DYNAMICS OF DATA-DRIVEN AMBIGUITY SETS FOR HYPERBOLIC CONSERVATION LAWS WITH UNCERTAIN INPUTS*

FRANCESCA BOSO*†, DIMITRIS BOSKOS*‡, JORGE CORTÉS‡, SONIA MARTÍNEZ‡, AND DANIEL M. TARTAKOVSKY†

Abstract. Ambiguity sets of probability distributions are used to hedge against uncertainty about the true probabilities of random quantities of interest (QoIs). When available, these ambiguity sets are constructed from both data (collected at the initial time and along the boundaries of the physical domain) and concentration-of-measure results on the Wasserstein metric. To propagate the ambiguity sets into the future, we use a physics-dependent equation governing the evolution of cumulative distribution functions (CDF) obtained through the method of distributions. This study focuses on the latter step by investigating the spatio-temporal evolution of data-driven ambiguity sets and their associated guarantees when the random QoIs they describe obey hyperbolic partial-differential equations with random inputs. For general nonlinear hyperbolic equations with smooth solutions, the CDF equation is used to propagate the upper and lower envelopes of pointwise ambiguity bands. For linear dynamics, the CDF equation allows us to construct an evolution equation for tighter ambiguity balls. We demonstrate that, in both cases, the ambiguity sets are guaranteed to contain the true (unknown) distributions within a prescribed confidence.

Key words. Uncertainty quantification, Wasserstein ambiguity sets, method of distributions

AMS subject classifications. 35R60,60H15,68T37,90C15,90C90

1. Introduction.

Hyperbolic conservation laws describe a wide spectrum of engineering applications ranging from multi-phase flows [8] to networked traffic [19]. The underlying dynamics is described by first-order hyperbolic partial differential equations (PDEs) with non-negligible parametric uncertainty, induced by factors such as limited and/or noisy measurements and random fluctuations of environmental attributes. Decisions based, in whole or in part, on predictions obtained from such models have to account for this uncertainty. The decision maker often has no distributional knowledge of the parametric uncertainties affecting the model and uses data—often noisy and insufficient—to make inferences about these distributions. Robust stochastic programming [2] calls for a quantifiable description of sets of probability measures, termed ambiguity sets, that contain the true (yet unknown) distribution with high confidence (e.g., [24, 13, 28]). The availability of such sets underpins distributionally robust optimization (DRO) formulations [2, 27] that are able of hedging against these uncertainties. Ambiguity sets are typically defined either through moment constraints [10] or statistical metric-like notions such as ϕ-divergences [1] and Wasserstein metrics [13], which allow the designer to identify distributions that are close to the nominal distribution in the prescribed metric. Ideally, ambiguity sets should be rich enough to contain the true distribution with high probability; be amenable to tractable reformulations; capture distribution variations relevant to the optimization problem without being overly conservative; and be data-driven. Wasserstein ambiguity sets have emerged as an appropriate choice because of two reasons. First, they provide computationally convenient dual reformulations of the associated DRO problems [13, 15]. Second, they penalize horizontal dislocations of the distributions [26], which considerably affect solutions of the stochastic optimization problems. Fur-

*Work supported by the DARPA Lagrange program through award N66001-18-2-4027. Both first authors contributed equally.

†Department of Energy Resources Engineering, Stanford University, ((fboso, tartakovsky) @stanford.edu).

‡Department of Mechanical and Aerospace Engineering, UC San Diego, ((dboskos, cortes, soniamd)@ucsd.edu).

thermore, data-driven Wasserstein ambiguity sets are accompanied by finite-sample guarantees of containing the true distribution with high confidence [14, 11, 33], resulting in DRO problems with prescribed out-of-sample performance. Our recent work [4, 5] has explored how ambiguity sets change under deterministic flow maps generated by ordinary differential equations, and used this information in dynamic DRO formulations. For these reasons, Wasserstein DRO formulations are utilized in a wide range of applications including distributed optimization [9], machine learning [3], traffic control [20], power systems [16], and logistics [17].

We consider two types of input ambiguity sets. The first is based on Wasserstein balls, whereas the second exploits CDF bands that contain the CDF of the true distribution with high probability. Our focus is on the spatio-temporal evolution of data-driven ambiguity sets (and their associated guarantees) when the random quantities they describe obey hyperbolic PDEs with random inputs. Many techniques can be used to propagate uncertainty affecting the inputs of a stochastic PDE to its solution. We use the method of distributions (MD) [30], which yields a deterministic evolution equation for the single-point cumulative distribution function (CDF) of a model output [6]. This method provides an efficient alternative to numerically demanding Monte Carlo simulations (MCS), which require multiple solutions of the PDE with repeated realizations of the random inputs. It is ideal for hyperbolic problems, for which other techniques (such us stochastic finite elements and stochastic collocation) can be slower than MCS [7]. In particular, when uncertainty in initial and boundary conditions is propagated by a hyperbolic deterministic PDE with a smooth solution, MD yields an exact CDF equation [31, 6]. Regardless of the uncertainty propagation technique, data can be used both to characterize the statistical properties of the input distributions and reduce uncertainty by assimilating observations into probabilistic model predictions via Bayesian techniques, e.g., [34].

The contributions of our study are threefold. First, we use data collected at the initial time and along the boundaries of the physical domain to build ambiguity sets that enjoy rigorous finite-sample guarantees for the input distributions. Specifically, we construct data-driven pointwise ambiguity sets for the *unknown* true distributions of parameterized random inputs, by transferring finite-sample guarantees for their associated Wasserstein distance in the parameter domain. The resulting ambiguity sets account for empirical information (from the data) without introducing arbitrary hypotheses on the distribution of the random parameters. Second, we design tools to propagate the ambiguity sets throughout space and time. The MD is employed to propagate each ambiguous distribution within the data-driven input ambiguity sets according to a physics-dependent CDF equation. For linear dynamics, we use the CDF equation to construct an evolution equation for the radius of ambiguity balls centered at the empirical distributions in the 1-Wasserstein (a.k.a. Kantorovich) metric. For a wider class of nonlinear hyperbolic equations with smooth solutions, we exploit the CDF equation to propagate the upper and lower envelopes of pointwise ambiguity bands. These are formed through upper and lower envelopes that contain all CDFs up to an assigned 1-Wasserstein distance from the empirical CDF. Third, we use these uncertainty propagation tools to obtain pointwise ambiguity sets across all locations of the space-time domain that contain their true distributions with prescribed probability. Our method can handle both types of input ambiguity sets (based on either Wasserstein balls or CDF bands), while maintaining their confidence guarantees upon propagation. This allows the decision maker to map their physics-driven stretching/shrinking under the PDE dynamics.

2. Preliminaries. Let $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and infinity norm in $\mathbb{R}^n$, respectively. The diameter of a set $S \subset \mathbb{R}^n$ is defined as $\operatorname{diam}(S) := \sup\{\|x - y\|_\infty \,|\, x, y \in S\}$. The Heaviside function $\mathcal{H} : \mathbb{R} \to \mathbb{R}$ is $\mathcal{H}(x) = 0$ for $x < 0$ and $\mathcal{H}(x) = 1$ for $x \geq 0$. We denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-algebra on $\mathbb{R}^d$, and by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For $\mu \in \mathcal{P}(\mathbb{R}^d)$, its support is the closed set $\operatorname{supp}(\mu) := \{x \in \mathbb{R}^d \,|\, \mu(U) > 0$ for each neighborhood U of $x\}$ or, equivalently, the smallest closed set with measure one. We denote by $\operatorname{Cdf}[P]$ the cumulative distribution function associated with the probability measure P on $\mathbb{R}$ and by $\mathcal{CD}(I)$ the set of all CDFs of scalar random variables whose induced probability measures are supported on the interval $I \subset \mathbb{R}$. Given $p \geq 1$, $\mathcal{P}_p(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) \,|\, \int_{\mathbb{R}^d} \|x\|^p d\mu < \infty\}$ is the set of probability measures in $\mathcal{P}(\mathbb{R}^d)$ with finite p-th moment. The Wasserstein distance of $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ is

$$W_p(\mu, \nu) := \Big(\inf_{\pi \in \mathcal{M}(\mu,\nu)} \Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \pi(dx, dy) \Big\} \Big)^{1/p},$$

where $\mathcal{M}(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν, respectively, also termed couplings. For scalar random variables, the Wasserstein distance W_p between two distributions μ and ν with CDFs F and G is, cf. [32], $W_p(\mu, \nu) = \big(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt \big)^{1/p}$, where F^{-1} denotes the generalized inverse of F, $F^{-1}(y) = \inf\{t \in \mathbb{R} \,|\, F(t) > y\}$. For $p = 1$, one can use the representation

$$W_1(\mu, \nu) = \int_{\mathbb{R}} |F(s) - G(s)| ds. \tag{2.1}$$

Given two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, and a measurable function Ψ from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$, the push-forward map $\Psi_\#$ assigns to each measure μ in $(\Omega, \mathcal{F})$ a new measure ν in $(\Omega', \mathcal{F}')$ defined by $\nu := \Psi_\# \mu$ iff $\nu(B) = \mu(\Psi^{-1}(B))$ for all $B \in \mathcal{F}'$. The map $\Psi_\#$ is linear and satisfies $\Psi_\# \delta_\omega = \delta_{\Psi(\omega)}$ with δ_ω the Dirac mass at $\omega \in \Omega$.

3. Problem formulation. We consider a hyperbolic model for $u(\mathbf{x}, t)$,

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{q}(u; \boldsymbol{\theta}_q)) = r(u; \boldsymbol{\theta}_r), \quad \mathbf{x} \in \Omega, \quad t > 0 \tag{3.1}$$

subject to initial and boundary conditions

$$\begin{aligned} &u(\mathbf{x}, t = 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \Omega \\ &u(\mathbf{x}, t) = u_b(\mathbf{x}, t), \quad \mathbf{x} \in \Gamma, \quad t > 0, \end{aligned} \tag{3.2}$$

restricting ourselves to problems with smooth solutions. Equation (3.1), with the given flux $\mathbf{q}(u; \boldsymbol{\theta}_q)$ and source term $r(u; \boldsymbol{\theta}_r)$, is defined on a d-dimensional semi-infinite spatial domain $\Omega \subset \mathbb{R}^d$, and by the parameters $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_r$, that can be spatially and/or temporally varying. The boundary function $u_b(\mathbf{x}, t)$ is prescribed at the upstream boundary Γ. For the sake of brevity, we do not consider different types of boundary conditions, although the procedure can be adjusted accordingly. Randomness in the initial and/or boundary conditions, $u_0(\mathbf{x})$ and $u_b(\mathbf{x}, t)$, renders (3.1) stochastic. We make the following hypotheses.

ASSUMPTION 3.1 (Deterministic dynamics). *We assume all parameters in* (3.1) *(i.e., all physical parameters specifying the flux* $\mathbf{q}$, $\boldsymbol{\theta}_q$*, and the source term* r, $\boldsymbol{\theta}_r$*) are deterministic, and the flux* $\mathbf{q}$ *is divergence-free once evaluated for a specific value of* $u(\mathbf{x}, t) = U$, $\nabla \cdot \mathbf{q}(U; \boldsymbol{\theta}_q) = 0$.

Assumption 3.2 (Existence and uniqueness of local solutions within a time horizon). *There exists $T \in (0, \infty]$ such that for each initial and boundary condition from their probability space, the solution $u(\mathbf{x}, t)$ of (3.1) is smooth and defined on $\Omega \times [0, T)$.*

Regarding Assumption 3.2, we refer to [25] for a theoretical treatment of local existence theorems. In the *absence of direct access to the distribution of the initial and boundary conditions*, we analyze their samples from independent realizations of (3.2). Specifically, we measure the initial condition u_0 for all $\mathbf{x} \in \Omega$ and get continuous measurements of u_b at each boundary point for all times (for instance, in a traffic flow scenario with Ω representing a long highway segment, a traffic helicopter might pass above the area at the same time each morning and take a photo from the segment that provides the initial condition for the traffic density u, whereas u at the segment boundary is continuously measured by a single-loop detector. Assumptions 3.1 and 3.2 require traffic conditions far from congestion, with deterministic parameters describing the flow, specifically maximum velocity and maximum traffic density). We are interested in exploiting the samples to construct ambiguity sets that contain the temporally- and spatially-variable one-point probability distributions of $u_0(t)$ and $u_b(\mathbf{x}, t)$ with high confidence. We consider initial and boundary conditions that are specified by a finite number of random parameters.

Assumption 3.3 (Input parameterization). *The initial and boundary conditions are parameterized by $\mathbf{a} := (a_1, \ldots, a_n)$ from a compact subset of $\mathbb{R}^n$, i.e., $u_0(\mathbf{x}) \equiv u_0(\mathbf{x}; \mathbf{a})$ and $u_b(\mathbf{x}, t) \equiv u_b(\mathbf{x}, t; \mathbf{a})$. The parameterizations are globally Lipschitz with respect to $\mathbf{a}$ for each initial position $\mathbf{x}$ and boundary pair $(\mathbf{x}, t)$. Specifically,*

$$|u_0(\mathbf{x}; \mathbf{a}) - u_0(\mathbf{x}; \mathbf{a}')| \leq L_0(\mathbf{x}) \|\mathbf{a} - \mathbf{a}'\| \quad \forall \mathbf{x} \in \Omega, \quad \mathbf{a}, \mathbf{a}' \in \mathbb{R}^n, \tag{3.3a}$$

$$|u_b(\mathbf{x}, t; \mathbf{a}) - u_b(\mathbf{x}, t; \mathbf{a}')| \leq L_b(\mathbf{x}, t) \|\mathbf{a} - \mathbf{a}'\| \quad \forall \mathbf{x} \in \Gamma, \quad t \geq 0, \quad \mathbf{a}, \mathbf{a}' \in \mathbb{R}^n, \tag{3.3b}$$

for some continuous functions $L_0 : \Omega \to \mathbb{R}_{\geq 0}$ and $L_b : \Gamma \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$.

We denote by $P_{\mathbf{a}}^{\text{true}}$ the distribution of the parameters in $\mathbb{R}^n$, by $P_{u_0(\mathbf{x})}^{\text{true}}$ the induced distribution of $u_0(\mathbf{x}; \mathbf{a})$ at the spatial point $\mathbf{x}$, and by $P_{u_b(\mathbf{x},t)}^{\text{true}}$ the distribution of $u_b(\mathbf{x}, t; \mathbf{a})$ at each boundary point $\mathbf{x}$ and time $t \geq 0$. We use the superscript 'true' to emphasize that we refer to the corresponding true distributions, that are unknown. We denote by $F_{u_0(\mathbf{x})}^{\text{true}} \equiv \text{Cdf}\left[P_{u_0(\mathbf{x})}^{\text{true}}\right]$ and $F_{u_b(\mathbf{x},t)}^{\text{true}} \equiv \text{Cdf}\left[P_{u_b(\mathbf{x},t)}^{\text{true}}\right]$ their associated CDFs and make the following hypothesis for data assimilation.

Assumption 3.4 (Input samples). *We have access to N independent pairs of initial and boundary condition samples, $(u_0^1, u_b^1), \ldots, (u_0^N, u_b^N)$, generated by corresponding independent realizations $\mathbf{a}^1, \ldots, \mathbf{a}^N$ of the parameters in Assumption 3.3.*

Under these hypotheses, we seek to derive *pointwise characterizations* of ambiguity sets for the CDF of u at each location $(\mathbf{x}, t)$ in space and time, starting with their characterization for the initial and boundary data. We are interested in defining the ambiguity sets in terms of plausible CDFs at each $(\mathbf{x}, t)$, and exploiting the known dynamics (3.1) to propagate the one-point CDFs of $u(\mathbf{x}, t)$ in space and time.

