PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Boso F, Tartakovsky DM. 2020 Learning on dynamic statistical manifolds. *Proc. R. Soc. A* **476**: 20200213. http://dx.doi.org/10.1098/rspa.2020.0213

Received: 26 March 2020
Accepted: 24 June 2020

Subject Areas:
mathematics

Keywords:
method of distributions, Bayesian inference, parameter identification

Author for correspondence:
Daniel M. Tartakovsky
e-mail: tartakovsky@stanford.edu

Learning on dynamic statistical manifolds

F. Boso and D. M. Tartakovsky

Department of Energy Resources Engineering, Stanford University, Stanford, CA 94305, USA

FB, 0000-0002-9066-0736; DMT, 0000-0001-9019-8935

Hyperbolic balance laws with uncertain (random) parameters and inputs are ubiquitous in science and engineering. Quantification of uncertainty in predictions derived from such laws, and reduction of predictive uncertainty via data assimilation, remain an open challenge. That is due to nonlinearity of governing equations, whose solutions are highly non-Gaussian and often discontinuous. To ameliorate these issues in a computationally efficient way, we use the method of distributions, which here takes the form of a deterministic equation for spatio-temporal evolution of the cumulative distribution function (CDF) of the random system state, as a means of forward uncertainty propagation. Uncertainty reduction is achieved by recasting the standard loss function, i.e. discrepancy between observations and model predictions, in distributional terms. This step exploits the equivalence between minimization of the square error discrepancy and the Kullback–Leibler divergence. The loss function is regularized by adding a Lagrangian constraint enforcing fulfilment of the CDF equation. Minimization is performed sequentially, progressively updating the parameters of the CDF equation as more measurements are assimilated.

1. Introduction

Robust and efficient quantification of parametric uncertainty in hyperbolic balance and conservations laws is hampered by their nonlinearity and solution structure, which typically possesses sharp gradients and often exhibits shocks and/or discontinuities. Many uncertainty quantification techniques (e.g. stochastic finite elements and stochastic collocation), which can be orders of magnitude faster than standard Monte Carlo simulations (MCS) when applied to elliptic and parabolic equations, often underperform on hyperbolic problems.

THE ROYAL SOCIETY PUBLISHING

The method of distributions (MD) [1] is an uncertainty quantification technique that is tailor-made for hyperbolic problems with random coefficients and inputs. Its goal is to derive a deterministic partial differential equation (PDE) for either the probability density function (PDF) or the cumulative distribution function (CDF) of the model output. In the presence of multiplicative noise introduced, e.g. by random parameter fields, MD requires a closure approximation, which is derived either via perturbation expansions or by resorting to phenomenology [2–4]. The method does not rely on a finite-term representation (e.g. via a truncated Karhunen–Loève expansion) of random parameter fields and, hence, does not suffer from the so-called curse of dimensionality [1,5]; its computational cost is independent of the correlation length of an input parameter [6] and can be orders of magnitude lower than that of MCS [2,4,7], and its accuracy increases as the correlation length decreases [1,8].

While MD enables one to quantify predictive uncertainty in hyperbolic models, assimilation of observations into probabilistic model predictions, e.g. by means of Bayes' rule, facilitates reduction of this uncertainty. Within this framework, the model provides a link between observed quantities and the estimates of the state, filtered through an observational map [9]. Direct application of Bayes' rule is often impractical because of the high dimensionality of a joint PDF of system states, and because of complex relations between parameters and states, and between states and observations [10, sec. 10.2]. For these reasons, a plethora of approximation techniques have been proposed. Some of these, e.g. maximum-likelihood estimation (MLE) [11] and maximum *a posteriori* estimation (MAP) [12], aim to identify the mode of a posterior distribution, which can be inadequate if the latter is highly non-Gaussian (e.g. multimodal), as is typical of nonlinear models. Ensemble Kalman filters (EnKF) [13] allow one to handle nonlinear PDEs but assume that their solutions are Gaussian. Other methods, e.g. Markov chain Monte Carlo [14] and particle filters [15], sample from the posterior directly and obviate the need for the Gaussianity and linearity assumptions. Like direct Bayesian updating, the methods of this class are computationally expensive because they rely on multiple forward solves of PDEs with uncertain (random) coefficients and/or auxiliary functions. Our goal is to eliminate this step by replacing it with MD.

