system, we use the BFN [3, 4] version of (3.1), where we iteratively solve the pair

$$(C.1a) \quad \frac{\partial \hat{F}_k}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \hat{F}_k = \lambda(F_{\text{obs}} - \hat{F}_k), \quad \hat{F}_k(t=0) = \hat{F}_{b,k-1}(t=0),$$

$$(C.1b) \quad \frac{\partial \hat{F}_{b,k}}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \hat{F}_{b,k} = -\lambda(F_{\text{obs}} - \hat{F}_{b,k}), \quad F_{b,k}(t=T) = \hat{F}_k(t=T).$$

Once the problem is initialized by assigning $\hat{F}_{b,0}(t=0)$, the forward iterative k th step (C.1a) yields approximate initial conditions $\hat{F}_k(t=T)$ for the backward problem (C.1b), which is to be solved backwards in time (i.e., from T to 0). Iterations are repeated until convergence of the forward and backward predictions of \hat{F} .

BFN yields a posterior estimator of $\hat{F}_0(U, \mathbf{x})$, which is then used as the initial condition for the CDF equation (2.5) with $\mathcal{M} \equiv 0$ to predict the behavior of the CDF $F_{\text{ss}}(U; \mathbf{x}, t)$ of a smooth solution $u(\mathbf{x}, t)$ at future times. The discrepancy between $F_{\text{ss}}(U; \mathbf{x}, t)$ and the CDF of subsequent observations, $|F_{\text{ss}}(U; \mathbf{x}, t) - F_{\text{obs}}(U; \mathbf{x}, t)|$, identifies the location and development of shocks and can yield a map of F_{kd} . This discrepancy can then be used to train a NN.

REFERENCES

- [1] M. ABADI, A. AGARWAL, P. BARHAM, E. BREVDO, Z. CHEN, C. CITRO, G. S. CORRADO, A. DAVIS, J. DEAN, M. DEVIN, S. GHEMAWAT, I. GOODFELLOW, A. HARP, G. IRVING, M. ISARD, Y. JIA, R. JOZEFOWICZ, L. KAISER, M. KUDLUR, J. LEVENBERG, D. MANÉ, R. MONGA, S. MOORE, D. MURRAY, C. OLAH, M. SCHUSTER, J. SHLENS, B. STEINER, I. SUTSKEVER, K. TALWAR, P. TUCKER, V. VANHOUCHE, V. VASUDEVAN, F. VIÉGAS, O. VINIYALS, P. WARDEN, M. WATTENBERG, M. WICKE, Y. YU, AND X. ZHENG, *TensorFlow: Large-scale Machine Learning on Heterogeneous Systems*, 2015, <https://www.tensorflow.org/>, software available from tensorflow.org.
- [2] A. A. ALAWADHI, F. BOSO, AND D. M. TARTAKOVSKY, *Method of distributions for water-hammer equations with uncertain parameters*, *Water Resour. Res.*, 54 (2018), pp. 9398–9411.
- [3] D. AUROUX AND J. BLUM, *Back and forth nudging algorithm for data assimilation problems*, *C. R. Math. Acad. Sci. Paris*, 340 (2005), pp. 873–878.
- [4] D. AUROUX AND M. NODET, *The back and forth nudging algorithm for data assimilation problems: Theoretical results on transport equations*, *ESAIM Control Optim. Calc. Var.*, 18 (2012), pp. 318–342.
- [5] S. BLANDIN, A. COUQUE, A. BAYEN, AND D. WORK, *On sequential data assimilation for scalar macroscopic traffic flow models*, *Phys. D*, 241 (2012), pp. 1421–1440.
- [6] F. BOSO, S. V. BRODYA, AND D. M. TARTAKOVSKY, *Cumulative distribution function solutions of advection-reaction equations with uncertain parameters*, *Proc. R. Soc. A*, 470 (2014), 20140189.
- [7] A.-C. BOULANGER, P. MOIREAU, B. PERTHAME, AND J. SAINT-MARIE, *Data assimilation for hyperbolic conservation laws. A Luenberger observer approach based on a kinetic description*, *Commun. Math. Sci.*, 13 (2015), pp. 587–622.
- [8] G. CRACIUN, A. BROWN, AND A. FRIEDMAN, *A dynamical system model of neurofilament transport in axons*, *J. Theoret. Biol.*, 237 (2005), pp. 316–322.
- [9] C. F. DAGANZO, *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*, *Transp. Res. B*, 28 (1994), pp. 269–287.
- [10] B. DESPRÉS, *From uncertainty propagation in transport equations to kinetic polynomials*, in *Uncertainty Quantification for Hyperbolic and Kinetic Equations*, S. Jin and L. Pareschi, eds., SEMA SIMAI Springer Ser. 14, Springer, Cham, 2017, pp. 127–150.
- [11] S. K. GODUNOV, *A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics*, *Mat. Sb.*, 89 (1959), pp. 271–306.
- [12] B. GREENSHIELDS, J. R. BIBBINS, W. S. CHANNING, AND H. MILLER, *A Study of Traffic Capacity*, in *Highway Research Board Proceedings*, Volume 14, National Research Council (USA), Highway Research Board, 1935.