Problem statement. *Given β, we seek to determine sets $\mathcal{P}_{\mathbf{x}}^0$, $\mathbf{x} \in \Omega$ and $\mathcal{P}_{\mathbf{x},t}^b$, $(\mathbf{x}, t) \in \Gamma \times \mathbb{R}_{\geq 0}$ of CDFs that contain the corresponding true CDFs $F_{u_0(\mathbf{x})}^{\text{true}}$ and $F_{u_b(\mathbf{x},t)}^{\text{true}}$ for the initial and boundary conditions, respectively, with confidence $1 - \beta$,*

$$\mathbb{P}(\{F_{u_0(\mathbf{x})}^{\text{true}} \in \mathcal{P}_{\mathbf{x}}^0 \; \forall \mathbf{x} \in \Omega\} \cap \{F_{u_b(\mathbf{x},t)}^{\text{true}} \in \mathcal{P}_{\mathbf{x},t}^b \; \forall (\mathbf{x}, t) \in \Gamma \times \mathbb{R}_{\geq 0}\}) \geq 1 - \beta.$$

We further seek to leverage the PDE dynamics to propagate the ambiguity sets of the initial and boundary data and obtain a pointwise characterization of ambiguity sets

$\mathcal{P}_{\mathbf{x},t}$ *containing the CDF of* $u(\mathbf{x},t)$ *at each* $\mathbf{x}\in\Omega$ *and* $t\in[0,T)$ *with confidence* $1-\beta$*,*

$$\mathbb{P}(F^{\text{true}}_{u(\mathbf{x},t)} \in \mathcal{P}_{\mathbf{x},t}\ \forall(\mathbf{x},t)\in\Omega\times[0,T)) \geq 1-\beta.$$

Section 4 exploits the compactly supported parameterization of the initial and boundary data to build ambiguity sets which enjoy rigorous finite-sample guarantees. Section 5 derives a deterministic PDE for the CDF of $u(\mathbf{x},t)$, which enables the investigation of how the difference between CDFs (and, by integration, their Wasserstein distance) evolves in space and time. Section 6 characterizes how the input ambiguity sets propagate in space and time under the same confidence guarantees.

4. Data-driven ambiguity sets for inputs.

Using Assumptions 3.3 and 3.4, at each $\mathbf{x}\in\Omega$ and boundary pair $(\mathbf{x},t)\in\Gamma\times\mathbb{R}_{\geq 0}$, we define empirical distributions

$$\widehat{P}^N_{u_0(\mathbf{x})} \equiv \widehat{P}^N_{u_0(\mathbf{x})}(\mathbf{a}^1,\dots,\mathbf{a}^N) := \frac{1}{N}\sum_{i=1}^N \delta_{u_0^i(\mathbf{x})} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{u_0(\mathbf{x};\mathbf{a}^i)},$$

$$\widehat{P}^N_{u_b(\mathbf{x},t)} \equiv \widehat{P}^N_{u_b(\mathbf{x},t)}(\mathbf{a}^1,\dots,\mathbf{a}^N) := \frac{1}{N}\sum_{i=1}^N \delta_{u_b^i(\mathbf{x},t)} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{u_b(\mathbf{x},t;\mathbf{a}^i)},$$

with associated CDFs $\widehat{F}^N_{u_0(\mathbf{x})} := \mathrm{Cdf}\big[\widehat{P}^N_{u_0(\mathbf{x})}\big]$ and $\widehat{F}^N_{u_b(\mathbf{x},t)} := \mathrm{Cdf}\big[\widehat{P}^N_{u_b(\mathbf{x},t)}\big]$. We employ these empirical distributions to build *pointwise ambiguity sets* based on concentration-of-measure results for the 1-Wasserstein distance. Specifically, we exploit compactness of the initial and boundary data parameterization together with the following confidence guarantees about the Wasserstein distance between the empirical and true distribution of compactly supported random variables (see [5]).

LEMMA 4.1 (Ambiguity radius). *Let* $(X_i)_{i\in\mathbb{N}}$ *be a sequence of i.i.d.* $\mathbb{R}^n$*-valued random variables that have a compactly supported distribution* μ *and let* $\rho := \mathrm{diam}(\mathrm{supp}(\mu))/2$*. Then, for* $p\geq 1$*,* $N\geq 1$*, and* $\epsilon>0$*,* $\mathbb{P}(W_p(\widehat{\mu}^N,\mu)\leq \epsilon_N(\beta,\rho))\geq 1-\beta$*, where*

$$\epsilon_N(\beta,\rho) := \begin{cases} \left(\frac{\ln(C\beta^{-1})}{c}\right)^{\frac{1}{2p}} \frac{\rho}{N^{\frac{1}{2p}}}, & \text{if } p > n/2,\\ h^{-1}\left(\frac{\ln(C\beta^{-1})}{cN}\right)^{\frac{1}{p}}\rho, & \text{if } p = n/2,\\ \left(\frac{\ln(C\beta^{-1})}{c}\right)^{\frac{1}{n}} \frac{\rho}{N^{\frac{1}{n}}}, & \text{if } p < n/2, \end{cases} \tag{4.1}$$

$\widehat{\mu}^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}$*, the constants* C *and* c *depend only on* p*,* n*, and* h^{-1} *is the inverse of* $h(x) = x^2/[\ln(2+1/x)]^2$*,* $x>0$*.*

This result quantifies the radius $\epsilon_N(\beta,\rho)$ of an ambiguity ball that contains the true distribution with high probability. The radius decreases with the number of samples and can be tuned by the confidence level $1-\beta$, allowing the decision maker to choose the desired level of conservativeness. The explicit determination of c and C in (4.1) through the analysis in [14] for the whole spectrum of data dimensions n and Wasserstein exponents p can become cumbersome. Nevertheless, (4.1) provides explicit ambiguity radius ratios for any pair of sample sizes once a confidence level is fixed. Recall that, according to Assumption 3.3, the mapping of the parameters to the initial and boundary data is globally Lipschitz. The following result, whose proof is given in Appendix A, is useful to quantify the Wasserstein distance between the true and empirical distribution at each input location.

Lemma 4.2 (Wasserstein distance under Lipschitz maps). *If $T:\mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz with constant $L>0$, namely, $\|T(x)-T(y)\| \le L\|x-y\|$, then for any pair of distributions μ, ν on $\mathbb{R}^n$ it holds that $W_p(\mu,\nu) \le L W_p(T_\#\mu, T_\#\nu)$.*

Using Lemmas 4.1 and 4.2 together with the finite-sample guarantees in the parameter domain, we next obtain a characterization of initial and boundary value ambiguity sets through pointwise Wasserstein balls. To express the ambiguity sets in terms of CDFs, we will interchangeably denote by $W_p(F_{X_1}, F_{X_2}) \equiv W_p(P_{X_1}, P_{X_2})$ the Wasserstein distance between any two scalar random variables X_1, X_2 with distributions P_{X_1}, P_{X_2} and associated CDFs $F_{X_1} = \mathrm{Cdf}[P_{X_1}]$, $F_{X_2} = \mathrm{Cdf}[P_{X_2}]$.

Proposition 4.3 (Input ambiguity sets). *Assume that N pairs of input samples are collected according to Assumption 3.4 and let*

$$\rho_{\mathbf{a}} := \mathrm{diam}(\mathrm{supp}(P^{\mathrm{true}}_{\mathbf{a}}))/2 \tag{4.2}$$

and $\bar{\mathbf{a}} \in \mathbb{R}^n$ such that $\|\mathbf{a}-\bar{\mathbf{a}}\|_\infty \le \rho_{\mathbf{a}}$ for all $\mathbf{a} \in \mathrm{supp}(P^{\mathrm{true}}_{\mathbf{a}})$. Given a confidence level $1-\beta$, define the ambiguity sets

$$\mathcal{P}^0_{\mathbf{x}} := \big\{F \in \mathcal{CD}([\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})]) \,|\, W_1(\widehat{F}^N_{u_0(\mathbf{x})}, F) \le L_0(\mathbf{x})\epsilon_N(\beta, \rho_{\mathbf{a}})\big\}$$
$$\mathcal{P}^b_{\mathbf{x},t} := \big\{F \in \mathcal{CD}([\alpha_b(\mathbf{x},t), \gamma_b(\mathbf{x},t)]) \,|\, W_1(\widehat{F}^N_{u_b(\mathbf{x},t)}, F) \le L_b(\mathbf{x},t)\epsilon_N(\beta, \rho_{\mathbf{a}})\big\},$$

for $\mathbf{x} \in \Omega$ and $\mathbf{x} \in \Gamma$, $t \ge 0$, respectively, where

$$[\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})] := [u_0(\mathbf{x};\bar{\mathbf{a}}) - \sqrt{n}L_0(\mathbf{x})\rho_{\mathbf{a}}, u_0(\mathbf{x};\bar{\mathbf{a}}) + \sqrt{n}L_0(\mathbf{x})\rho_{\mathbf{a}}] \tag{4.3a}$$
$$[\alpha_b(\mathbf{x},t), \gamma_b(\mathbf{x},t)] := [u_b(\mathbf{x},t;\bar{\mathbf{a}}) - \sqrt{n}L_b(\mathbf{x},t)\rho_{\mathbf{a}}, u_b(\mathbf{x},t;\bar{\mathbf{a}}) + \sqrt{n}L_b(\mathbf{x},t)\rho_{\mathbf{a}}], \tag{4.3b}$$

and $L_0(\mathbf{x})$, $L_b(\mathbf{x},t)$, and $\epsilon_N(\beta,\rho_{\mathbf{a}})$ are given by (3.3a), (3.3b), and (4.1). Then,

$$\mathbb{P}(\{F^{\mathrm{true}}_{u_0(\mathbf{x})} \in \mathcal{P}^0_{\mathbf{x}}\ \forall \mathbf{x} \in \Omega\} \cap \{F^{\mathrm{true}}_{u_b(\mathbf{x},t)} \in \mathcal{P}^b_{\mathbf{x},t}\ \forall (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\ge 0}\}) \ge 1-\beta. \tag{4.4}$$

Proof. For the selected confidence $1-\beta$, we get from Lemma 4.1 with $p=1$ that

$$\mathbb{P}(W_1(\widehat{P}^N_{\mathbf{a}}, P^{\mathrm{true}}_{\mathbf{a}}) \le \epsilon_N(\beta,\rho_{\mathbf{a}})) \ge 1-\beta. \tag{4.5}$$

Denoting by $u_0[\mathbf{x}]$ the mapping $\mathbf{a} \mapsto u_0[\mathbf{x}](\mathbf{a}) := u_0(\mathbf{x};\mathbf{a})$, it follows from elementary properties of the pushforward map given in section 2 that $\widehat{P}^N_{u_0(\mathbf{x})} = u_0[\mathbf{x}]_\#\widehat{P}^N_{\mathbf{a}}$ and $P^{\mathrm{true}}_{u_0(\mathbf{x})} = u_0[\mathbf{x}]_\# P^{\mathrm{true}}_{\mathbf{a}}$, where $\widehat{P}^N_{\mathbf{a}} := \frac{1}{N}\sum_{i=1}^N \delta_{\mathbf{a}^i}$. Thus, we obtain from the Lipschitz hypothesis (3.3a) and Lemma 4.2 that

$$W_1(\widehat{P}^N_{u_0(\mathbf{x})}, P^{\mathrm{true}}_{u_0(\mathbf{x})}) \le L_0(\mathbf{x}) W_1(\widehat{P}^N_{\mathbf{a}}, P^{\mathrm{true}}_{\mathbf{a}}), \quad \forall \mathbf{x} \in \Omega.$$

Since $P^{\mathrm{true}}_{u_0(\mathbf{x})} = u_0[\mathbf{x}]_\# P^{\mathrm{true}}_{\mathbf{a}}$, we get from (3.3a), (4.3a), and the selection of $\bar{\mathbf{a}}$ that $P^{\mathrm{true}}_{u_0(\mathbf{x})}$ is supported on $[\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})]$, and hence, that $F^{\mathrm{true}}_{u_0(\mathbf{x})} \in \mathcal{CD}([\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})])$ for all $\mathbf{x} \in \Omega$. Analogously, we have that

$$W_1(\widehat{P}^N_{u_b(\mathbf{x},t)}, P^{\mathrm{true}}_{u_b(\mathbf{x},t)}) \le L_b(\mathbf{x},t) W_1(\widehat{P}^N_{\mathbf{a}}, P^{\mathrm{true}}_{\mathbf{a}})$$

and $F^{\mathrm{true}}_{u_b(\mathbf{x},t)} \in \mathcal{CD}([\alpha_b(\mathbf{x},t), \gamma_b(\mathbf{x},t)])$ for all $(\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\ge 0}$. Consequently

$$\{W_1(\widehat{P}^N_{\mathbf{a}}, P^{\mathrm{true}}_{\mathbf{a}}) \le \epsilon_N(\beta,\rho_{\mathbf{a}})\} \subset \{W_1(\widehat{P}^N_{u_0(\mathbf{x})}, P^{\mathrm{true}}_{u_0(\mathbf{x})}) \le L_0(\mathbf{x})\epsilon_N(\beta,\rho_{\mathbf{a}})\ \forall \mathbf{x} \in \Omega\}$$
$$\cap \{W_1(\widehat{P}^N_{u_b(\mathbf{x},t)}, P^{\mathrm{true}}_{u_b(\mathbf{x},t)}) \le L_b(\mathbf{x},t)\epsilon_N(\beta,\rho_{\mathbf{a}})\ \forall (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\ge 0}\}.$$

Thus, since each $F^{\mathrm{true}}_{u_0(\mathbf{x})} \in \mathcal{CD}([\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})])$ and $F^{\mathrm{true}}_{u_b(\mathbf{x},t)} \in \mathcal{CD}([\alpha_b(\mathbf{x},t), \gamma_b(\mathbf{x},t)])$, we deduce (4.4) from the definitions of the ambiguity sets. ∎

We next consider an alternative characterization of the ambiguity sets, which enables the exploitation of a propagation tool applicable to a wider class of PDE dynamics, yet at the cost of increased conservativeness. These ambiguity sets are built using pointwise confidence bands (thereinafter termed *ambiguity bands*), enclosed between upper and lower CDF envelopes that contain the true CDF at each spatio-temporal location with prescribed probability. We rely on the next result, whose proof is given in Appendix A, providing upper and lower CDF envelopes for any CDF F and distance ρ, cf. Figure 1, so that the CDF of any distribution with 1-Wasserstein distance at most ρ from F is pointwise between these envelopes.

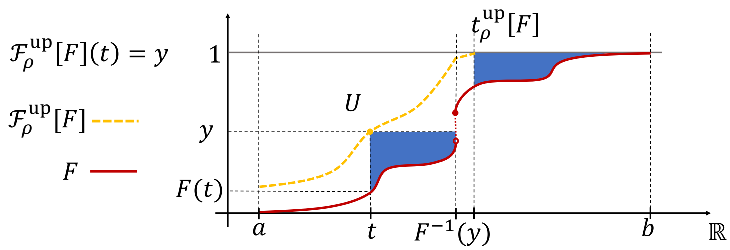

FIG. 1. *Illustration of the upper CDF envelope $\mathcal{F}_\rho^{\mathrm{up}}[F]$ (in yellow) of F (in red). For each point (t, y) in the graph of $\mathcal{F}_\rho^{\mathrm{up}}[F]$, the blue area enclosed among the lines parallel to the axes that originate from (t, y) and F is equal to ρ.*

LEMMA 4.4 (Upper and lower CDF envelopes). *Let $F \in \mathcal{CD}([a, b])$, define*

$$t_\rho^{\mathrm{up}}[F] \equiv t_{\rho,[a,b]}^{\mathrm{up}}[F] := \sup \Big\{ \tau \in [a, b] \,\Big|\, \int_\tau^b (1 - F(t))dt \geq \rho \Big\}$$

$$t_\rho^{\mathrm{low}}[F] \equiv t_{\rho,[a,b]}^{\mathrm{low}}[F] := \inf \Big\{ \tau \in [a, b] \,\Big|\, \int_a^\tau F(t)dt \geq \rho \Big\}$$

for any $0 < \rho \leq \min\{\int_a^b F(t)dt, \int_a^b (1 - F(t))dt\}$, and the corresponding upper and lower CDF envelopes $\mathcal{F}_\rho^{\mathrm{up}}[F] \equiv \mathcal{F}_{\rho,[a,b]}^{\mathrm{up}}[F]$ and $\mathcal{F}_\rho^{\mathrm{low}}[F] \equiv \mathcal{F}_{\rho,[a,b]}^{\mathrm{low}}[F]$

$$\mathcal{F}_\rho^{\mathrm{up}}[F](t) := \begin{cases} 0, & \text{if } t \in (-\infty, a) \\ \sup \Big\{ z \in [F(t), 1] \,\Big|\, \int_{F(t)}^z (F^{-1}(y) - t)dy \leq \rho \Big\}, & \text{if } t \in [a, t_\rho^{\mathrm{up}}[F]) \\ 1, & \text{if } t \in [t_\rho^{\mathrm{up}}[F], \infty), \end{cases}$$

$$\mathcal{F}_\rho^{\mathrm{low}}[F](t) := \begin{cases} 0, & \text{if } t \in (-\infty, t_\rho^{\mathrm{low}}[F]) \\ \inf \Big\{ z \in [0, F(t)] \,\Big|\, \int_z^{F(t)} (t - F^{-1}(y))dy \leq \rho \Big\}, & \text{if } t \in [t_\rho^{\mathrm{low}}[F], b) \\ 1, & \text{if } t \in [b, \infty). \end{cases}$$

Then, both $\mathcal{F}_\rho^{\mathrm{up}}[F]$ and $\mathcal{F}_\rho^{\mathrm{low}}[F]$ are continuous CDFs in $\mathcal{CD}([a, b])$ and for any $F' \in \mathcal{CD}([a, b])$ with $W_1(F, F') \leq \rho$, it holds that

$$\mathcal{F}_\rho^{\mathrm{low}}[F](t) \leq F'(t) \leq \mathcal{F}_\rho^{\mathrm{up}}[F](t), \quad \forall t \in \mathbb{R}. \tag{4.6}$$

We rely on Lemma 4.4 to obtain in the next result ambiguity bands for the inputs that share the confidence guarantees with the ambiguity sets of Proposition 4.3.