Variational formulation recasts some of the methods described above (MLE, MAP, analysis step in EnKF) as a minimization problem in which a cost (loss) function contains the average distance between measurements and a model's predictions; parameter estimation is then accomplished by minimizing this loss function with respect to the model's parameters (and their statistical moments). This variational formulation belongs to a broader class of optimization methods, sometimes termed variational inference (VI) [16], that approximate Bayesian posterior densities by imposing closeness (in the Kullback–Leibler divergence sense) to the target density. Key innovations of our method are to reformulate the loss function in distributional terms using a different discrepancy metric and to confine both the prior and the posterior distributions to a dynamic statistical manifold defined by a deterministic CDF equation. Minimization is done with respect to variables used to parametrize the closure terms in the CDF equation; these variables are, in turn, expressed in terms of the statistical properties of the uncertain parameters and/or auxiliary functions of the original model.

Resulting PDE-constrained optimization problems can be solved with several techniques [17]. We employ a machine learning approach [18–20], which approximates a PDE's solution with a neural network whose coefficients are obtained by minimizing the resulting residual. This component of our algorithm places it in the burgeoning field variously known as physics-informed machine learning or data-aware modelling. Its goal is to overcome the scarcity of experimental data inherent in many physical systems by fusing physical constraints and observations. It is worthwhile emphasizing though that optimization techniques other than the one mentioned above can be used in our Bayesian data assimilation algorithm.

In §2, we formulate a data assimilation problem for hyperbolic PDEs with uncertain parameters and/or auxiliary functions, and introduce MD as a *forecast* step in Bayesian updating. Section 3 contains a novel *analysis* step, in which MD is used as a constraint to reduce parametric uncertainty; technical details are provided in appendix A. We refer to this combination of forecast

and analysis as the data-aware method of distributions (DA-MD). In §4, we test our approach on a linear inhomogeneous hyperbolic equation; this setting admits both exact and approximate Bayesian updates of the random parameters (either spatially uniform or variable) and, hence, enables us to verify the method's accuracy. Finally, in §5, we summarize the main findings and discuss future directions.

2. Forecast: method of distributions

While the data assimilation approach introduced here is applicable to other problems, we formulate it in §2a for hyperbolic PDEs with uncertain (random) parameters and/or auxiliary functions. This setting simplifies the derivation of a deterministic CDF equation used in §2b as the forecast step in Bayesian data assimilation.

(a) Problem formulation

We consider a smooth state variable $u(\mathbf{x},t):\Omega\times\mathbb{R}^+\to\mathbb{R}$, whose dynamics is governed by a nonlinear hyperbolic PDE

$$\frac{\partial u}{\partial t}+\nabla\cdot\mathbf{q}(u;\boldsymbol{\theta}_q)=r(u;\boldsymbol{\theta}_r),\quad \mathbf{x}\in\Omega,\quad t>0. \tag{2.1a}$$

This equation is subject to the initial condition

$$u(\mathbf{x},t=0)=u_0(\mathbf{x}),\quad \mathbf{x}\in\Omega, \tag{2.1b}$$

and, if the d-dimensional domain $\Omega\subset\mathbb{R}^d$ is bounded, to appropriate boundary conditions along the domain boundary $\partial\Omega$. The flux, $\mathbf{q}(u):\mathbb{R}\to\mathbb{R}^d$, and the source term, $r(u):\mathbb{R}\to\mathbb{R}$, are parametrized by $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_r$, respectively. These real-valued parameters can either be constant or vary in space ($\mathbf{x}$) and time (t). The functions $\mathbf{q}(u)$ and $r(u)$ are either linear or nonlinear, as long as the solution of (2.1) does not develop shocks.[1] For example, $u(\mathbf{x},t)$ is the concentration of a reactive solute advected by a flow velocity $\mathbf{v}(\mathbf{x})$, while undergoing chemical transformations; in this setting, $\mathbf{q}(u)=\mathbf{v}(\mathbf{x})u$ is the advective flux parametrized by $\mathbf{v}(\mathbf{x})$, and $r(u)$ represents a chemical reaction parametrized by a reaction rate constant k.

Incomplete or noisy measurements of the parameters $\boldsymbol{\theta}=\{\boldsymbol{\theta}_q,\boldsymbol{\theta}_r\}$ render them uncertain; this uncertainty is quantified by treating $\boldsymbol{\theta}$ as random fields and random variables. Additionally, auxiliary functions, such as the initial state $u_0(\mathbf{x})$ and boundary functions, can be uncertain/random. In the following, $\tilde{\boldsymbol{\theta}}$ denotes the complete set of random inputs, comprised of both $\boldsymbol{\theta}$ and auxiliary functions. This randomness renders $u(\mathbf{x},t)$, a solution of (2.1), random as well. Rather than computing low statistical moments of $u(\mathbf{x},t)$ (e.g. its ensemble mean $\bar{u}(\mathbf{x},t)$ and standard deviation $\sigma_u(\mathbf{x},t)$ that are commonly used to obtain an unbiased estimator of a system's dynamics and to quantify the corresponding predictive uncertainty, respectively), our goal is to compute its one-point CDF $F_u(U;\mathbf{x},t)\equiv\mathbb{P}[u(\mathbf{x},t)\le U]$, where $U\in\Omega_U\subseteq\mathbb{R}$. The value space for the random variable $u(\mathbf{x},t)$, $\Omega_U=[U_{\min},U_{\max}]$, identifies the support of the CDF $F_u(U;\cdot)$. The latter can be either infinite ($\Omega_U=\mathbb{R}$, with $U_{\min}=-\infty$ and $U_{\max}=+\infty$) or finite ($U_{\min},U_{\max}\in\mathbb{R}$ such that $U_{\min}<U_{\max}$).