- [13] J. E. GUYER, D. WHEELER, AND J. A. WARREN, *FiPy: Partial differential equations with Python*, Comput. Sci. Engrg., 11 (2009), pp. 6–15.
- [14] J. C. HERRERA AND A. M. BAYEN, *Incorporation of Lagrangian measurements in freeway traffic state estimation*, Transport. Res. B Meth., 44 (2010), pp. 460–481.
- [15] K. HORNIK, M. STINCHCOMBE, AND H. WHITE, *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks*, Neural Net., 3 (1990), pp. 551–560.
- [16] A. H. JAZWINSKI, *Stochastic Processes and Filtering*, Dover, Mineola, NY, 1970.
- [17] S. JIN AND L. PARESCHI, *Uncertainty Quantification for Hyperbolic and Kinetic Equations*, SEMA SIMAI Springer Series, 14. Springer, Cham, 2017.
- [18] S. JIN, D. XIU, AND X. ZHU, *A well-balanced stochastic Galerkin method for scalar hyperbolic balance laws with random inputs*, J. Sci. Comput., 67 (2016), pp. 1198–1218.
- [19] S. LAKSHMIVARAHAN AND J. M. LEWIS, *Nudging Methods: A Critical Overview*, in Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), Springer, 2013, pp. 27–57.
- [20] J. LE NY AND G. J. PAPPAS, *Differentially private filtering*, IEEE Trans. Automat. Control, 59 (2014), pp. 341–354.
- [21] L. LEI AND J. P. HACKER, *Nudging, ensemble, and nudging ensembles for data assimilation in the presence of model error*, Mon. Weather Rev., 143 (2015), pp. 2600–2610.
- [22] P.-L. LIONS, B. PERTHAME, AND E. TADMOR, *A kinetic formulation of multidimensional scalar conservation laws and related equations*, J. Amer. Math. Soc., 7 (1994), pp. 169–191.
- [23] B. PERTHAME, *Kinetic Formulation of Conservation Laws*, Oxford Lecture Series in Math. Appl. 21, Oxford University Press, New York, 2002.
- [24] M. RAISSI, *Deep hidden physics models: Deep learning of nonlinear partial differential equations*, J. Mach. Learn. Res., 19 (2018), 25.
- [25] D. R. STAUFFER AND N. L. SEAMAN, *Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data*, Mon. Weather Rev., 118 (1990), pp. 1250–1277.
- [26] D. M. TARTAKOVSKY, M. DENTZ, AND P. C. LICHTNER, *Probability density functions for advective-reactive transport in porous media with uncertain reaction rates*, Water Resour. Res., 45 (2009), W07414, <https://doi.org/10.1029/2008WR007383>.
- [27] D. M. TARTAKOVSKY AND P. A. GREMAUD, *Method of distributions for uncertainty quantification*, in Handbook of Uncertainty Quantification, R. Ghanem, D. Higdon, and H. Owhadi, eds., Springer, Cham, 2017, pp. 763–783, https://doi.org/10.1007/978-3-319-12385-1_27.
- [28] E. F. TORO, *NUMERICA, a Library of Source Codes for Teaching, Research and Applications*, CD-ROM, Numeritek Ltd., Cheshire, UK, 1999.
- [29] E. F. TORO, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer-Verlag, Berlin, Heidelberg, 2013.
- [30] D. VENTURI, D. M. TARTAKOVSKY, A. M. TARTAKOVSKY, AND G. E. KARNIADAKIS, *Exact PDF equations and closure approximations for advective-reactive transport*, J. Comput. Phys., 243 (2013), pp. 323–343.
- [31] P. VIDARD, F.-X. L. DIMET, AND A. PIACENTINI, *Determination of optimal nudging coefficients*, Tellus A: Dynamic Meteorology and Oceanography, 55 (2003), pp. 1–15.
- [32] P. WANG AND D. M. TARTAKOVSKY, *Uncertainty quantification in kinematic wave models*, J. Comput. Phys., 231 (2012), pp. 7868–7880, <https://doi.org/10.1016/j.jcp.2012.07.030>.
- [33] P. WANG, D. M. TARTAKOVSKY, K. D. JARMAN, JR., AND A. M. TARTAKOVSKY, *CDF solutions of Buckley-Leverett equation with uncertain parameters*, Multiscale Model. Simul., 11 (2013), pp. 118–133, <https://doi.org/10.1137/120865574>.