COROLLARY 4.5 (Input ambiguity bands). *Assume N pairs of input samples are collected according to Assumption 3.4 and let $\rho_{\mathbf{a}}$ and $\bar{\mathbf{a}}$ as in the statement of*

Proposition 4.3. Given a confidence level $1-\beta$, define the ambiguity sets

$$\mathcal{P}_{\mathbf{x}}^{0,\mathrm{Env}} := \big\{F \in \mathcal{CD}(\mathbb{R}) \,|\, \mathcal{F}^{\mathrm{low}}_{\rho_0(\mathbf{x}),[\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})]}[\widehat{F}^N_{u_0(\mathbf{x})}](U) \le F(U) \le \mathcal{F}^{\mathrm{up}}_{\rho_0(\mathbf{x}),[\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})]}[\widehat{F}^N_{u_0(\mathbf{x})}](U) \;\forall U \in \mathbb{R}\big\},$$

$$\mathcal{P}_{\mathbf{x},t}^{b,\mathrm{Env}} := \big\{F \in \mathcal{CD}(\mathbb{R}) \,|\, \mathcal{F}^{\mathrm{low}}_{\rho_b(\mathbf{x},t),[\alpha_b(\mathbf{x},t),\gamma_b(\mathbf{x},t)]}[\widehat{F}^N_{u_b(\mathbf{x},t)}](U) \le F(U) \le \mathcal{F}^{\mathrm{up}}_{\rho_b(\mathbf{x},t),[\alpha_b(\mathbf{x},t),\gamma_b(\mathbf{x},t)]}[\widehat{F}^N_{u_b(\mathbf{x},t)}](U) \;\forall U \in \mathbb{R}\big\},$$

for $\mathbf{x} \in \Omega$ and $(\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\ge 0}$, respectively, where

$$\rho_0(\mathbf{x}) := L_0(\mathbf{x})\epsilon_N(\beta,\rho_{\mathbf{a}}) \tag{4.8a}$$

$$\rho_b(\mathbf{x},t) := L_b(\mathbf{x},t)\epsilon_N(\beta,\rho_{\mathbf{a}}), \tag{4.8b}$$

and $[\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})]$, $[\alpha_b(\mathbf{x},t),\gamma_b(\mathbf{x},t)]$, $\epsilon_N(\beta,\rho_{\mathbf{a}})$ given by (4.3a), (4.3b), and (4.1). Then

$$\mathbb{P}(\{F^{\mathrm{true}}_{u_0(\mathbf{x})} \in \mathcal{P}_{\mathbf{x}}^{0,\mathrm{Env}} \;\forall \mathbf{x} \in \Omega\} \cap \{F^{\mathrm{true}}_{u_b(\mathbf{x},t)} \in \mathcal{P}_{\mathbf{x},t}^{b,\mathrm{Env}} \;\forall (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\ge 0}\}) \ge 1-\beta. \tag{4.9}$$

Proof. By (4.4) and (4.9), it suffices to show that $\mathcal{P}_{\mathbf{x}}^0 \subset \mathcal{P}_{\mathbf{x}}^{0,\mathrm{Env}}$ and $\mathcal{P}_{\mathbf{x},t}^b \subset \mathcal{P}_{\mathbf{x},t}^{b,\mathrm{Env}}$ for all $\mathbf{x} \in \Omega$ and $(\mathbf{x},t) \in \Omega \times \mathbb{R}_{\ge 0}$, respectively, with $\mathcal{P}_{\mathbf{x}}^0$ and $\mathcal{P}_{\mathbf{x},t}^b$ given in Proposition 4.3. Let $\mathbf{x} \in \Omega$ and $F \in \mathcal{P}_{\mathbf{x}}^0$. Then, we get from the definition of $\mathcal{P}_{\mathbf{x}}^0$ and (4.8a) that $F \in \mathcal{CD}([\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})])$ and $W_1(\widehat{F}^N_{u_0},F) \le L_0(\mathbf{x})\epsilon_N(\beta,\rho_{\mathbf{a}}) = \rho_0(\mathbf{x})$. Thus, since $F \in \mathcal{CD}([\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})])$, we can invoke Lemma 4.4 and deduce from (4.6) that $F \in \mathcal{P}_{\mathbf{x}}^{0,\mathrm{Env}}$. Analogously, $\mathcal{P}_{\mathbf{x},t}^b \subset \mathcal{P}_{\mathbf{x},t}^{b,\mathrm{Env}}$ for all $(\mathbf{x},t) \in \Omega \times \mathbb{R}_{\ge 0}$. ∎

REMARK 4.6 (Confidence bands for components of non-scalar random variables). Confidence bands for *scalar* random variables are well-studied in the statistics literature [22]. Their construction has been originally based on the Kolmogorov-Smirnov test [18], [29], for which rigorous confidence guarantees have been introduced in [12] and further refined in [21]. A key difference of our approach is that we obtain analogous guarantees for an infinite (in fact uncountable) number of random variables, indexed by all spatio-temporal locations. This is achievable by using the Wasserstein ball guarantees in the finite-dimensional but in general *non-scalar* parameter space. Therefore, resorting to traditional confidence band guarantees [21] is possible only in the restrictive case where we consider a single random parameter for the inputs. □

We next present explicit constructions for the upper and lower CDF envelopes of the empirical CDF. For $n,m \in \mathbb{N}$ and $t \in \mathbb{R}$, we use the conventions $[n:m] = \emptyset$ when $m < n$ and $[t,t) = \emptyset$. The proof of the following result is given in Appendix A.

PROPOSITION 4.7 (Upper CDF envelope for discrete distributions). *Let $\widehat{F} \in \mathcal{CD}([a,b])$ be the CDF of a discrete distribution with positive mass c_i at a finite number of points t_i, $i \in [1:N]$ satisfying $a =: t_0 \le t_1 < \cdots < t_N \le b$ and define $b_{i,j} := \sum_{k=j}^{i}(t_k - t_j)c_k$, for $0 \le j \le i \le N$, (with $b_{i,j} = 0$ for any other $i,j \in \mathbb{N}_0$). Given $\rho > 0$ with $b_{N,0} = \sum_{i=1}^{N}(t_i - a)c_i > \rho$, let $j_1 := 0$, $i_1 := \min\{i \in [1:N] \,|\, b_{i,0} \ge \rho\}$ and*

$$j_{k+1} := \max\{j \in [j_k : i_k] \,|\, b_{i_k,j} \ge \rho\} + 1, \quad k = 1,\ldots,k_{\max}$$

$$i_{k+1} := \min\{i \in [i_k + 1 : N] \,|\, b_{i,j_{k+1}} \ge \rho\}, \quad k = 1,\ldots,k_{\max}-1,$$

where $k_{\max} := \min\{k \in \mathbb{N} \,|\, b_{N,j_{k+1}} \le \rho\}$. Then, all indices j_k, i_k are well defined and

$$j_k < j_{k+1} \le i_k < i_{k+1} \quad \forall k \in [1:k_{\max}], \tag{4.10}$$

where $i_{k_{\max}+1} := N+1$*. Also, for each* $k \in [1:k_{\max}]$*, let*

$$\Delta t_\ell := \frac{\rho - b_{\ell, j_{k+1}}}{\sum_{l=j_{k+1}}^{\ell} c_l}, \quad \tau_\ell := t_{j_{k+1}} - \Delta t_\ell, \quad \ell \in [i_k : i_{k+1}-1]$$

$$\Delta y_\ell := \frac{\rho - b_{i_k - 1, \ell}}{t_{i_k} - t_\ell}, \quad y_\ell := \sum_{l=1}^{i_k - 1} c_l + \Delta y_\ell, \quad \ell \in [j_k : j_{k+1}-1].$$

Then, τ_ℓ *are defined for all* $\ell \in [i_1 : N]$ *and form a strictly increasing sequence with*

$$\begin{aligned} t_0 = t_{j_1} &\le \cdots \le t_{j_2-1} \le \tau_{i_1} \le \cdots \le \tau_{i_2-1} < t_{j_2} \le \cdots \\ &\le t_{j_k} \le \cdots \le t_{j_{k+1}-1} \le \tau_{i_k} \le \cdots \le \tau_{i_{k+1}-1} < t_{j_{k+1}} \le \cdots \\ &\le t_{j_{k_{\max}}} \le \cdots \le t_{j_{k_{\max}+1}-1} \le \tau_{i_{k_{\max}}} \le \cdots \\ &\le \tau_{i_{k_{\max}+1}-1} = \tau_N < t_{j_{k_{\max}+1}} \le t_{i_{k_{\max}}} < t_N. \end{aligned} \tag{4.11}$$

Further, the upper CDF envelope $\widehat{F}^{\mathrm{up}} \equiv \mathcal{F}_\rho^{\mathrm{up}}[\widehat{F}]$ *of* $\widehat{F}$ *is given as*

$$\widehat{F}^{\mathrm{up}}(t) =$$

$$\begin{cases} 0 & \text{if } t \in (-\infty, a), \\ z_\ell + (y_\ell - z_\ell)\frac{t_{i_k} - t_\ell}{t_{i_k} - t} & \text{if } t \in [t_\ell, t_{\ell+1}), \ell \in [j_k : j_{k+1}-2], k \in [1:k_{\max}], \\ & \text{if } t \in [t_{j_{k+1}-1}, \tau_{i_k}), \ell = j_{k+1}-1, k \in [1:k_{\max}], \\ z_{j_{k+1}-1} + (z_\ell - z_{j_{k+1}-1})\frac{t_{\ell+1} - \tau_\ell}{t_{\ell+1} - t} & \text{if } t \in [\tau_\ell, \tau_{\ell+1}), \ell \in [i_k : i_{k+1}-2], k \in [1:k_{\max}], \\ & \text{if } t \in [\tau_{i_{k+1}-1}, t_{j_{k+1}}), \ell = i_{k+1}-1, k \in [1:k_{\max}], \\ 1 & \text{if } t \in [\tau_N, \infty), \end{cases}$$

where $z_\ell := \sum_{l=0}^{\ell} c_l$*,* $\ell \in [0:N]$ *and* $c_0 := 0$*.*

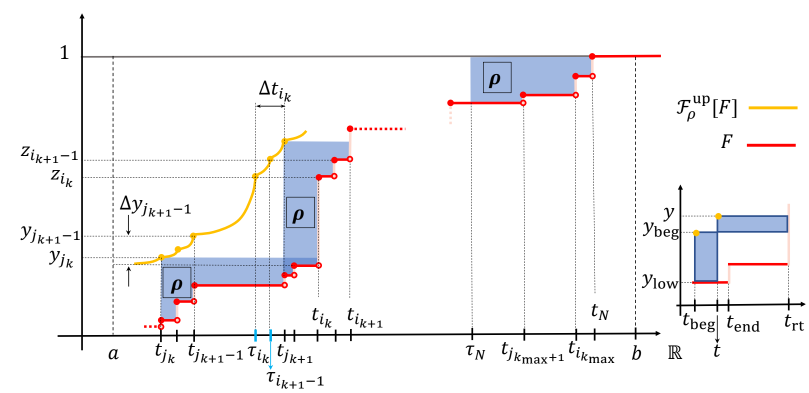

FIG. 2. *Illustration of how the upper CDF envelope* $\mathcal{F}_\rho^{\mathrm{up}}[F]$ *(in yellow) is constructed for a discrete distribution with a finite number of atoms.*

Proposition 4.7 is illustrated in Figure 2. To construct lower CDF envelopes, we introduce the reflection $\mathcal{F}^{\mathrm{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[F]$ of a function F around the point $(\frac{a+b}{2}, \frac{1}{2})$, i.e., $\mathcal{F}^{\mathrm{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[F](t) := 1 - F(a+b-t)$, $t \in \mathbb{R}$. We also define the right-continuous version $\mathrm{rc}[G]$ of an increasing function G by $\mathrm{rc}[G](t) := \lim_{s \searrow t} G(s)$, that satisfies $\int_a^t G(s)ds = \int_a^t \mathrm{rc}[G](s)ds$. Combining this with the fact that $G^{-1} \equiv (\mathrm{rc}[G])^{-1}$ when

G is increasing, we deduce from Lemma 4.4 that the upper and lower CDF envelopes of a CDF F are well defined and, in fact, are the same with those of any increasing function G agreeing with F everywhere except from its points of discontinuity, i.e., with $\mathrm{rc}[G] = F$. The next result explicitly constructs lower CDF envelopes by reflecting the upper CDF envelopes of reflected CDFs. Its proof is given in Appendix A.

LEMMA 4.8 (Lower CDF envelope via reflection). *Let $F \in \mathcal{CD}([a,b])$ and $\rho > 0$ with $\rho \le \int_a^b F(t)dt$. Then, the lower CDF envelope of F satisfies*

$$\mathcal{F}_\rho^{\mathrm{low}}[F] = \mathcal{F}^{\mathrm{refl}}_{(\frac{a+b}{2},\frac{1}{2})}\big[\mathcal{F}_\rho^{\mathrm{up}}\big[\mathcal{F}^{\mathrm{refl}}_{(\frac{a+b}{2},\frac{1}{2})}[F]\big]\big].$$

Using Lemma 4.8, one can leverage Proposition 4.7 to obtain the lower CDF envelope $\mathcal{F}_\rho^{\mathrm{low}}[F]$ of a discrete distribution $F \in \mathcal{CD}([a,b])$ with mass $c_i > 0$ at a finite number of points $a =: t_0 \le t_1 < \cdots < t_N \le b$ for any $\rho > 0$ with $\rho \le \int_a^b F(t)dt$.

5. CDFs and 1-Wasserstein Distance propagation via the Method of Distributions.

Here we develop the necessary tools to propagate in space and time the input ambiguity sets constructed in section 4. To obtain an evolution equation for the single-point cumulative distribution function $F_{u(\mathbf{x},t)}$ of $u(\mathbf{x},t)$, we introduce the random variable $\Pi(U,\mathbf{x},t) = \mathcal{H}(U - u(\mathbf{x},t))$, parameterized by $U \in \mathbb{R}$. The ensemble mean of Π over all possible realizations of u at a point $(\mathbf{x},t)$ is the single-point CDF

$$\langle \Pi(U,\mathbf{x},t)\rangle = F_{u(\mathbf{x},t)}.$$

The dependence of $F_{u(\mathbf{x},t)}$ on $U \in \mathbb{R}$ is implied. We henceforth use the notation $\widetilde{\Omega} \equiv \mathbb{R}\times\Omega$, $\widetilde{\Gamma} \equiv \mathbb{R}\times\Gamma$, and $\widetilde{\mathbf{x}} \equiv (U,\mathbf{x}) \in \mathbb{R}\times\mathbb{R}^d$. Using the Method of Distributions [30], one can obtain the next result, whose derivation is summarized in Appendix B.