The model (2.1) is supplemented with N_{meas} measurements of the state variable $u(\mathbf{x},t)$ collected at selected space–time points $(\mathbf{x},t)_m$ with $m=1,\dots,N_{\text{meas}}$. These data, $\mathbf{d}_{1:N_{\text{meas}}}=\{d_1,\dots,d_{N_{\text{meas}}}\}$, are assumed to differ from the corresponding exact model predictions $u[(\mathbf{x},t)_m]$ by a random measurement error ε_m,

$$d_m=u[(\mathbf{x},t)_m]+\varepsilon_m,\quad m=1,\dots,N_{\text{meas}}. \tag{2.2}$$

The measurement errors are assumed to have zero mean, $\mathbb{E}[\varepsilon_m]=0$, and to be mutually uncorrelated, $\mathbb{E}[\varepsilon_m\varepsilon_n]=0$ for all $m\neq n$. A complete probabilistic description of the data is encapsulated in the PDF $f_L(d_m|u[(\mathbf{x},t)_m]=U)$, which is also known as likelihood function. In the

[1]The presence of shocks and discontinuities complicates the derivation of CDF equations [4,21,22], obfuscating our focus on data assimilation.

absence of measurement errors, the observational PDF is given by the Dirac distribution $\delta(\cdot)$, i.e. $f_L(d_m|u[(\mathbf{x},t)_m]=U)=\delta(U-d_m)$.

(b) Cumulative distribution function equation

Direct numerical computation of the CDF $F_u(U;\mathbf{x},t)$, e.g. via MCS of (2.1), is computationally expensive. Instead, we use MD to derive a $(d+1)$-dimensional linear PDE for F_u (see appendix A for details),

$$\frac{\partial F_u}{\partial t}+\boldsymbol{\mathcal{Q}}(U;\mathbf{x},t)\cdot\widetilde{\nabla}F_u=\widetilde{\nabla}\cdot\left[\boldsymbol{\mathcal{D}}(U;\mathbf{x},t)\widetilde{\nabla}F_u\right],\quad(\mathbf{x},U)\in\widetilde{\Omega},\quad t>0. \tag{2.3}$$

This deterministic PDE is defined in the augmented space $\tilde{\Omega}=\Omega\cup\Omega_U$.This equation is subject to initial and boundary conditions that reflect uncertainty in the initial and boundary conditions for the original problem (2.1). Additional boundary conditions are defined for $\partial\Omega_U$, $F_u(U_{\min};\cdot)=0$ and $F_u(U_{\max};\cdot)=1$; they stem from the definition of a CDF.

In general, derivation of (2.3) requires a closure approximation, such as the perturbation expansion used in appendix A. Notable exceptions of practical significance include a scenario of random inputs (initial and boundary conditions) but deterministic parameters $\boldsymbol{\theta}$;[2] in this case, (2.3) is exact and its coefficients are given by (appendix A)

$$\boldsymbol{\mathcal{Q}}(U;\mathbf{x},t)=\{\dot{\mathbf{q}}(U;\boldsymbol{\theta}_q),r(U;\mathbf{x},t)\},\quad\boldsymbol{\mathcal{D}}(U;\mathbf{x},t)=\mathbf{0}, \tag{2.4}$$

where $\dot{\mathbf{q}}(U)=\mathrm{d}\mathbf{q}(U)/\mathrm{d}U$. When the model parameters $\boldsymbol{\theta}$ are random, i.e. when the CDF equation (2.3) is inexact, the coefficients $\boldsymbol{\mathcal{Q}}$ and $\boldsymbol{\mathcal{D}}$ depend on a set $\boldsymbol{\varphi}$ of statistical parameters that characterize the randomness of $\boldsymbol{\theta}$. This set consists of the shape parameters of PDFs of $\tilde{\boldsymbol{\theta}}$, i.e. their means, variances and correlation lengths. Together with $(\mathbf{x},t)$ and the statistical characteristics of the random auxiliary functions, these parameters represent the coordinates $\tilde{\boldsymbol{\varphi}}$ of a manifold of distributions $\mathcal{F}(F_u)$, whose dynamics is governed by the CDF equation (2.3). Each point in this finite-dimensional coordinate space $\tilde{\boldsymbol{\varphi}}$ uniquely identifies a distribution [24].

The use of perturbative closures to derive a CDF equation raises several questions about its accuracy and robustness, which have been the subject of previous investigations. First, even though the coefficient of variation (CV) of the model parameters serves as a perturbation parameter, the resulting CDF equations for many applications remain accurate for relatively large values of CV [2,8,25]. Second, the coefficients of perturbation-based CDF equations, such as $\boldsymbol{\mathcal{Q}}$ and $\boldsymbol{\mathcal{D}}$ in (2.3), depend only on the low-order statistical moments (such as $\boldsymbol{\varphi}$) of the model parameters, rather than their full PDFs. By using an advection–reaction equation as a test case, we show in appendix A that the resulting CDF equation is distributionally robust, giving consistent predictions of the system state's CDF regardless of whether the model coefficient (spatially varying reaction rate) has a Gaussian, lognormal, or uniform PDF. Third, the accuracy of perturbation-based CDF equations depends on correlation lengths of the model parameters: these equations are often exact for white noise (zero correlation) and become progressively less so as the correlation lengths increase. If the correlation lengths are large, perturbation-based closures can be replaced with truncated Karhunen–Loéve expansions of the random parameter fields, leading to accurate/exact CDF equations [6].

In summary, we use the CDF equation (2.3) as an efficient forecasting tool, which propagates parametric uncertainty in space and in time through a physical model. It represents a counterpart of a set of ensemble members or particles in the context of EnKF or particle filter, respectively. Its accuracy and computational efficiency *vis-à-vis* MCS have been throughly investigated [2,4,7].

3. Analysis: sequential Bayesian update on dynamic manifolds

We use MD as a constraint for the analysis step, during which observations of the system state are used to refine the knowledge of the meta-parameters $\boldsymbol{\varphi}$. Specifically, our novel analysis step

[2]When both the inputs and parameters are deterministic, the strategy of transforming a nonlinear d-dimensional hyperbolic PDE into its linear $(d+1)$-dimensional counterpart is referred to as kinetic formulation of a hyperbolic conservation law [23].

involves minimization of the discrepancy between the 'observational' CDF $\hat{F}_u(U;(\mathbf{x},t)_m)$ in each measurement location ($m=1,\ldots,N_{\text{meas}}$) and the corresponding 'estimate' CDF $F_u(U;\boldsymbol{\varphi};(\mathbf{x},t)_m)$:

$$\boldsymbol{\varphi}^{(m)} = \underset{\boldsymbol{\varphi}}{\operatorname{argmin}} \|\hat{F}_u(U;(\mathbf{x},t)_m) - F_u(U;\boldsymbol{\varphi};(\mathbf{x},t)_m)\|_2 \quad \text{subject to} \quad F_u \in \mathcal{F}, \tag{3.1}$$

where

$$\|\hat{F}_u(U;(\mathbf{x},t)_m) - F_u(U;\boldsymbol{\varphi};(\mathbf{x},t)_m)\|_2 = \left(\int_{\Omega_U} (\hat{F}_u(U;(\mathbf{x},t)_m) - F_u(U;\boldsymbol{\varphi};(\mathbf{x},t)_m))^2 \mathrm{d}U\right)^{1/2}.$$

The analysis step, i.e. minimization of (3.1), is performed sequentially for each observation m, so that all the distributions above are uni-variate. Formulation (3.1) is at the core of our data assimilation strategy and requires a thorough explanation.

Remark 3.1. *MD constraint*: The estimate distribution $F_u(U;\boldsymbol{\varphi};(\mathbf{x},t)_m)$ is a solution of the CDF equation (2.3) subject to appropriate initial/boundary conditions. This boundary value problem is parametrized by the set of parameters $\boldsymbol{\varphi}$, over which the discrepancy minimization is performed. In other words, (3.1) identifies the parameters of the CDF equation that yield a CDF F_u in the measurement location as close as possible to the observational CDF $\hat{F}_u$. This implies that the minimization is performed on the manifold of distributions obeying the CDF equation. This observation is further elaborated upon in §3b. Reliance on MD obviates the need for both Gaussianity assumption for the system states and the linearity requirement for the physical model, as long as it is possible to develop a reliable and accurate CDF equation.