THEOREM 5.1 (Physics-driven CDF equation [6]). *Let $F_{u_0(\mathbf{x})}$, $\mathbf{x} \in \Omega$, and $F_{u_b(\mathbf{x},t)}$, $(\mathbf{x},t) \in \Gamma\times\mathbb{R}_{\ge 0}$, be the CDFs of the initial and boundary conditions in (3.2). Under Assumptions 3.1 and 3.2, the CDF $F_{u(\mathbf{x},t)}$ as a solution of (3.1) obeys*

$$\frac{\partial F_{u(\mathbf{x},t)}}{\partial t} + \mathbf{\Lambda}\cdot\widetilde{\nabla} F_{u(\mathbf{x},t)} = 0, \quad \widetilde{\mathbf{x}} \in \widetilde{\Omega}, t \in (0,T) \tag{5.1}$$

with $\mathbf{\Lambda} = (\dot{\boldsymbol{q}}(U;\boldsymbol{\theta}_q), r(U;\boldsymbol{\theta}_r))$ and $\widetilde{\nabla} = (\nabla, \partial/\partial U)$, with $\dot{\mathbf{q}} = \partial\mathbf{q}/\partial U$, and subject to initial and boundary conditions $F_{u_0(\mathbf{x})}$ and $F_{u_b(\mathbf{x},t)}$, respectively.

The CDF evolution is governed by the linear hyperbolic PDE (5.1), which is specific for the physical model (3.1). The next result exploits the properties of (5.1) to obtain an upper bound across space and time on the difference between two CDFs.

COROLLARY 5.1 (Propagation of upper bound on difference between CDFs). *Consider a pair of input CDFs $F^1_{u_0(\mathbf{x})}$, $F^2_{u_0(\mathbf{x})}$, $\mathbf{x} \in \Omega$, and $F^1_{u_b(\mathbf{x},t)}$, $F^2_{u_b(\mathbf{x},t)}$, $(\mathbf{x},t) \in \Gamma\times\mathbb{R}_{\ge 0}$ such that*

$$|e_0(\widetilde{\mathbf{x}})| \ge |\varepsilon_0(\widetilde{\mathbf{x}})| = |F^1_{u_0(\mathbf{x})} - F^2_{u_0(\mathbf{x})}|, \quad \forall \widetilde{\mathbf{x}} \in \widetilde{\Omega}$$

$$|e_b(\widetilde{\mathbf{x}},t)| \ge |\varepsilon_b(\widetilde{\mathbf{x}},t)| = |F^1_{u_b(\mathbf{x},t)} - F^2_{u_b(\mathbf{x},t)}|, \quad \forall (\widetilde{\mathbf{x}},t) \in \widetilde{\Gamma}\times\mathbb{R}_{\ge 0}. \tag{5.2}$$

Then, it holds that

$$|e(\widetilde{\mathbf{x}},t)| \ge |F^1_{u(\mathbf{x},t)} - F^2_{u(\mathbf{x},t)}| = |\varepsilon(\widetilde{\mathbf{x}},t)|, \quad \forall (\widetilde{\mathbf{x}},t) \in \widetilde{\Omega}\times[0,T), \tag{5.3}$$

where $F^1_{u(\mathbf{x},t)}$ *and* $F^2_{u(\mathbf{x},t)}$ *are the solutions of* (5.1) *for the corresponding initial and boundary data, with* $e(\widetilde{\mathbf{x}},t)$ *obeying*

$$\begin{aligned}
&\frac{\partial |e|}{\partial t} + \boldsymbol{\Lambda}\cdot\widetilde{\nabla}|e| = 0, && \widetilde{\mathbf{x}}\in\widetilde{\Omega}, t>0\\
&|e(\widetilde{\mathbf{x}},t=0)| = |e_0(\widetilde{\mathbf{x}})|, && \widetilde{\mathbf{x}}\in\widetilde{\Omega}\\
&|e(\widetilde{\mathbf{x}},t)| = |e_b(\widetilde{\mathbf{x}},t)|, && \widetilde{\mathbf{x}}\in\widetilde{\Gamma}, t>0
\end{aligned}\tag{5.4}$$

Proof. Exploiting the linearity of (5.1), one can write an equation for the difference $\varepsilon(\widetilde{\mathbf{x}},t) = F^1_{u(\mathbf{x},t)} - F^2_{u(\mathbf{x},t)}$,

$$\begin{aligned}
&\frac{\partial \varepsilon}{\partial t} + \boldsymbol{\Lambda}\cdot\widetilde{\nabla}\varepsilon = 0, && \widetilde{\mathbf{x}}\in\widetilde{\Omega}, t\in(0,T)\\
&\varepsilon(\widetilde{\mathbf{x}},t=0) = \varepsilon_0(\widetilde{\mathbf{x}}), && \widetilde{\mathbf{x}}\in\widetilde{\Omega}\\
&\varepsilon(\widetilde{\mathbf{x}},t) = \varepsilon_b(\widetilde{\mathbf{x}},t), && \widetilde{\mathbf{x}}\in\widetilde{\Gamma}, t>0
\end{aligned}\tag{5.5}$$

where $\varepsilon_0(\widetilde{\mathbf{x}}) = F^1_{u_0(\mathbf{x})} - F^2_{u_0(\mathbf{x})}$ and $\varepsilon_b(\widetilde{\mathbf{x}},t) = F^1_{u_b(\mathbf{x},t)} - F^2_{u_b(\mathbf{x},t)}$ are the initial and boundary differences, resp. (5.5) can be expressed as the ODE system $\frac{\mathrm{d}\varepsilon}{\mathrm{d}s} = 0$, $\frac{\mathrm{d}\widetilde{\mathbf{x}}}{\mathrm{d}s} = \boldsymbol{\Lambda}$, $s>0$ with initial/boundary conditions assigned at the intersection between the characteristic lines and the noncharacteristic surface delimiting the space-time domain. Pointwise input differences $\varepsilon_0(\widetilde{\mathbf{x}})$ and $\varepsilon_b(\widetilde{\mathbf{x}},t)$ are conserved and propagate rigidly along deterministic characteristic lines, hence retaining the sign set by the input. Since the system dynamics does not change the sign of ε along the deterministic characteristic lines, ε and $|\varepsilon|$ obey the same dynamics

$$\begin{aligned}
&\frac{\partial |\varepsilon|}{\partial t} + \boldsymbol{\Lambda}\cdot\widetilde{\nabla}|\varepsilon| = 0, && \widetilde{\mathbf{x}}\in\widetilde{\Omega}, t\in(0,T)\\
&|\varepsilon(\widetilde{\mathbf{x}},t=0)| = |\varepsilon_0(\widetilde{\mathbf{x}})|, && \widetilde{\mathbf{x}}\in\widetilde{\Omega}\\
&|\varepsilon(\widetilde{\mathbf{x}},t)| = |\varepsilon_b(\widetilde{\mathbf{x}},t)|, && \widetilde{\mathbf{x}}\in\widetilde{\Gamma}\times\mathbb{R}_{\geq 0}.
\end{aligned}\tag{5.6}$$

For $e_0(\widetilde{\mathbf{x}},t)$ and $e_b(\widetilde{\mathbf{x}},t)$ as in (5.2), and $|e(\widetilde{\mathbf{x}},t)|$ obeying (5.4), (5.6) implies (5.3). ◻

The next result shows that propagation in space and time of CDFs is monotonic.

COROLLARY 5.2 (Propagation of CDFs is monotonic). *Consider a pair of input CDFs* $F^1_{u_0(\mathbf{x})}$, $F^2_{u_0(\mathbf{x})}$, $\mathbf{x}\in\Omega$, *and* $F^1_{u_b(\mathbf{x},t)}$, $F^2_{u_b(\mathbf{x},t)}$, $(\mathbf{x},t)\in\Gamma\times\mathbb{R}_{\geq 0}$ *such that*

$$\begin{aligned}
&F^1_{u_0(\mathbf{x})} \geq F^2_{u_0(\mathbf{x})} && \forall\widetilde{\mathbf{x}}\in\widetilde{\Omega}\\
&F^1_{u_b(\mathbf{x},t)} \geq F^2_{u_b(\mathbf{x},t)} && \forall(\widetilde{\mathbf{x}},t)\in\widetilde{\Gamma}\times\mathbb{R}_{\geq 0}
\end{aligned}\tag{5.7}$$

Furthermore, we assume $F^1_{u(\mathbf{x},t)}$ *and* $F^2_{u(\mathbf{x},t)}$ *to be solutions of* (5.1) *with* $F^1_{u_0(\mathbf{x})}, F^1_{u_b(\mathbf{x},t)}$ *and* $F^2_{u_0(\mathbf{x})}, F^2_{u_b(\mathbf{x},t)}$ *initial and boundary conditions, respectively. Then, it holds that*

$$F^1_{u(\mathbf{x},t)} \geq F^2_{u(\mathbf{x},t)}, \forall\widetilde{\mathbf{x}}\in\widetilde{\Omega}\times[0,T).\tag{5.8}$$

Proof. The discrepancy $\varepsilon(\widetilde{\mathbf{x}},t) = F^1_{u(\mathbf{x},t)} - F^2_{u(\mathbf{x},t)}$ obeys (5.5). Given non-negative initial and boundary conditions, consistently with (5.7), it holds that $\varepsilon(\widetilde{\mathbf{x}},t)\geq 0$ for all $\widetilde{\mathbf{x}}\in\widetilde{\Omega}, t\in(0,T)$, hence (5.8). ◻

The CDF equation (5.1) provides a computational tool for the space-time propagation of the CDFs of the inputs. If the governing equation (3.1) is linear, we show next that one can obtain an evolution equation in the form of a PDE for the 1-Wasserstein distance between each pair of distributions describing the same underlying physical process.

THEOREM 5.2 (Physics-driven 1-Wasserstein discrepancy equation). *Consider a pair of distributions $F^1_{u(\mathbf{x},t)}$ and $F^2_{u(\mathbf{x},t)}$ obeying (5.1), and assume linearity of (3.1). Then, the 1-Wasserstein discrepancy between $F^1_{u(\mathbf{x},t)}$ and $F^2_{u(\mathbf{x},t)}$ defined by (2.1), $\omega_1(\mathbf{x},t) = \int_{\mathbb{R}} |F^1_{u(\mathbf{x},t)} - F^2_{u(\mathbf{x},t)}| dU$, obeys*

$$\begin{aligned} &\frac{\partial \omega_1}{\partial t} + \dot{\boldsymbol{q}} \cdot \nabla \omega_1 - \dot{r}\, \omega_1 = 0, && \mathbf{x} \in \Omega, t > 0 \\ &\omega_1(\mathbf{x}, t=0) = \omega_0(\mathbf{x}), && \mathbf{x} \in \Omega \\ &\omega_1(\mathbf{x},t) = \omega_b(\mathbf{x},t), && \mathbf{x} \in \Gamma, t > 0, \end{aligned} \tag{5.9}$$

with $\omega_0(\mathbf{x}) = \int_{\mathbb{R}} |F^1_{u_0(\mathbf{x})} - F^2_{u_0(\mathbf{x})}| dU$ and $\omega_b = \int_{\mathbb{R}} |F^1_{u_b(\mathbf{x},t)} - F^2_{u_b(\mathbf{x},t)}| dU$ the input discrepancies.

Proof. (5.9) follows from (5.4) by integration along $U \in \mathbb{R}$ assuming $F^1_{u(\mathbf{x},t)}(U = \pm\infty) = F^2_{u(\mathbf{x},t)}(U = \pm\infty)$, for all $\mathbf{x} \in \Omega, t > 0$, accounting for the linearity of $\mathbf{q}(U)$ and $r(U)$. ∎

Corollary 5.1 and the following Corollary 5.3 take advantage of the linearity and hyperbolic structure of (5.4) and (5.9), respectively, and identify a dynamic bound for the evolution of the pointwise CDF absolute difference and their 1-Wasserstein distance, respectively, once the corresponding discrepancies are set at the initial time and along the boundaries.

COROLLARY 5.3 (Physics-driven 1-Wasserstein dynamic bound). *Consider the input CDF pairs $F^1_{u_0(\mathbf{x})}$, $F^2_{u_0(\mathbf{x})}$, $\mathbf{x} \in \Omega$, and $F^1_{u_b(\mathbf{x},t)}$, $F^2_{u_b(\mathbf{x},t)}$, $(\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}$. Let $w(\mathbf{x},t)$ be the solution of (5.9) with initial and boundary conditions satisfying*

$$\begin{aligned} &w_0(\mathbf{x}) \geq \omega_0(\mathbf{x}) = W_1\big(F^1_{u_0(\mathbf{x})}, F^2_{u_0(\mathbf{x})}\big) \quad \forall \mathbf{x} \in \Omega \\ &w_b(\mathbf{x},t) \geq \omega_b(\mathbf{x},t) = W_1\big(F^1_{u_b(\mathbf{x},t)}, F^2_{u_b(\mathbf{x},t)}\big) \quad \forall (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}. \end{aligned} \tag{5.10}$$

Then, it holds that

$$\omega_1(\mathbf{x},t) = W_1\big(F^1_{u(\mathbf{x},t)}, F^2_{u(\mathbf{x},t)}\big) \leq w(\mathbf{x},t) \quad \forall (\mathbf{x},t) \in \Omega \times \mathbb{R}_{\geq 0}, \tag{5.11}$$

where $F^1_{u(\mathbf{x},t)}$ and $F^2_{u(\mathbf{x},t)}$ are the solutions of (5.1) for the corresponding initial and boundary distributions.

Proof. (5.11) follows from condition (5.10) and having $w(\mathbf{x},t)$ and $\omega_1(\mathbf{x},t)$ that fulfill (5.9) with conditions w_0, w_b and ω_0, ω_b, respectively. ∎

6. Ambiguity set propagation under finite-sample guarantees. Here we combine the results from sections 4 and 5 to build pointwise ambiguity sets for the distribution of $u(\mathbf{x},t)$ over the whole spatio-temporal domain. We first consider the general PDE model (3.1) and study how the input ambiguity bands of Corollary 4.5 propagate in space and time using the CDF equation (5.1).

THEOREM 6.1 (Ambiguity band evolution via the CDF dynamics). *Assume that N pairs of input samples are collected according to Assumption 3.4. Consider a confidence $1-\beta$ and the CDFs*

$$F^{\text{low}}_{u_0(\mathbf{x})} := \mathcal{F}^{\text{low}}_{\rho_0(\mathbf{x}),[\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})]}\big[\widehat{F}^N_{u_0(\mathbf{x})}\big], \quad \mathbf{x} \in \Omega$$

$$F^{\mathrm{low}}_{u_b(\mathbf{x},t)} := \mathcal{F}^{\mathrm{low}}_{\rho_b(\mathbf{x},t),[\alpha_b(\mathbf{x},t),\gamma_b(\mathbf{x},t)]}\big[\widehat{F}^N_{u_b(\mathbf{x},t)}\big], \quad (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}$$
$$F^{\mathrm{up}}_{u_0(\mathbf{x})} := \mathcal{F}^{\mathrm{up}}_{\rho_0(\mathbf{x}),[\alpha_0(\mathbf{x}),\gamma_0(\mathbf{x})]}\big[\widehat{F}^N_{u_0(\mathbf{x})}\big], \quad \mathbf{x} \in \Omega$$
$$F^{\mathrm{up}}_{u_b(\mathbf{x},t)} := \mathcal{F}^{\mathrm{up}}_{\rho_b(\mathbf{x},t),[\alpha_b(\mathbf{x},t),\gamma_b(\mathbf{x},t)]}\big[\widehat{F}^N_{u_b(\mathbf{x},t)}\big], \quad (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0},$$

with $[\alpha_0(\mathbf{x}), \gamma_0(\mathbf{x})]$, $[\alpha_b(\mathbf{x},t), \gamma_b(\mathbf{x},t)]$ *and* $\rho_0(\mathbf{x})$, $\rho_b(\mathbf{x},t)$ *as given in* (4.3a), (4.3b) *and* (4.8a), (4.8b), *respectively. Let* $F^{\mathrm{low}}_{u(\mathbf{x},t)}$ *and* $F^{\mathrm{up}}_{u(\mathbf{x},t)}$ *be the solutions of* (5.1) *with the corresponding input CDFs above and define the ambiguity sets*

$$\mathcal{P}^{\mathrm{Env}}_{\mathbf{x},t} := \big\{F \in \mathcal{CD}(\mathbb{R}) \,|\, F^{\mathrm{low}}_{u(\mathbf{x},t)} \leq F \leq F^{\mathrm{up}}_{u(\mathbf{x},t)} \;\forall U \in \mathbb{R}\big\}, \quad \mathbf{x} \in \Omega, t \in [0,T).$$

Then $\mathbb{P}(F^{\mathrm{true}}_{u(\mathbf{x},t)} \in \mathcal{P}^{\mathrm{Env}}_{\mathbf{x},t} \;\forall (\mathbf{x},t) \in \Omega \times [0,T)) \geq 1-\beta$.