Remark 3.2. *Observational CDFs*: We construct the observational CDF,

$$\hat{F}_u(U;(\mathbf{x},t)_m) = \int_{U_{\min}}^{U} \hat{f}_u(U;(\mathbf{x},t)_m)\,\mathrm{d}U,$$

via Bayesian update of the corresponding PDF $\hat{f}_u$ at each space–time measurement point m,

$$\hat{f}_u(U;(\mathbf{x},t)_m|d_m) \propto f_L(d_m|u[(\mathbf{x},t)_m]=U) f_u(U;\boldsymbol{\varphi}^{(m-1)};(\mathbf{x},t)_m). \tag{3.2}$$

The prior PDF $f_u(U;\boldsymbol{\varphi}^{(m-1)};(\mathbf{x},t)_m)$ is computed from a solution of the CDF equation (2.3) whose parameters $\boldsymbol{\varphi}^{(m-1)}$ are computed in the previous assimilation step. This procedure provides a *local* update of the system state's PDF in the sense that it yields no information on the surrounding locations or on the future time evolution of the state.

Remark 3.3. *Sequential update*: The sequential update of the observational PDF $\hat{f}_u$ allows us to obtain final estimates for the MD parameters $\boldsymbol{\varphi}$ that are conditional on all assimilated observations [26]. It is employed both to reduce the dimensionality of the CDFs/PDFs involved and to facilitate real-time update of the estimates as new measurements become available [10, p. 101]. At each step, or for each data point, $m=1,\ldots,N_{\text{meas}}$, we follow the following procedure.

- For $m=1$, the MD parameters $\boldsymbol{\varphi}^{(0)}$ are initialized to define the prior and to compute (3.2). The normalization constant that specifies $\hat{f}_u$ is obtained by (numerical) integration, $C_1 = \int f_L(d_1|U) f_u(U;\boldsymbol{\varphi}^{(0)};(\mathbf{x},t)_1)\,\mathrm{d}U$.
- For $m>1$, each update (3.2) accounts for conditioning on all previous measurements up to the current one, $\mathbf{d}_{1:m}$, such that

$$\hat{f}_u(U;(\mathbf{x},t)_m|\mathbf{d}_{1:m}) \propto f_L(\mathbf{d}_{1:m}|U) f_u(U;\boldsymbol{\varphi}^{(m-1)};(\mathbf{x},t)_m). \tag{3.3}$$

  This step implies that the prior distribution in the current measurement location m obeys the CDF equation (2.3). If observation errors are mutually uncorrelated, then $f_L(\mathbf{d}_{1:m}|U) =$

$\prod_{i=1}^{m} f_L(d_i|U)$ and

$$\hat{f}_u(U;(\mathbf{x},t)_m|\mathbf{d}_{1:m}) \propto \prod_{i=1}^{m-1} f_L(d_i|U) f_L(d_m|U) f_u(U;\boldsymbol{\varphi}^{(m-1)};(\mathbf{x},t)_m)$$

$$\propto f_L(d_m|U)\hat{f}_u(U;(\mathbf{x},t)_m|\mathbf{d}_{1:m-1}). \tag{3.4}$$

Here, $\hat{f}_u(U;(\mathbf{x},t)_m|\mathbf{d}_{1:m-1})$ is approximated by a solution of the CDF equation in $(\mathbf{x},t)_m$ with parameters $\boldsymbol{\varphi}^{(m-1)}$ from the previous iterative step. In other words, a solution of the CDF equation (2.3) with parameters $\boldsymbol{\varphi}^{(m-1)}$ serves as prior.

At the end of this sequential assimilation procedure, the CDF equation (2.3) with parameters $\boldsymbol{\varphi}^{(N_{\text{meas}})}$ allows us to predict the future dynamics of the CDF $F_u(U;\cdot)$, i.e. to make a probabilistic forecast.

Remark 3.4. *Choice of the discrepancy metric*: Our reliance on the squared L^2 norm (also known as Cramer's distance [27]),

$$\|F_1(U) - F_2(U)\|_2^2 = \int_{U_{\min}}^{U_{\max}} [F_1(U) - F_2(U)]^2 \mathrm{d}U,$$

as a measure of discrepancy between any two CDFs, $F_1(U)$ and $F_2(U)$, facilitates numerical minimization of the loss function in (3.1) with a technique described in §3a. We deploy it in place of a commonly used Kullback–Leibler (KL) divergence,

$$D_{\text{KL}}(F_1,F_2) = \int_{U_{\min}}^{U_{\max}} f_1(U) \ln \frac{f_1(U)}{f_2(U)} \mathrm{d}U, \quad \text{with } f_1(U) = \frac{\partial F_1}{\partial U}, \quad f_2(U) = \frac{\partial F_2}{\partial U}$$

for the following reasons. According to Pinsker's inequality [28,29], $D_{\text{KL}}[F_1,F_2] \geq (1/2)\|F_1 - F_2\|_1^2$, where $\|\cdot\|_1$ is the L^1 norm. Since $\|F_1 - F_2\|_1 \geq \|F_1 - F_2\|_2$ [30, Prop. 1.5], this yields $D_{\text{KL}}(F_1,F_2) \geq (1/2)\|F_1 - F_2\|_2^2$. Since $D_{\text{KL}}(F_1,F_2)$ and $\|F_1 - F_2\|_2$ share the same minimum (for $F_1 \equiv F_2$ both metrics are equal to zero), a solution of the minimization problem (3.1) would also minimize the corresponding loss function based on the KL divergence. Moreover, it is advantageous to employ MD in its CDF form, rather than its PDF form, because of the straightforward assignment of the boundary conditions along $\partial\Omega_U$ and smoother solutions.

Remark 3.5. *Relationship to VI techniques*: Our method aims at approximating posterior densities in a Bayesian sense via a minimization procedure. As such, it connects with VI techniques, which use optimization to identify one joint density—chosen to belong to a specified family of approximate densities—which is close to the target posterior in KL divergence terms [16]. We choose a physics-based family of plausible distributions, which obey the CDF equation parametrized with a finite set of parameters. Constraining distributions to a dynamic manifold allows us to consider sequentially the update of single-point distributions: updated parameters can be used, in combination with the CDF equation, to obtain forecast predictions in different space–time locations. Moreover, it reduces drastically (to one) the dimensionality of the posterior distribution to be updated at each assimilation step.[3]

(a) Loss function minimization

The PDE-constrained optimization problem (3.1) can be solved with several techniques [17]. If the CDF equation (2.3) admits an analytical solution, e.g. if the system parameters $\boldsymbol{\theta}$ are deterministic and the initial and/or boundary functions are random, $F_u(U;\boldsymbol{\varphi})$ can be expressed as a (semi)explicit function of the statistical parameters, $\boldsymbol{\varphi}_0$ and $\boldsymbol{\varphi}_b$, characterizing the initial and boundary CDFs F_0 and F_b, respectively. Section 4a deals with such a scenario; it serves to verify

[3]In this regard, we mention the work [24], where the reduction in complexity of statistical models is quantified by exploiting relevant embedding constraints specifying geodesic motion on curved statistical manifolds.

the reliability of our approach by comparing its performance with that of the standard Bayesian update.

When the CDF equation (2.3) has to be solved numerically, we follow [19,31] to approximate its solution, $F_u(U;\tilde{\boldsymbol{\varphi}})$, with a neural network $F_{\mathrm{NN}}(U;\tilde{\boldsymbol{\varphi}})$ whose coefficients (weights and biases) are computed by minimizing the residual

$$R = \frac{\partial F_{\mathrm{NN}}}{\partial t} + (\boldsymbol{\mathcal{Q}} - \tilde{\nabla}\cdot\boldsymbol{\mathcal{D}})\cdot\tilde{\nabla}F_{\mathrm{NN}} - \boldsymbol{\mathcal{D}}\tilde{\Delta}F_{\mathrm{NN}} \tag{3.5}$$

at a set of N_{res} points $\{(\mathbf{x},t)_r\}_{r=1}^{N_{\mathrm{res}}}$; the initial and boundary conditions are enforced at a finite set of N_{aux} points $\{(U,\mathbf{x},t)_r\}_{r=1}^{N_{\mathrm{aux}}}$. The derivatives in (3.5) are computed via automatic differentiation, as implemented in `TensorFlow` [32]. This procedure replaces the PDE-constrained minimization problem (3.1) with an optimization problem

$$\boldsymbol{\varphi}^{(m)} = \underset{\boldsymbol{\varphi}}{\mathrm{argmin}}\big\{\|\hat{F}(U;(\mathbf{x},t)_m) - F_{\mathrm{NN}}(U;(\mathbf{x},t)_m,\boldsymbol{\varphi})\|_2 + \mathrm{MSE}_R(\boldsymbol{\varphi}) + \mathrm{MSE}_B(\boldsymbol{\varphi})\big\}, \tag{3.6}$$

where

$$\mathrm{MSE}_R(\boldsymbol{\varphi}) = \frac{1}{N_{\mathrm{res}}}\sum_{r=1}^{N_{\mathrm{res}}}\|R((\mathbf{x},t)_r;\boldsymbol{\varphi})\|_2$$

and

$$\mathrm{MSE}_B(\boldsymbol{\varphi}) = \frac{1}{N_{\mathrm{aux}}}\sum_{i=1}^{N_{\mathrm{aux}}}\|F_{\mathrm{NN}}((U,\mathbf{x},t)_i,\boldsymbol{\varphi}) - F_{\mathrm{inp}}((U,\mathbf{x},t)_i)\|_2,$$

where F_{inp} represents the prescribed CDFs of either the initial state or the boundary functions along $\partial\tilde{\Omega}$. The NN function approximation via minimization enjoys convergence guarantees in the chosen L^2 norm (e.g. [33,34]). A solution of (3.6) provides a CDF surrogate (a 'trained' NN) and the set of optimal parameters $\boldsymbol{\varphi}$. The surrogate can then be used to update predictions and for forecast (not pursued here).