Proof. Let

$$A := \{(\mathbf{a}^1,\dots,\mathbf{a}^N) \in \mathbb{R}^{Nn} \,|\, F^{\mathrm{true}}_{u_0(\mathbf{x})} \in \mathcal{P}^{0,\mathrm{Env}}_{\mathbf{x}}(\mathbf{a}^1,\dots,\mathbf{a}^N) \;\forall \mathbf{x} \in \Omega$$
$$\wedge\; F^{\mathrm{true}}_{u_b(\mathbf{x},t)} \in \mathcal{P}^{b,\mathrm{Env}}_{\mathbf{x},t}(\mathbf{a}^1,\dots,\mathbf{a}^N) \;\forall (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}\},$$

with $\mathcal{P}^{0,\mathrm{Env}}_{\mathbf{x}}$ and $\mathcal{P}^{b,\mathrm{Env}}_{\mathbf{x},t}$ as given in Corollary 4.5, where we emphasize their dependence on the parameter realizations. Then, we have from (4.9) that

$$\mathbb{P}((\mathbf{a}^1,\dots,\mathbf{a}^N) \in A) \geq 1-\beta. \tag{6.1}$$

Next, let $(\mathbf{a}^1,\dots,\mathbf{a}^N) \in A$ and $\widehat{F}^N_{u_0(\mathbf{x})} \equiv \widehat{F}^N_{u_0(\mathbf{x})}(\mathbf{a}^1,\dots,\mathbf{a}^N)$, $\mathbf{x} \in \Omega$, $\widehat{F}^N_{u_b(\mathbf{x},t)} \equiv \widehat{F}^N_{u_b(\mathbf{x},t)}(\mathbf{a}^1,\dots,\mathbf{a}^N)$, $(\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}$ be the associated empirical input CDFs. These generate the corresponding lower CDF envelopes $F^{\mathrm{low}}_{u_0(\mathbf{x})} \equiv F^{\mathrm{low}}_{u_0(\mathbf{x})}(\mathbf{a}^1,\dots,\mathbf{a}^N)$ and $F^{\mathrm{low}}_{u_b(\mathbf{x},t)} \equiv F^{\mathrm{low}}_{u_b(\mathbf{x},t)}(\mathbf{a}^1,\dots,\mathbf{a}^N)$ given in the statement, and we deduce from the definitions of A and the ambiguity sets $\mathcal{P}^{0,\mathrm{Env}}_{\mathbf{x}}$, $\mathcal{P}^{b,\mathrm{Env}}_{\mathbf{x},t}$ that $F^{\mathrm{true}}_{u_0(\mathbf{x})}(U) \geq F^{\mathrm{low}}_{u_0(\mathbf{x})}(U)$ for all $U \in \mathbb{R}, \mathbf{x} \in \Omega$ and $F^{\mathrm{true}}_{u_b(\mathbf{x},t)}(U) \geq F^{\mathrm{low}}_{u_b(\mathbf{x},t)}(U)$ for all $U \in \mathbb{R}, (\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}$. Thus, we obtain from Corollary 5.2 applied with $F^1_u \equiv F^{\mathrm{true}}_u$ and $F^2_u \equiv F^{\mathrm{low}}_u$ that

$$F^{\mathrm{true}}_{u(\mathbf{x},t)}(U) \geq F^{\mathrm{low}}_{u(\mathbf{x},t)}(U) \quad \forall U \in \mathbb{R}, (\mathbf{x},t) \in \Omega \times [0,T).$$

Analogously, we get that $F^{\mathrm{true}}_{u(\mathbf{x},t)}(U) \leq F^{\mathrm{up}}_{u(\mathbf{x},t)}(U)$ for all $U \in \mathbb{R}, (\mathbf{x},t) \in \Omega \times [0,T)$, and we deduce from the definition of the ambiguity sets $\mathcal{P}^{\mathrm{Env}}_{\mathbf{x},t}$ in the statement that

$$F^{\mathrm{true}}_{u(\mathbf{x},t)} \in \mathcal{P}^{\mathrm{Env}}_{\mathbf{x},t}(\mathbf{a}^1,\dots,\mathbf{a}^N) \quad \forall U \in \mathbb{R}, (\mathbf{x},t) \in \Omega \times [0,T).$$

The result now follows from (6.1). ∎

Under linearity of the dynamics, we can exploit Corollary 5.3 to propagate the tighter Wasserstein input ambiguity balls of Proposition 4.3.

Theorem 6.2 (Ambiguity set evolution for linear dynamics). *Assume that PDE* (3.1) *is linear and* N *pairs of input samples are collected according to Assumption 3.4. Consider a confidence level* $1-\beta$ *and let* $w(\mathbf{x},t)$ *be the solution of* (5.9) *with* $w_0(\mathbf{x}) = L_0(\mathbf{x})\epsilon_N(\beta,\rho_{\mathbf{a}})$, $\mathbf{x} \in \Omega$ *and* $w_b(\mathbf{x},t) = L_b(\mathbf{x},t)\epsilon_N(\beta,\rho_{\mathbf{a}})$, $(\mathbf{x},t) \in \Gamma \times \mathbb{R}_{\geq 0}$, *and* $L_0(\mathbf{x})$, $L_b(\mathbf{x},t)$, $\rho_{\mathbf{a}}$, *and* $\epsilon_N(\beta,\rho_{\mathbf{a}})$ *given by* (3.3a), (3.3b), (4.2), *and* (4.1). *Let* $\widehat{F}^N_{u(\mathbf{x},t)}$ *be the solution of* (5.1) *with the empirical input CDFs* $\widehat{F}^N_{u_0(\mathbf{x})}$ *and* $\widehat{F}^N_{u_b(\mathbf{x},t)}$ *as given in section 4 and define the ambiguity sets*

$$\mathcal{P}_{\mathbf{x},t} := \big\{F \in \mathcal{CD}(\mathbb{R}) \,|\, W_1(\widehat{F}^N_{u(\mathbf{x},t)}, F) \leq w(\mathbf{x},t)\big\}, \quad \mathbf{x} \in \Omega, t \in \mathbb{R}_{\geq 0}.$$

Then $\mathbb{P}(F^{\mathrm{true}}_{u(\mathbf{x},t)} \in \mathcal{P}_{\mathbf{x},t} \;\forall (\mathbf{x},t) \in \Omega \times \mathbb{R}_{\geq 0}) \geq 1-\beta$.

Proof. Let $A := \{(\mathbf{a}^1, \dots, \mathbf{a}^N) \in \mathbb{R}^{Nn} \,|\, F^{\text{true}}_{u_0(\mathbf{x})} \in \mathcal{P}^0_{\mathbf{x}}(\mathbf{a}^1, \dots, \mathbf{a}^N) \;\forall \mathbf{x} \in \Omega \;\wedge\; F^{\text{true}}_{u_b(\mathbf{x},t)} \in \mathcal{P}^b_{\mathbf{x},t}(\mathbf{a}^1, \dots, \mathbf{a}^N) \;\forall (\mathbf{x}, t) \in \Gamma \times \mathbb{R}_{\geq 0}\}$, with $\mathcal{P}^0_{\mathbf{x}}$ and $\mathcal{P}^b_{\mathbf{x},t}$ as given in Proposition 4.3. Then, we have from (4.4) that (6.1) holds. Next, let $(\mathbf{a}^1, \dots, \mathbf{a}^N) \in A$ and $\widehat{F}^N_{u_0(\mathbf{x})} \equiv \widehat{F}^N_{u_0(\mathbf{x})}(\mathbf{a}^1, \dots, \mathbf{a}^N)$, $\mathbf{x} \in \Omega$, $\widehat{F}^N_{u_b(\mathbf{x},t)} \equiv \widehat{F}^N_{u_b(\mathbf{x},t)}(\mathbf{a}^1, \dots, \mathbf{a}^N)$, $(\mathbf{x}, t) \in \Gamma \times \mathbb{R}_{\geq 0}$ be the associated input CDFs. From the definition of $\mathcal{P}^0_{\mathbf{x}}$, $\mathcal{P}^b_{\mathbf{x},t}$ and w_0, w_b we get

$$W_1\big(\widehat{F}^N_{u_0(\mathbf{x})}, F^{\text{true}}_{u_0(\mathbf{x})}\big) \leq w_0(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega$$
$$W_1\big(\widehat{F}^N_{u_b(\mathbf{x},t)}, F^{\text{true}}_{u_b(\mathbf{x},t)}\big) \leq w_b(\mathbf{x}, t) \quad \forall (\mathbf{x}, t) \in \Gamma \times \mathbb{R}_{\geq 0}.$$

Thus, applying Corollary 5.3 with $F^1 \equiv \widehat{F}^N_u$ and $F^2 \equiv F^{\text{true}}_u$, $W_1\big(\widehat{F}^N_{u(\mathbf{x},t)}, F^{\text{true}}_{u(\mathbf{x},t)}\big) \leq w(\mathbf{x}, t)$, for all $(\mathbf{x}, t) \in \Omega \times \mathbb{R}_{\geq 0}$, and it follows from the definition of $\mathcal{P}_{\mathbf{x},t}$ that

$$F^{\text{true}}_{u(\mathbf{x},t)} \in \mathcal{P}_{\mathbf{x},t}(\mathbf{a}^1, \dots, \mathbf{a}^N) \quad \forall (\mathbf{x}, t) \in \Omega \times \mathbb{R}_{\geq 0}.$$

Combining this with (6.1) for A as given in this proof yields the result. ∎

7. Numerical example. In this section, we illustrate the use of the ambiguity propagation tools developed above in a numerical example. We consider a one-dimensional version of (3.1) with linear

$$q(u) = u, \quad \text{and} \quad r(u; \theta_r) = \theta_r u, \quad \theta_r \in \mathbb{R}, \tag{7.1}$$

defined in $\Omega = \mathbb{R}_{\geq 0}$ and subject to the following initial and boundary conditions

$$u(x, 0) = u_0 = a_1 + a_2, \quad x \geq 0$$
$$u(0, t) = u_b(t) = a_1 + a_2\,(1 + a_3 \sin(2\pi t))\,, \quad t \geq 0 \tag{7.2}$$

(note that this fulfills the most restrictive conditions of Theorem 5.2). Because of (7.2), in the following we drop the dependence of the input and boundary conditions from x. Randomness is introduced by the finite set of $(n = 3)$ i.i.d. uncertain parameters $\mathbf{a} = (a_1, a_2, a_3)$, which vary in $[0, 1]^n$; according to (4.2), $\rho_{\mathbf{a}} = 1/2$. We choose a uniform distribution to be the data-generating distribution for $\mathbf{a}$. Both u_0 and $u_b(t)$ are random non-negative variables which are defined on the compact supports $[0, 2]$ and $[0, 2 + \max(0, \sin(2\pi t))]$, respectively.

7.1. Shape and size of the input ambiguity sets. We consider data-driven 1-Wasserstein ambiguity sets for the parameters $\mathbf{a}$, which are constructed according to Lemma 4.1 using $p = 1$ and $n = 3$. We choose the radius $\epsilon_N(\beta, \rho_{\mathbf{a}})$ in (4.1) for a given sample size N and a fixed β. Threshold radii for different size of the sample N and identical confidence level $1-\beta$ can be constructed in relative terms, as exemplified in [5]. By adjusting $\epsilon_N(\beta, \rho_{\mathbf{a}})$, the decision-maker determines the level of conservativeness of the ambiguity set, and the distributional robustness as a consequence. The ambiguity sets for the parameters are scaled into pointwise ambiguity sets for the inputs following Proposition 4.3, via the definition of the Lipschitz constants

$$\rho_0 = L_0 \epsilon_N(\beta, \rho_{\mathbf{a}}), \qquad \text{with } L_0 := \sqrt{2},$$
$$\rho_b(t) = L_b(t) \epsilon_N(\beta, \rho_{\mathbf{a}}), \quad \text{with } L_b(t) := \sqrt{2 + 2\sin^2(2\pi t) + 2\max(0, \sin(2\pi t))}. \tag{7.3}$$

Second, we construct conservative ambiguity envelopes for the initial and the boundary conditions characterized by a 1-Wasserstein discrepancy larger than ρ_0 and $\rho_b(t)$, respectively, according to Proposition 4.7. These upper and lower envelopes define an ambiguity band which enjoys the same performance guarantees as the previously defined 1-Wasserstein ambiguity sets. We denote with $\rho_0^{\mathrm{Env}} \geq \rho_0$ and $\rho_b^{\mathrm{Env}}(t) \geq \rho_b(t)$ the 1-Wasserstein discrepancy between the upper and lower distributions defining the initial and boundary ambiguity bands, respectively.

For both inputs, the maximum pointwise Wasserstein distance $\rho_{0,\max}$ and $\rho_{b,\max}(t)$ corresponds to the local size of the support. 1-Wasserstein discrepancies larger than the maximum value denote uniformative ambiguity sets. For the chosen scenario, $\rho_{0,\max} = 2$ and $\rho_{b,\max}(t) = 2 + \max(0, \sin(2\pi t))$ for the initial and the boundary values, respectively. A comparison of $\rho_b(t)$, $\rho_b^{\mathrm{Env}}(t)$ and $\rho_{b,\max}(t)$ is presented in Figure 3 for different sample sizes N and identical confidence level $1-\beta$. The corresponding values for the initial condition can be read in the same figure at $t = 0$ because of the imposed continuity between initial and boundary conditions at $t = 0$. Regardless of the chosen shape of the ambiguity set, larger N determines smaller ambiguity sets characterized by smaller 1-Wasserstein discrepancies. By construction, 1-Wasserstein ambiguity sets defined through (7.3) are sharper than the corresponding ambiguity bands drawn geometrically via Proposition 4.7 at all times. The temporal behavior of $\rho_b(t)$ is determined by the Lipschitz scaling function $L_b(t)$ in (7.3); in this case it is periodic and bounded. Figures 4 and 5 show the corresponding ambiguity bands

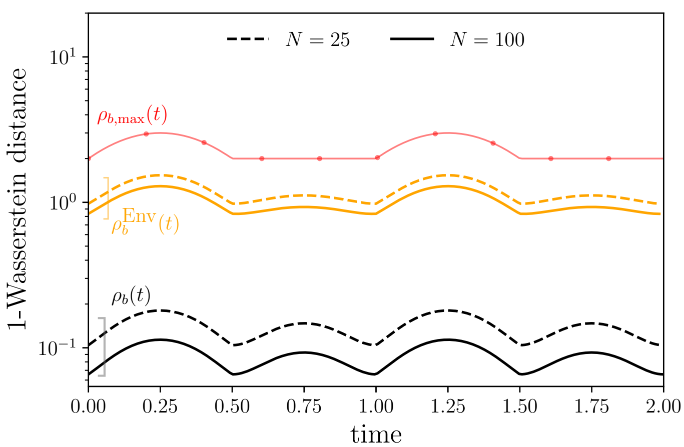

Fig. 3. *Characteristic 1-Wasserstein distances for the pointwise ambiguity sets for $u_b(0,t)$. Black lines correspond to the $\rho_b(t)$ bounds set in Corollary 4.5 and used to define 1-Wasserstein ambiguity sets. Yellow lines indicate $\rho_b^{Env}(t)$, the sample-dependent 1-Wasserstein discrepancy between envelopes defined via the Proposition 4.7 procedure. The line pattern indicates the size of the data sample N, as listed in the legend. The maximum theoretical 1-Wasserstein discrepancy for $u_b(0,t)$, $\rho_{b,\max}(t)$, is also drawn (red circles).*

for u_0 and $u_b(t)$ at a given time t, respectively, for the same values of sample size N and identical confidence level $1-\beta$. Both upper and lower envelopes are data-driven, i.e., they depend on the empirical distribution of a specific sample. We also show the 1-Wasserstein discrepancy between the upper and lower envelopes.

7.2. Propagation of the ambiguity set. Pointwise 1-Wasserstein distances for the inputs can be propagated in space and in time to describe the behavior of the ambiguous distributions using (5.9), under the assumption of linear dynamics. Solving (5.9) yields a quantitative measure of the stretch/shrink of the ambiguity ball in each space-time location. True (unknown) distributions as well as their empirical approximations describing the given physical dynamics evolve according to (5.1); the

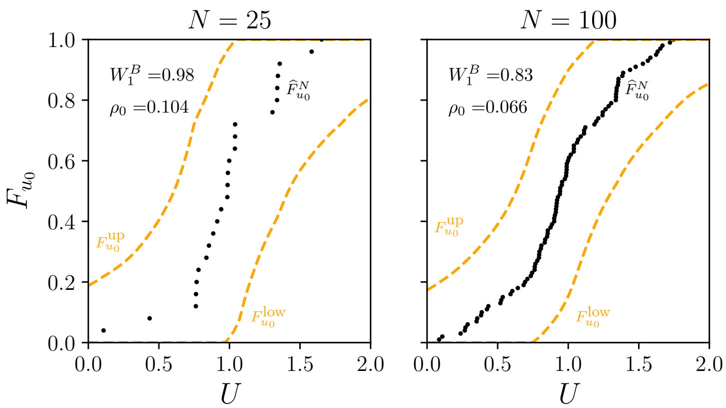

FIG. 4. *Ambiguity band for the distributions of u_0 for different sample size N and identical confidence level $1-\beta$. We use $\theta_r = -1$. Scatter points represent the empirical distribution $\widehat{F}_{u_0}^N$. Dashed yellow lines represent the conservative envelopes (with respect to a minimum 1-Wasserstein distance ρ_0) constructed according to Proposition 4.7. The 1-Wasserstein discrepancies for the ambiguity band - computed between the upper and the lower envelope - are reported in the corresponding panels, also indicating ρ_0.*

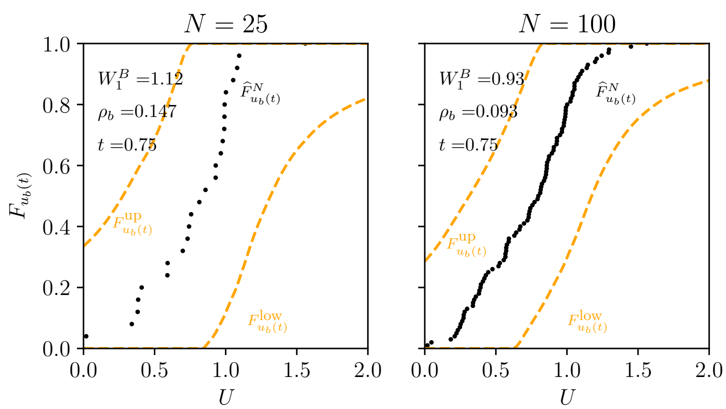

FIG. 5. *Ambiguity band for the distributions of $u_b(t)$ at $t = 0.75$ for different sample size N and identical confidence level $1-\beta$. We use $\theta_r = -1$. Scatter points represent the empirical distribution $\widehat{F}_{u_b(t)}^N$. Dashed yellow lines represent the conservative envelopes (with respect to a minimum 1-Wasserstein distance $\rho_b(t)$) constructed according to Proposition 4.7. The 1-Wasserstein discrepancies for the ambiguity band are reported in the corresponding panels, also indicating $\rho_b(t)$.*

latter provide an anchor for the pointwise ambiguity balls in $(\mathbf{x}, t)$. In Figure 6 we present the solution of (5.9), $w_1(x,t)$, solved using ρ_0 and $\rho_b(t)$ as defined in (7.3) as initial and boundary conditions, respectively. The ambiguity ball shrinks with respect to the input conditions as an effect of a depletion dynamics imposed by (3.1) with the given choice of $\theta_r = -1$. As expected, the smaller the sample size N, the larger the radius of the ambiguity ball as quantified by $w_1(x,t)$.

The dynamic evolution of ambiguity bands is determined by the evolution of the upper and lower envelopes for the input samples, cf. Proposition 4.7, for given sample size N and confidence level $1-\beta$. The envelopes evolve according to (5.1), thus requiring no linearity assumption for (3.1). As such, ambiguity bands, while possibly

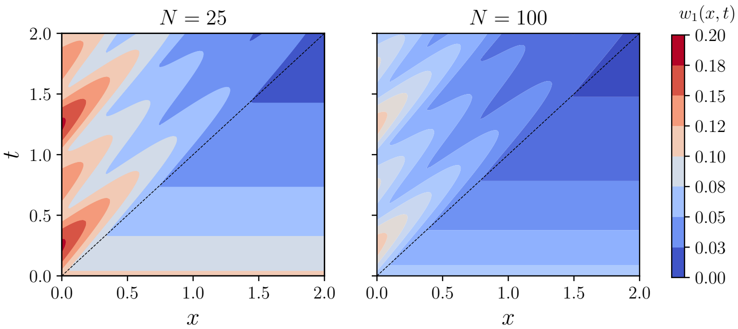

FIG. 6. *$w_1(x,t)$ as a solution of (5.9) with $w_0(x) = \rho_0$ and $w_b(x,t) = \rho_b(t)$ for different sample size N ($N = 25$ in the left panel, and $N = 100$ in the right panel) and identical confidence level $1-\beta$. The dotted line represents the domain partition between regions where information originates from either the initial or the boundary condition. We use $\theta_r = -1$.*

being more conservative than 1-Wasserstein ambiguity sets in terms of size, can be evolved for a wider class of hyperbolic equations. Ambiguity bands are equipped with 1-Wasserstein measures, as the 1-Wasserstein distance between the upper and the lower envelope represents the maximum distance between any pair of distributions within the band, and it is constructed to be always larger or equal than the local radius of the corresponding ambiguity ball. Confidence guarantees established for the inputs (Corollary 4.5) withstand propagation, as demonstrated in Theorem 6.1.

For a given choice of N, we compare the propagation of 1-Wasserstein ambiguity sets with input conditions defined by (7.3) to the data-driven dynamic ambiguity bands constructed via Proposition 4.7 and subject to the input envelopes represented in Figures 4 and 5. The corresponding w_1 maps are shown in Figure 7 (top row). In both cases, the pointwise 1-Wasserstein distance undergoes the same dynamics established by (5.9), but subject to different inputs (represented in Figure 3). In each spatial location, it is possible to track the temporal behavior of the ambiguity set size for both shapes, as shown for two representative locations in Figure 7 (bottom row). The size of both ambiguity sets decreases from the maximum imposed at the initial time for $t < x$, and reflects the temporal signature of the boundary, dampened as an effect of depletion dynamics introduced by (7.1) with $\theta_r = -1$, for $t > x$.

8. Conclusions.

We have provided computational tools in the form of PDEs for the space-time propagation of pointwise ambiguity sets for random variables obeying hyperbolic conservation laws. The initial and boundary conditions of these propagation PDEs depend on the data-driven characterization of the ambiguity sets at the initial time and along the physical boundaries of the spatial domain. We have introduced both 1-Wasserstein ambiguity balls and ambiguity bands, formed through upper and lower CDF envelopes containing all distributions with an assigned 1-Wasserstein distance from their empirical CDFs. The former are propagated by evolving the ambiguity radius according to a dynamic law that can be derived exactly in the case of linear physical models. The latter are propagated by solving the CDF equation for both the upper and the lower CDF envelope defining the ambiguity band. In this second case, both linear and non linear physical processes can be described exactly in CDF terms, provided that no shock develops in the physical model solution. The performance guarantees for the input ambiguity sets of both types are demonstrated to withstand propagation through the physical dynamics. These computational tools allow the modeler to map the physics-driven stretch/ shrink of the ambiguity sets

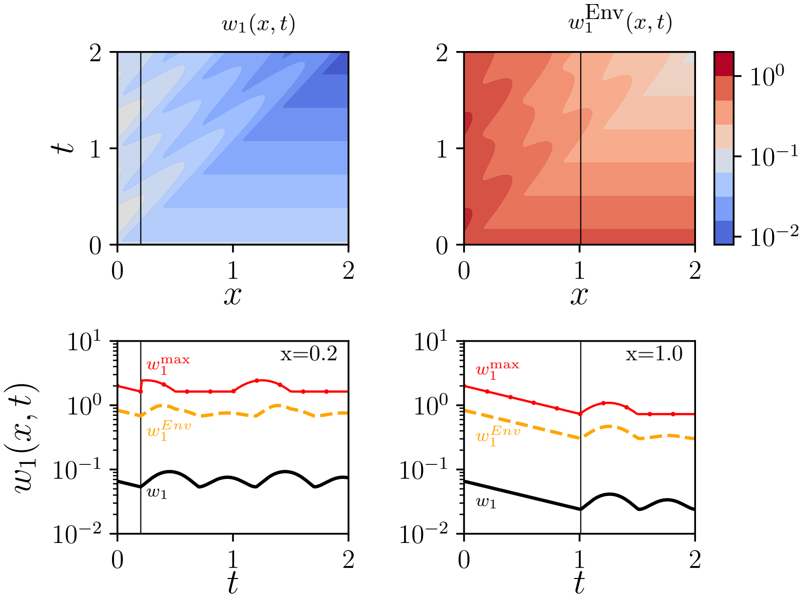

FIG. 7. *Top row: 1-Wasserstein distance maps for the radius of the ambiguity balls $w_1(x,t)$ with input radii* (7.3) *(left), and the ambiguity band $w_1^{\mathrm{Env}}(x,t)$ (right), where $w_1^{\mathrm{Env}}(x,t) = W_1(F^{low}_{u(x,t)}, F^{up}_{u(x,t)})$. Bottom row: 1-Wasserstein distance profiles at given locations $x = \{0.2, 1.\}$. The black solid line reflects the 1-Wasserstein ambiguity radius $w_1(x,t)$, whereas the yellow dashed line represents the 1-Wasserstein distance of the ambiguity band, $w_1^{\mathrm{Env}}(x,t)$. The maximum theoretical 1-Wasserstein discrepancy is also drawn, $w_1^{\max}(x,t)$ (marked red line). The location of the cross-sections is indicated in the top-row contour plots in the corresponding column ($x = 0.2$ and $x = 1$, respectively), whereas the demarcation line $t = x$ is indicated in the bottom panels. Parameters are set to: $N = 100$, $\theta_r = -1$.*

size, enabling dynamic evaluations of distributional robustness. Future research will consider systems of conservation laws with joint one-point CDFs, the characterization of ambiguity sets when shocks are formed under nonlinear dynamics, the assimilation of data collected within the space-time domain, the application of these results in distributionally robust optimization problems, and sharper concentration-of-measure results to reduce conservativeness of the ambiguity sets for small numbers of samples.

Appendix A. Technical proofs from Section 4. We collect here basic properties of generalized CDF inverses used in the following:

(GI1) $F(t) < y \Rightarrow t < F^{-1}(y)$;

(GI2) $F(t_1) \leq y \leq F(t_2) \Rightarrow t_1 \leq F^{-1}(y) \leq t_2$;

(GI3) $t < F^{-1}(y) \Rightarrow F(t) < y$;

(GI4) $F(t) = F(t_1)\ \forall t \in [t_1, t_2) \wedge F(t_1) < y \leq F(t_2) \Rightarrow F^{-1}(y) = t_2$.

Proof of Lemma 4.2. Let $\widehat{T} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^m$ with $\widehat{T}(x,y) = (T(x), T(y))$, consider an optimal coupling π for which the infimum in the definition of the distance $W_p(\mu,\nu)$ is attained, and define $\widehat{\pi} := \widehat{T}_{\#}\pi = \pi \circ \widehat{T}^{-1}$. Then, it follows that $\widehat{\pi}(A \times \mathbb{R}^m) = (\pi \circ \widehat{T}^{-1})(A \times \mathbb{R}^m) = \pi(T^{-1}(A) \times T^{-1}(\mathbb{R}^m)) = \mu(T^{-1}(A)) = T_{\#}\mu(A)$. Hence, $T_{\#}\mu$ is a marginal of $\widehat{\pi}$ and similarly $T_{\#}\nu$, i.e., $\widehat{\pi}$ is a coupling between $T_{\#}\mu$ and $T_{\#}\nu$. Let $\phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ with $\phi(x,y) = \|x-y\|^p$ and $\widehat{T}$ as given above. Then, we obtain from the change of variables formula and the Lipschitz hypothesis that

$$\begin{aligned}(\text{LHS}) &= \int_{\mathbb{R}^m\times\mathbb{R}^m} \|\widehat{x}-\widehat{y}\|^p \widehat{\pi}(d\widehat{x},d\widehat{y}) = \int_{\mathbb{R}^m\times\mathbb{R}^m} \phi(\widehat{x},\widehat{y})\widehat{\pi}(d\widehat{x},d\widehat{y}) \\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} \phi\circ\widehat{T}(x,y)\pi(dx,dy) = \int_{\mathbb{R}^n\times\mathbb{R}^n} \phi(T(x),T(y))\pi(dx,dy) \\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} \|T(x)-T(y)\|^p\pi(dx,dy) \le \int_{\mathbb{R}^n\times\mathbb{R}^n} L^p\|x-y\|^p\pi(dx,dy) = (\text{RHS}).\end{aligned}$$

Thus, we get $W_p^p(T_\#\mu,T_\#\nu) \le (\text{LHS}) \le (\text{RHS}) = L^pW_p^p(\mu,\nu)$, implying the result. ◻

Proof of Lemma 4.4. We show that $\mathcal{F}_\rho^{\text{up}}[F]$ is continuous and increasing, and hence, it is also a CDF, as it takes values in $[0,1]$ (the proof for $\mathcal{F}_\rho^{\text{low}}[F]$ is analogous). Notice first that due to (GI1), i.e., that $F(t) < y \Rightarrow t < F^{-1}(y)$, the mapping $z \mapsto \int_{F(t)}^z (F^{-1}(y)-t)dy$ is strictly increasing for $z\in[F(t),1]$. Combining this fact with continuity of $z\mapsto \int_{F(t)}^z(F^{-1}(y)-t)dy$, we deduce existence of a unique $z\in[F(t),1]$ so that $\mathcal{F}_\rho^{\text{up}}[F](t)=z$ and $\int_{F(t)}^z(F^{-1}(y)-t)dy=\rho$ for all $t\in[a,t_\rho^{\text{up}}[F])$. To show that $\mathcal{F}_\rho^{\text{up}}[F]$ is increasing, let $a\le t_1<t_2<t_\rho^{\text{up}}[F]$ with $\mathcal{F}_\rho^{\text{up}}[F](t_1)=z_1$ and $\mathcal{F}_\rho^{\text{up}}[F](t_2)=z_2$ and assume w.l.o.g. that $F(t_2)<z_1$. Then, we have that

$$\rho = \int_{F(t_1)}^{z_1}(F^{-1}(y)-t_1)dy \ge \int_{F(t_2)}^{z_1}(F^{-1}(y)-t_1)dy > \int_{F(t_2)}^{z_1}(F^{-1}(y)-t_2)dy,$$

where we exploited that F is increasing in the first inequality. Thus, we get that $z_2>z_1$, because also $\int_{F(t_2)}^{z_2}(F^{-1}(y)-t_2)dy=\rho$. To prove continuity, let $t_\nu\to t\in[a,t_\rho^{\text{up}}[F])$ and $\{z_\nu\}_{\nu\in\mathbb{N}}$ with $\mathcal{F}_\rho^{\text{up}}[F](t_\nu)=z_\nu$. Then, we have that

$$\begin{aligned}\int_{F(t_\nu)}^{z_\nu}(F^{-1}(y)-t_\nu)dy &= \int_{F(t)}^{z_\nu}(F^{-1}(y)-t)dy \\ &\quad + \int_{F(t_\nu)}^{F(t)}(F^{-1}(y)-t)dy + \int_{F(t_\nu)}^{z_\nu}(t-t_\nu)dy,\end{aligned}$$

or equivalently, $\int_{F(t)}^{z_\nu}(F^{-1}(y)-t)dy = \rho - \int_{F(t_\nu)}^{F(t)}(F^{-1}(y)-t)dy - \int_{F(t_\nu)}^{z_\nu}(t-t_\nu)dy$. Since $0\le F(t_\nu)<z_\nu\le 1$, and $t_\nu\to t$ we get that $\int_{F(t_\nu)}^{z_\nu}(t-t_\nu)dy\to 0$. For the other term, we have w.l.o.g. that $F(t_\nu)\le y\le F(t)$. It then follows from (GI2) that $t_\nu\le F^{-1}(y)\le t$ and therefore $\left|\int_{F(t_\nu)}^{F(t)}(F^{-1}(y)-t)dy\right| \le \int_{F(t_\nu)}^{F(t)}|t_\nu-t|dy\to 0$. Thus,

$$\int_{F(t)}^{z_\nu}(F^{-1}(y)-t)dy \to \rho = \int_{F(t)}^{z}(F^{-1}(y)-t)dy \tag{A.1}$$

for a unique $z\in[F(t),1]$. Since $z'\mapsto\int_{F(t)}^{z'}(F^{-1}(y)-t)dy$ is strictly increasing (near z) and continuous, its inverse is well defined and continuous (see e.g., [35, Theorem 5, Page 168]). Thus, we get from (A.1) that $z_\nu\to z$, establishing continuity of $\mathcal{F}_\rho^{\text{up}}[F]$.