(b) Information-geometric interpretation

A family of distributions satisfying the CDF equation (2.3) defines a dynamic statistical manifold $\mathcal{F}[F_u;\tilde{\boldsymbol{\varphi}}]$. Each point in this space, with coordinates $\tilde{\boldsymbol{\varphi}} = (\mathbf{x},t,\boldsymbol{\varphi})$, uniquely identifies a physics-informed CDF $F_u(U;\mathbf{x},t)$ of the model's output $u(\mathbf{x},t)$ at each space–time point $(\mathbf{x},t)$. The manifold $\mathcal{F}$ is differentiable in all coordinate directions and equipped with a Riemannian metric. The latter takes the form of the Fisher information metric (FIM), a $(d+1+N_\varphi)\times(d+1+N_\varphi)$ matrix whose components are [35, p. 33]

$$g_{jk}(\tilde{\boldsymbol{\varphi}}) = \int \frac{\partial \ln f_u(U;\tilde{\boldsymbol{\varphi}})}{\partial\tilde{\varphi}_j}\frac{\partial \ln f_u(U;\tilde{\boldsymbol{\varphi}})}{\partial\tilde{\varphi}_k} f_u(U;\tilde{\boldsymbol{\varphi}})\,\mathrm{d}U, \quad j,k = 1,\ldots,d+1+N_\varphi, \tag{3.7}$$

where $N_{\tilde{\varphi}} = d+1+N_\varphi$ is the number of manifold coordinates, with N_φ statistical parameters in the CDF equation (2.3).[4] The local curvature of the manifold, g_{jk}, represents a Euclidean metric (a distance on the manifold $\mathcal{F}$) upon an appropriate change of variable. FIM quantifies the differential amount of information between two infinitesimally close points on a manifold; it is formally computed as the second derivative of the KL divergence of distributions $F_u(U;\tilde{\boldsymbol{\varphi}})$ and $F_u(U;\tilde{\boldsymbol{\varphi}}')$ with $\tilde{\boldsymbol{\varphi}}' \to \tilde{\boldsymbol{\varphi}}$ [36].

The significance of FIM and its geometric implications [37] will be explored elsewhere. Here, we focus on the calculation of the information gain achieved during each step of the data assimilation process. Specifically, we express an mth analysis step in geometrical terms as a change of the coordinates on the statistical manifold $\mathcal{F}$, from $\tilde{\boldsymbol{\varphi}}^{(m-1)}$ to $\tilde{\boldsymbol{\varphi}}^{(m)}$, and quantify the corresponding information gain by $D_{\mathrm{KL}}[F_u(U;\varphi^{(m)}),F_u(U;\varphi^{(m-1)})]$. This quantity is computed as a post-processing step for comparative analysis.

[4]This definition assumes the existence of the PDF f_u; for hyperbolic PDEs (2.1) with smooth solutions, it does exist and satisfies a PDF equation corresponding to the CDF equation (2.3) [1,6,8].

4. Numerical experiments

Let us consider a scalar $u(x,t):\mathbb{R}^+\times\mathbb{R}^+\to\mathbb{R}^+$, whose dynamics satisfies a one-dimensional dimensionless advection–reaction equation

$$\frac{\partial u}{\partial t}+\frac{\partial q(u)}{\partial x}=r(x,u),\quad q\equiv vu,\quad r\equiv -k(x)u;\quad x>0,\ t>0, \tag{4.1a}$$

subject to initial and boundary conditions

$$u(x,t=0)=u_0;\quad u(x=0,t)=u_{\mathrm{b}}+s(t),\quad s(t)=a\sin(2\pi\nu t+\phi)\,. \tag{4.1b}$$

This problem describes, for example, advection of a solute that undergoes linear decay; in this example, u represents the normalized solute concentration, v is the normalized flow velocity along a streamline and k is the normalized reaction rate. In the simulations reported below, we set $v=1$, $a=0.1$, $\nu=1$ and $\phi=3\pi/2$. In the first test, k is a deterministic constant, while the uniform initial state u_0 and baseline state u_{b} are random variables. In the other two tests, both u_0 and u_{b} are deterministic, and k is alternatively treated either as a random constant or as a spatially varying random field.