Next, let $F'\in\mathcal{CD}([a,b])$ with $W_1(F,F')\le\rho$. Equivalently, $\int_a^b|F'(t)-F(t)|dt\le\rho$. We show (4.6) by contradiction. Assume w.l.o.g. that the upper bound in (4.6) is violated, and there exists t^* with $F'(t^*)>\mathcal{F}_\rho^{\text{up}}[F](t^*)$. Then necessarily $t^*\in[a,t_\rho^{\text{up}}[F])$, and since $F'(t^*)>F(t^*)$, (GI1) implies that $F^{-1}(F'(t^*))>t^*$. Hence,

$[t^*, F^{-1}(F'(t^*)))$ is nonempty and we get from (GI3) that $F'(t) \geq F(t)$ for all $t \in [t^*, F^{-1}(F'(t^*)))$. Consequently, we obtain

$$\begin{aligned}
\rho &\geq \int_a^b |F'(t) - F(t)|dt \geq \int_{t^*}^{F^{-1}(F'(t^*))} |F'(t) - F(t)|dt \\
&= \int_{t^*}^{F^{-1}(F'(t^*))} (F'(t) - F(t))dt \geq \int_{t^*}^{F^{-1}(F'(t^*))} (F'(t^*) - F(t))dt \\
&= \int_{F(t^*)}^{F'(t^*)} (F^{-1}(y) - t^*)dy > \int_{F(t^*)}^{\mathcal{F}_\rho^{\text{up}}[F](t^*)} (F^{-1}(y) - t^*)dy = \rho,
\end{aligned}$$

which is a contradiction. ∎

Proof of Proposition 4.7. We break the proof into several steps.

Step 1: all indices j_k and i_k are well defined and satisfy (4.10). We need to establish that the min and max operations for the definitions of these indices are not taken over the empty set. To show this for all $k \in [1 : k_{\max}]$, we verify the following Induction Hypothesis (IH):

(IH) For each $k \in [1 : k_{\max}]$, j_k, i_k are well defined, $j_k < i_k$, and $b_{i_k,j_k} \geq \rho$.

All properties of (IH) can be directly checked for $k = 1$ by the definition of j_1 and i_1, and the assumption $b_{N,0} > \rho$. For the general case, let $k \leq k_{\max} - 1$ and assume that (IH) is fulfilled. Then, j_{k+1} is well defined because $b_{i_k,j_k} \geq \rho$ by (IH). To show this also for i_{k+1} we first establish that $i_k < N$. Indeed, assume on the contrary that $i_k = N$. Then, from the definition of j_{k+1} we have that $b_{i_k,j_{k+1}} < \rho$ and we get from the definition of $k_{\max}$ that $k \geq k_{\max}$, which is a contradiction. Since $i_k < N$, $[i_{k+1} : N]$ is nonempty. Combining this with the fact that $b_{N,j_{k+1}} > \rho$, which follows from the definition of $k_{\max}$ and our assumption $k < k_{\max}$, we deduce that the minimum in the definition of i_{k+1} is taken over a non-empty set. Hence, i_{k+1} is well defined. In addition, we get from the definitions of j_{k+1} and i_{k+1} that $j_{k+1} < i_{k+1}$ and from the definition of i_{k+1} that $b_{i_{k+1},j_{k+1}} \geq \rho$. Thus, we have shown (IH). Finally, $j_{k_{\max}+1}$ is also well defined because $b_{i_{k_{\max}},j_{k_{\max}}} \geq \rho$ by (IH). Having established that j_k and i_k are well defined for all $k \in [1 : k_{\max}+1]$, (4.10) follows directly from their expressions.

Step 2: establishing (4.11). By the definition of j_{k+1}, we get

$$b_{i_k,j_{k+1}} < \rho \quad \forall k \in [1 : k_{\max}]. \tag{A.2}$$

In addition, we have that

$$b_{i_{k+1}-1,j_{k+1}} < \rho \quad \forall k \in [0 : k_{\max}]. \tag{A.3}$$

For $k = 0$ this follows from the definition of j_1 and i_1. To show it also for $k \in [1 : k_{\max}]$ we consider two cases. If $b_{i_k+1,j_{k+1}} \geq \rho$, then, by definition, $i_{k+1} = i_k + 1$ and we get from (A.2) that $b_{i_{k+1}-1,j_{k+1}} = b_{i_k,j_{k+1}} < \rho$. In the other case where $b_{i_k+1,j_{k+1}} < \rho$, (A.3) follows directly from the definition of i_{k+1}. Next, note that due to (4.10) and the fact that $i_{k_{\max}+1} = N + 1$, the times τ_ℓ are indeed defined for all $\ell \in [i_1 : N]$. In addition, for each $k \in [1 : k_{\max}]$ we get from (A.3) that $\rho - b_{\ell,j_{k+1}} > 0$ for all $\ell \in [i_k : i_{k+1} - 1]$. Hence, Δt_ℓ is positive and strictly decreasing with $\ell \in [i_k : i_{k+1} - 1]$ and we have from the definition of the τ_ℓ's that

$$\tau_\ell < \tau_{\ell'} \quad \forall k \in [1 : k_{\max}], \ell, \ell' \in [i_k : i_{k+1} - 1] \text{ with } \ell < \ell' \tag{A.4}$$

$$\tau_{i_{k+1}-1} < t_{j_{k+1}} \quad \forall k \in [1 : k_{\max}]. \tag{A.5}$$

By the definition of j_{k+1} we further obtain that

$$b_{i_k,j_{k+1}-1} \geq \rho \quad \forall k \in [1:k_{\max}]. \tag{A.6}$$

From the latter and the definition of Δt_{i_k}, which implies that $\Delta t_{i_k} \sum_{l=j_{k+1}}^{i_k} c_l + b_{i_k,j_{k+1}} = \rho$, we get that $b_{i_k,j_{k+1}-1} \geq \Delta t_{i_k} \sum_{l=j_{k+1}}^{i_k} c_l + b_{i_k,j_{k+1}}$, or equivalently, that

$$\sum_{l=j_{k+1}-1}^{i_k} (t_l - t_{j_{k+1}-1})c_l - \sum_{l=j_{k+1}}^{i_k} (t_l - t_{j_{k+1}})c_l \geq \Delta t_{i_k} \sum_{l=j_{k+1}}^{i_k} c_l \Leftrightarrow$$
$$\sum_{l=j_{k+1}}^{i_k} (t_{j_{k+1}} - t_{j_{k+1}-1})c_l \geq \Delta t_{i_k} \sum_{l=j_{k+1}}^{i_k} c_l \Leftrightarrow t_{j_{k+1}} - t_{j_{k+1}-1} \geq \Delta t_{i_k}.$$

Thus, we deduce from the definition of τ_ℓ with $\ell \equiv i_k$ that $\tau_{i_k} \geq t_{j_{k+1}-1}$ for $k \in [1:k_{\max}]$. Using this, and recalling that $\{t_\ell\}_{\ell=0}^N$ are strictly increasing, we get from (4.10), (A.4), and (A.5), that $\{\tau_\ell\}_{\ell=j_1}^N$ are strictly increasing and (4.11) is satisfied.

Step 3: verification of the formula for $\widehat{F}^{\mathrm{up}}$ for $t \in (-\infty, a) \cup [\tau_N, \infty)$. For $t \in (-\infty, a)$, it follows directly from the definition of the upper CDF envelope. To establish it also when $t \in [\tau_N, \infty)$, it suffices again from the definition of the upper CDF envelope to show that $\tau_N = t_\rho^{\mathrm{up}}[\widehat{F}]$, with t_ρ^{up} given in the statement of Lemma 4.4. To show this, note that since by (4.11) $t_{j_{k_{\max}+1}-1} \leq \tau_N < t_{j_{k_{\max}+1}}$, we have

$$\int_{\tau_N}^b (1 - \widehat{F}(t))dt = \int_{\tau_N}^{t_N} (1 - \widehat{F}(t))dt = \int_{\tau_N}^{t_{j_{k_{\max}+1}}} (1 - \widehat{F}(t))dt$$
$$+ \int_{t_{j_{k_{\max}+1}}}^{t_N} (1 - \widehat{F}(t))dt = (t_{j_{k_{\max}+1}} - \tau_N) \sum_{l=j_{k_{\max}+1}}^{N} c_l + b_{N,j_{k_{\max}+1}},$$

which, in turn, equals $\Delta t_N \sum_{l=j_{k_{\max}+1}}^N c_l + b_{N,j_{k_{\max}+1}}$. Thus, we get from the definition of Δt_N that $\int_{\tau_N}^b (1 - \widehat{F}(t))dt = \frac{\rho - b_{N,j_{k_{\max}+1}}}{\sum_{l=j_{k_{\max}+1}}^N c_l} \sum_{l=j_{k_{\max}+1}}^N c_l + b_{N,j_{k_{\max}+1}} = \rho$, and hence $\tau_N = \sup\{\tau \in [a,b] \,|\, \int_\tau^b (1 - \widehat{F}(t))dt \geq \rho\} = t_\rho^{\mathrm{up}}[\widehat{F}]$. It remains to verify the formula for $\widehat{F}^{\mathrm{up}}$ for all intermediate intervals, which are of the form $[t_{\mathrm{beg}}, t_{\mathrm{end}})$. To each of these intervals we also associate a right time-instant t_{rt}. For each $k \in [1:k_{\max}]$, t_{beg}, t_{end}, and t_{rt} are given by one of the following cases.

Case 1) $t_{\mathrm{beg}} = t_\ell$ and $t_{\mathrm{end}} = t_{\ell+1}$ with $\ell \in [j_k : j_{k+1} - 2]$, and $t_{\mathrm{rt}} = t_{i_k}$;
Case 2) $t_{\mathrm{beg}} = t_{j_{k+1}-1}$, $t_{\mathrm{end}} = \tau_{i_k}$, and $t_{\mathrm{rt}} = t_{i_k}$;
Case 3) $t_{\mathrm{beg}} = \tau_\ell$ and $t_{\mathrm{end}} = \tau_{\ell+1}$ with $\ell \in [i_k : i_{k+1} - 2]$, and $t_{\mathrm{rt}} = t_{\ell+1}$;
Case 4) $t_{\mathrm{beg}} = \tau_{i_{k+1}-1}$, $t_{\mathrm{end}} = t_{j_{k+1}}$, and $t_{\mathrm{rt}} = t_{i_{k+1}}$.

One can readily check from the formula for $\widehat{F}^{\mathrm{up}}$ that these cases cover all intermediate intervals. To verify the formula for all $[t_{\mathrm{beg}}, t_{\mathrm{end}})$ we will exploit the following fact:

Fact I) For each of the Cases 1)–4) and pair (t, y) with $t \in (t_{\mathrm{beg}}, t_{\mathrm{end}})$ and $y = \widehat{F}^{\mathrm{up}}(t)$, it holds that $\widehat{F}^{-1}(y) = t_{\mathrm{rt}}$.

Step 4: Proof of Fact I. Recall that

$$\widehat{F}^{\mathrm{up}}(t) = \sup\left\{ z \in [\widehat{F}(t), 1] \,\middle|\, \int_{\widehat{F}(t)}^z (\widehat{F}^{-1}(y) - t)dy \leq \rho \right\} \tag{A.7}$$

and note that

$$\int_{\widehat{F}(t_j)}^{\widehat{F}(t_i)} (\widehat{F}^{-1}(y) - t_j)dy = b_{i,j} \quad \forall\, 0 \le j \le i \le N. \tag{A.8}$$

We first consider Case 1). Let $t \in (t_\ell, t_{\ell+1})$ with $\ell \in [j_k : j_{k+1} - 2]$. Then, we have from (4.10) and (A.8) that

$$\int_{\widehat{F}(t)}^{\widehat{F}(t_{i_k})} (\widehat{F}^{-1}(y) - t)dy \ge \int_{\widehat{F}(t_{j_{k+1}-1})}^{\widehat{F}(t_{i_k})} (\widehat{F}^{-1}(y) - t_{j_{k+1}-1})dy = b_{i_k, j_{k+1}-1} \ge \rho,$$
$$\int_{\widehat{F}(t)}^{\widehat{F}(t_{i_k-1})} (\widehat{F}^{-1}(y) - t)dy \le \int_{\widehat{F}(t_{j_k})}^{\widehat{F}(t_{i_k-1})} (\widehat{F}^{-1}(y) - t_{j_k})dy = b_{i_k-1, j_k} < \rho,$$

where we exploited (A.6) and (A.3) for each last inequality, respectively. Thus, it follows from (A.7) that $\widehat{F}(t_{i_k-1}) < \widehat{F}^{\mathrm{up}}(t) \le \widehat{F}(t_{i_k})$, which implies by (GI4) that $\widehat{F}^{-1}(\widehat{F}^{\mathrm{up}}(t)) = t_{i_k} \equiv t_{\mathrm{rt}}$. For Case 2), let $t \in (t_{j_{k+1}-1}, \tau_{i_k})$. Then, we get from (A.8) and the definition of τ_{i_k} that

$$\int_{\widehat{F}(t)}^{\widehat{F}(t_{i_k})} (\widehat{F}^{-1}(y) - t)dy \ge \int_{\widehat{F}(t_{j_{k+1}-1})}^{\widehat{F}(t_{i_k})} (\widehat{F}^{-1}(y) - \tau_{i_k})dy = \int_{\widehat{F}(t_{j_{k+1}-1})}^{\widehat{F}(t_{i_k})} (\widehat{F}^{-1}(y) - t_{j_{k+1}})dy$$
$$+ \int_{\widehat{F}(t_{j_{k+1}-1})}^{\widehat{F}(t_{i_k})} (t_{j_{k+1}} - \tau_{i_k})dy = b_{i_k, j_{k+1}} + \Delta t_{i_k} \sum_{l=j_{k+1}}^{i_k} c_l = \rho,$$

whereas by arguing precisely as in Case 1), we get that $\int_{\widehat{F}(t)}^{\widehat{F}(t_{i_k-1})} (\widehat{F}^{-1}(y) - t)dy < \rho$. Thus, we deduce $\widehat{F}(t_{i_k-1}) < \widehat{F}^{\mathrm{up}}(t) \le \widehat{F}(t_{i_k})$, and hence, by (GI4), $\widehat{F}^{-1}(\widehat{F}^{\mathrm{up}}(t)) = t_{i_k} \equiv t_{\mathrm{rt}}$. The proof of Fact I for Cases 3) and 4) follows similar arguments and exploits the orderings (4.10) and (4.11), and we omit it for space reasons.

Step 5: verification of the formula for $\widehat{F}^{\mathrm{up}}$ for $t \in [a, \tau_N)$. Let any interval $(t_{\mathrm{beg}}, t_{\mathrm{end}})$ as given by Cases 1)–4), let $t \in (t_{\mathrm{beg}}, t_{\mathrm{end}})$, $\{t_\nu\}_{\nu \in \mathbb{N}} \subset (t_{\mathrm{beg}}, t_{\mathrm{end}})$ with $t_\nu \searrow t_{\mathrm{beg}}$, and denote $y \equiv \widehat{F}^{\mathrm{up}}(t)$, $y_\nu \equiv \widehat{F}^{\mathrm{up}}(t_\nu)$, $\nu \in \mathbb{N}$. Due to Fact I, $\widehat{F}^{-1}(y) = t_{\mathrm{rt}}$, $\widehat{F}^{-1}(y_\nu) = t_{\mathrm{rt}}$ for all $\nu \in \mathbb{N}$. We use this together with $z = \widehat{F}^{\mathrm{up}}(t) \Leftrightarrow \int_t^{\widehat{F}^{-1}(z)} (z - \widehat{F}(s))ds = \rho$ and the continuity of $\widehat{F}^{\mathrm{up}}$ (which implies $y_\nu \to y_{\mathrm{beg}} \equiv \widehat{F}^{\mathrm{up}}(t_{\mathrm{beg}})$) to get

$$\int_t^{F^{-1}(y)} (y - \widehat{F}(s))ds = \int_{t_\nu}^{F^{-1}(y_\nu)} (y_\nu - \widehat{F}(s))ds \quad \forall \nu \in \mathbb{N} \Leftrightarrow$$
$$\int_t^{t_{\mathrm{rt}}} (y - \widehat{F}(s))ds = \int_{t_\nu}^{t_{\mathrm{rt}}} (y_\nu - \widehat{F}(s))ds \quad \forall \nu \in \mathbb{N} \Leftrightarrow$$
$$\int_t^{t_{\mathrm{rt}}} (y - \widehat{F}(s))ds = \int_{t_{\mathrm{beg}}}^{t_{\mathrm{rt}}} (y_{\mathrm{beg}} - \widehat{F}(s))ds \Leftrightarrow$$
$$\int_t^{t_{\mathrm{rt}}} (y - y_{\mathrm{beg}})ds + \int_t^{t_{\mathrm{rt}}} (y_{\mathrm{beg}} - \widehat{F}(s))ds = \int_{t_{\mathrm{beg}}}^{t} (y_{\mathrm{beg}} - y_{\mathrm{low}})ds + \int_t^{t_{\mathrm{rt}}} (y_{\mathrm{beg}} - \widehat{F}(s))ds \Leftrightarrow$$
$$(y - y_{\mathrm{beg}})(t_{\mathrm{rt}} - t) = (y_{\mathrm{beg}} - y_{\mathrm{low}})(t - t_{\mathrm{beg}}),$$

with $y_{\text{low}} = \widehat{F}(t_{\text{beg}})$, cf. Figure 2. Hence, $y = y_{\text{beg}} + (y_{\text{beg}} - y_{\text{low}})\frac{t - t_{\text{beg}}}{t_{\text{rt}} - t} = y_{\text{low}} + (y_{\text{beg}} - y_{\text{low}})\frac{t_{\text{rt}} - t_{\text{beg}}}{t_{\text{rt}} - t}$. The proof is completed by verifying the formula for $\widehat{F}^{\text{up}}$ at t_{beg} for each interval given by Cases 1)–4), which follows from the definitions of y_ℓ and z_ℓ. ◻

Proof of Lemma 4.8. We exploit the following equivalences for any $F \in \mathcal{CD}([a,b])$ and pair (t,y) in the graph of its lower and upper CDF envelopes:

$$y = \mathcal{F}_\rho^{\text{low}}[F](t) \Leftrightarrow \int_{F^{-1}(y)}^{t} (F(s) - y)ds = \rho \tag{A.9a}$$

$$y = \mathcal{F}_\rho^{\text{up}}[F](t) \Leftrightarrow \int_{t}^{F^{-1}(y)} (y - F(s))ds = \rho. \tag{A.9b}$$

We also use the following elementary results about the left inverse of a CDF $F \in \mathcal{CD}(\mathbb{R})$, defined by $F_{\text{left}}^{-1}(y) := \inf\{t \in \mathbb{R} \,|\, F(t) \geq y\}$.

Fact II) For any $y \in (0,1)$, $F^{-1}(1-y) = a + b - \widetilde{F}_{\text{left}}^{-1}(y)$, where $\widetilde{F} \equiv \mathcal{F}^{\text{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[F]$.

Fact III) For any $y \in [0,1]$ and $t \in \mathbb{R}$, $\int_t^{F_{\text{left}}^{-1}(y)} (y - F(s))ds = \int_t^{F^{-1}(y)} (y - F(s))ds$.

Next, let $F \in \mathcal{CD}([a,b])$ and denote $\widetilde{F} \equiv \mathcal{F}^{\text{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[F]$ and $\widetilde{F}^{\text{up}} \equiv \mathcal{F}_\rho^{\text{up}}[\widetilde{F}]$. To prove the result, we show that $\mathcal{F}_\rho^{\text{low}}[F](t) = \mathcal{F}^{\text{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[\widetilde{F}^{\text{up}}](t)$ for any t for which these values are in $(0,1)$. Let $y = 1 - \widetilde{F}^{\text{up}}(a+b-t) = \mathcal{F}^{\text{refl}}_{(\frac{a+b}{2}, \frac{1}{2})}[\widetilde{F}^{\text{up}}](t) \in (0,1)$. We show that $\int_{F^{-1}(y)}^t (F(s) - y)ds = \rho$, which by (A.9a) implies that $\mathcal{F}_\rho^{\text{low}}[F](t) = y$. Indeed,

$$\begin{aligned}
\int_{F^{-1}(y)}^{t} (F(s) - y)ds &= \int_{F^{-1}(1 - \widetilde{F}^{\text{up}}(a+b-t))}^{t} (F(s) - (1 - \widetilde{F}^{\text{up}}(a+b-t)))ds \\
&= \int_{a+b-\widetilde{F}_{\text{left}}^{-1}(\widetilde{F}^{\text{up}}(a+b-t))}^{t} (F(s) - (1 - \widetilde{F}^{\text{up}}(a+b-t)))ds \\
&= \int_{a+b-t}^{\widetilde{F}_{\text{left}}^{-1}(\widetilde{F}^{\text{up}}(a+b-t))} (\widetilde{F}^{\text{up}}(a+b-t) - \widetilde{F}(s))ds \\
&= \int_{a+b-t}^{\widetilde{F}^{-1}(\widetilde{F}^{\text{up}}(a+b-t))} (\widetilde{F}^{\text{up}}(a+b-t) - \widetilde{F}(s))ds = \rho,
\end{aligned}$$

where we used Fact II in the second equality, that the reflection around $(\frac{a+b}{2}, \frac{1}{2})$, i.e., the change of variables $(t,y) \mapsto (a+b-t, 1-y)$ is an isometry in the third equality, Fact III in the fourth equality, and the equivalent characterization (A.9b) for $y = \mathcal{F}_\rho^{\text{up}}[F](t)$ in the last equality. ◻

Proof of Fact II. Let $y \in (0,1)$. Then

$$\begin{aligned}
F^{-1}(1-y) &= \inf\{t \in \mathbb{R} \,|\, F(t) > 1-y\} = \inf F^{-1}((1-y, \infty)) \\
&= \sup F^{-1}((-\infty, 1-y]) = \sup\{t \in \mathbb{R} \,|\, F(t) \leq 1-y\} \\
&= \sup\{t \in \mathbb{R} \,|\, 1 - \widetilde{F}(a+b-t) \leq 1-y\} \\
&= \sup\{a+b-\tau, \tau \in \mathbb{R} \,|\, 1 - \widetilde{F}(\tau) \leq 1-y\} \\
&= a+b+\sup\{-\tau, \tau \in \mathbb{R} \,|\, \widetilde{F}(\tau) \geq y\} \\
&= a+b-\inf\{\tau \in \mathbb{R} \,|\, \widetilde{F}(\tau) \geq y\} = a+b-\widetilde{F}_{\text{left}}^{-1}(y),
\end{aligned}$$

where we used F is increasing and $\inf I = \sup I^c$ for any intervals I, I^c with $I \cup I^c = \mathbb{R}$ in the third equality. ◻

Proof of Fact III. To show the result we will prove that $\int_{F_{\text{left}}^{-1}(y)}^{F^{-1}(y)}(y - F(s))ds = 0$. Since $F^{-1}(y) \geq F_{\text{left}}^{-1}(y)$, it suffices to consider the case of strict inequality. Then, the result follows directly from the fact that $F(s) = y$ for any $s \in (F_{\text{left}}^{-1}(y), F^{-1}(y))$, which can be readily checked by the definitions of F^{-1} and F_{left}^{-1}. ∎

Appendix B. Derivation of the CDF equation. An equation for the Cumulative Distribution Function of $u(\mathbf{x}, t)$, solution of (3.1), obeying Assumption 3.1 and Assumption 3.2, is obtained via the Method of Distributions in three steps. First, we rely on the following inequalities for the newly introduced random variable $\Pi(\widetilde{\mathbf{x}}, t)$

$$\frac{\partial \Pi}{\partial t} = -\frac{\partial \Pi}{\partial U}\frac{\partial u}{\partial t}, \quad \nabla \Pi = -\frac{\partial \Pi}{\partial U}\nabla u. \tag{B.1}$$

Second, we multiply (3.1) by $-\frac{\partial \Pi}{\partial U}$ and, accounting for (B.1), we obtain a stochastic PDE for $\Pi(U, \mathbf{x}, t)$:

$$\frac{\partial \Pi}{\partial t} + \dot{\mathbf{q}}(U) \cdot \nabla \Pi = -\frac{\partial \Pi}{\partial U} r(U), \quad \mathbf{x} \in \Omega, U \in \mathbb{R}, t > 0, \tag{B.2}$$

with $\dot{\mathbf{q}} = \partial \mathbf{q}/\partial U$. This formulation is exact in case of smooth solutions of (3.1) [23] and whenever $\nabla \cdot \mathbf{q}(U) = 0$. (B.2) is defined in an augmented $(d+1)$-dimensional space $\widetilde{\Omega} = \Omega \times \mathbb{R}$, and it is subject to initial and boundary conditions that follow from the initial and boundary conditions of the original model

$$\begin{aligned} \Pi(U, \mathbf{x}, t = 0) = \Pi_0 = \mathcal{H}(U - u_0(\mathbf{x})), \quad & \widetilde{\mathbf{x}} \in \widetilde{\Omega} \\ \Pi(U, \mathbf{x}, t) = \Pi_b(U, \mathbf{x}, t) = \mathcal{H}(U - u_b(t)), \quad & \mathbf{x} \in \Gamma, U \in \Omega_U, t > 0. \end{aligned}$$

Finally, since the ensemble average of Π is the CDF of u, $F_{u(\mathbf{x},t)} = \langle \Pi(U, \mathbf{x}, t) \rangle$, ensemble averaging of (B.2) yields (5.1). This equation is subject to initial and boundary conditions along $(\Gamma \times \mathbb{R})$

$$\begin{aligned} F_{u(\mathbf{x},t)} = F_{u_0(\mathbf{x})}, \quad & \widetilde{\mathbf{x}} \in \widetilde{\Omega}, t = 0 \\ F_{u(\mathbf{x},t)} = F_{u_b(\mathbf{x},t)}, \quad & \mathbf{x} \in \Gamma, U \in \mathbb{R}, t > 0. \end{aligned} \tag{B.3}$$

The relaxation of Assumptions 3.1 and 3.2 leads to different (and often approximated) CDF equations: we refer to [6, 7] for a complete discussion.

REFERENCES

[1] A. Ben-Tal, D. D. Hertog, A. D. Waegenaere, B. Melenberg, and G. Rennen, *Robust solutions of optimization problems affected by uncertain probabilities*, Manage. Sci., 59 (2013), p. 341357.

[2] D. Bertsimas, D. B. Brown, and C. Caramanis, *Theory and applications of robust optimization*, SIAM Rev., 53 (2011), p. 464501.

[3] J. Blanchet, Y. Kang, and K. Murthy, *Robust Wasserstein profile inference and applications to machine learning*, J. Appl. Prob., 56 (2019), pp. 830–857.

[4] D. Boskos, J. Cortés, and S. Martinez, *Data-driven ambiguity sets with probabilistic guarantees for dynamic processes*, IEEE Trans. Aut. Contr., (2019). Submitted. Available at https://arxiv.org/abs/1909.11194.

[5] D. Boskos, J. Cortés, and S. Martínez, *Dynamic evolution of distributional ambiguity sets and precision tradeoffs in data assimilation*, in European Control Conference, Naples, Italy, June 2019, pp. 2252–2257.

[6] F. Boso, S. V. Broyda, and D. M. Tartakovsky, *Cumulative distribution function solutions of advection-reaction equations with uncertain parameters*, Proc. Roy. Soc. A, 470 (2014), p. 20140189.

[7] F. Boso and D. M. Tartakovsky, *Data-informed method of distributions for hyperbolic conservation laws*, SIAM J. Sci. Comput., 42 (2020), pp. A559–A583.

[8] S. E. Buckley, M. Leverett, et al., *Mechanism of fluid displacement in sands*, Trans. AIME, 146 (1942), pp. 107–116.

[9] A. Cherukuri and J. Cortés, *Cooperative data-driven distributionally robust optimization*, IEEE Trans. Aut. Contr., 65 (2020). To appear.

[10] E. Delage and Y. Ye, *Distributionally robust optimization under moment uncertainty with application to data-driven problems*, Operations Research, 58 (2010), p. 595612.

[11] S. Dereich, M. Scheutzow, and R. Schottstedt, *Constructive quantization: Approximation by empirical measures*, Annales de l'Institut Henri Poincar, Probabilits et Statistiques, 49 (2013), p. 11831203.

[12] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, The Annals of Mathematical Statistics, (1956), pp. 642–669.

[13] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations*, Mathematical Programming, 171 (2018), pp. 115–166.

[14] N. Fournier and A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure*, Probability Theory and Related Fields, 162 (2015), p. 707738.

[15] R. Gao and A. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, arXiv preprint arXiv:1604.02199, (2016).

[16] Y. Guo, K. Baker, E. DallAnese, Z. Hu, and T. H. Summers, *Data-based distributionally robust stochastic optimal power flowPart I: Methodologies*, IEEE Transactions on Power Systems, 34 (2018), pp. 1483–1492.

[17] R. Jiang, M. Ryu, and G. Xu, *Data-driven distributionally robust appointment scheduling over Wasserstein balls*, arXiv preprint arXiv:1907.03219, (2019).

[18] A. N. Kolmogorov, *Sulla determinazione empírica di uma legge di distribuzione*, Giornale dell' Istituto Italiano degli Attuari, 4 (1933).

[19] J.-P. Lebacque, *First-order macroscopic traffic flow models: Intersection modeling, network modeling*, in Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory, University of Maryland, College Park, 2005.

[20] D. Li, D. Fooladivanda, and S. Martínez, *Data-driven variable speed limit design for highways via distributionally robust optimization*, in European Control Conference, Napoli, Italy, June 2019, pp. 1055–1061.

[21] P. Massart, *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality*, The Annals of Probability, (1990), pp. 1269–1283.

[22] A. B. Owen, *Nonparametric likelihood confidence bands for a distribution function*, Journal of the American Statistical Association, 90 (1995), pp. 516–521.

[23] B. Perthame, *Kinetic Formulation of Conservation Laws*, vol. 21, Oxford University Press, 2002.

[24] G. Pflug and D. Wozabal, *Ambiguity in portfolio selection*, Quantitative Finance, 7 (2007), pp. 435–442.

[25] R. Racke, *Lectures on Nonlinear Evolution Equations: Initial Value Problems*, vol. 19, Springer, 1992.

[26] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Springer, 2015.

[27] A. Shapiro, *Distributionally robust stochastic programming*, SIAM Journal on Optimization, 27 (2017), pp. 2258–2275.

[28] A. Shapiro and S. Ahmed, *On a class of minimax stochastic programs*, SIAM Journal on Optimization, 14 (2004), pp. 1237–1249.

[29] N. V. Smirnov, *Approximate laws of distribution of random variables from empirical data*, Uspekhi Matematicheskikh Nauk, 10 (1944), pp. 179–206.

[30] D. M. Tartakovsky and P. A. Gremaud, *Method of distributions for uncertainty quantification*, Handbook of Uncertainty Quantification, (2017), pp. 763–783.

[31] D. Venturi, D. M. Tartakovsky, A. M. Tartakovsky, and G. E. Karniadakis, *Exact pdf equations and closure approximations for advective-reactive transport*, Journal of Computational Physics, 243 (2013), pp. 323–343.

[32] C. Villani, *Topics in Optimal Transportation*, no. 58, American Mathematical Society, 2003.

[33] J. Weed and F. Bach, *Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance*, Bernoulli, 25 (2019), pp. 2620–2648.

[34] C. K. Wikle and L. M. Berliner, *A Bayesian tutorial for data assimilation*, Physica D: Nonlinear Phenomena, 230 (2007), pp. 1–16.

[35] V. A. Zorich, *Mathematical Analysis I*, Springer, 2003.