In all three experiments, datasets $\mathbf{d}=\{d_1,\dots,d_{N_{\mathrm{meas}}}\}$ are generated in accordance with (2.2) by adding Gaussian white noise, $\mathcal{N}(0,\sigma_\varepsilon)$, to a solution of (4.1) with a given choice of model parameters. The likelihood function, $f_L(d_m|u(x,t)_m)$ with $m=1,\dots,d_{N_{\mathrm{meas}}}$, is assumed to be Gaussian.

The CDF equation for (4.1) is derived, and the accuracy and robustness of the underlying closure approximations analysed, in [2] for the three scenarios described above. Appendix A contains a brief summary of these results.

(a) Uncertain initial and boundary conditions

Let u_0 and u_{b} be random uncorrelated random variables with (prior) PDFs $f_{u_0}(U_0)$ and $f_{u_{\mathrm{b}}}(U_{\mathrm{b}})$. Then the random initial and boundary states $u(x,t=0)$ and $u(x=0,t)$ are characterized by respective CDFs $F_0(U;\boldsymbol{\varphi}_0)$ and $F_{\mathrm{b}}(U;t,\boldsymbol{\varphi}_{\mathrm{b}})$ with shape parameters $\boldsymbol{\varphi}_0$ and $\boldsymbol{\varphi}_{\mathrm{b}}$. In the absence of other sources of uncertainty, CDF $F_u(U;x,t)$ of the random state $u(x,t)$ in (4.1) satisfies *exactly* a PDE

$$\frac{\partial F_u}{\partial t}+\frac{\partial F_u}{\partial x}-kU\frac{\partial F_u}{\partial U}=0, \tag{4.2a}$$

subject to initial and boundary conditions

$$F_u(U;x,0)=F_0,\quad F_u(U;0,t)=F_{\mathrm{b}},\quad F_u(U_{\min};x,t)=0\quad\text{or}\quad F_u(U_{max};x,t)=1. \tag{4.2b}$$

This boundary-value problem admits an analytical solution, with either F_0 or F_{b} that are propagated along deterministic characteristic lines. The dynamic manifold $\mathcal{F}$ of the resulting CDFs F_u has coordinates $\tilde{\boldsymbol{\varphi}}=\{x,t,\boldsymbol{\varphi}_0,\boldsymbol{\varphi}_{\mathrm{b}}\}$. The analysis step of DA-MD takes place on this statistical manifold. Each measurement contributes to uncertainty reduction of either $\boldsymbol{\varphi}_0$ or $\boldsymbol{\varphi}_{\mathrm{b}}$ (i.e. sharpens either f_{u_0} or $f_{u_{\mathrm{b}}}$), depending on the data location $(x,t)_m$. Half of these N_{meas} measurements are collected at locations informing the initial condition, i.e. $(x/t)_m>1$), and the other half at locations informing the boundary condition, i.e. $(x/t)_m<1$.

To verify the accuracy of DA-MD, we compare its predictions of the optimal parameters $\boldsymbol{\varphi}^{(N_{\mathrm{meas}})}$ with those given by the Bayesian posterior joint PDF:

$$\begin{aligned}\hat{f}_{u_0,u_{\mathrm{b}}}(U_0,U_{\mathrm{b}}|\mathbf{d}_{1:N_{\mathrm{meas}}})&=\hat{f}_{u_0}(U_0|\mathbf{d}_{1:N_{\mathrm{meas}}})\hat{f}_{u_{\mathrm{b}}}(U_{\mathrm{b}}|\mathbf{d}_{1:N_{\mathrm{meas}}})\\&\propto f_L(\mathbf{d}_{1:N_{\mathrm{meas}}}|\mathbf{u}[(x,t)_{1:N_{\mathrm{meas}}};U_0,U_{\mathrm{b}}]f_{u_0}(U_0)f_{u_{\mathrm{b}}}(U_{\mathrm{b}})\\&\approx\prod_{m=1}^{N_{\mathrm{meas}}}f_L(d_m|u[(x,t)_m;U_0,U_{\mathrm{b}}]f_{u_0}(U_0)f_{u_{\mathrm{b}}}(U_{\mathrm{b}}).\end{aligned} \tag{4.3}$$

To facilitate the Bayesian update, we take F_{u_0} and $F_{u_{\mathrm{b}}}$ to be Gaussian, fully specified by their respective means and standard deviations, $\boldsymbol{\varphi}_0=\{\mu_0,\sigma_0\}$ and $\boldsymbol{\varphi}_{\mathrm{b}}=\{\mu_{\mathrm{b}},\sigma_{\mathrm{b}}\}$. Then, (4.3) yields

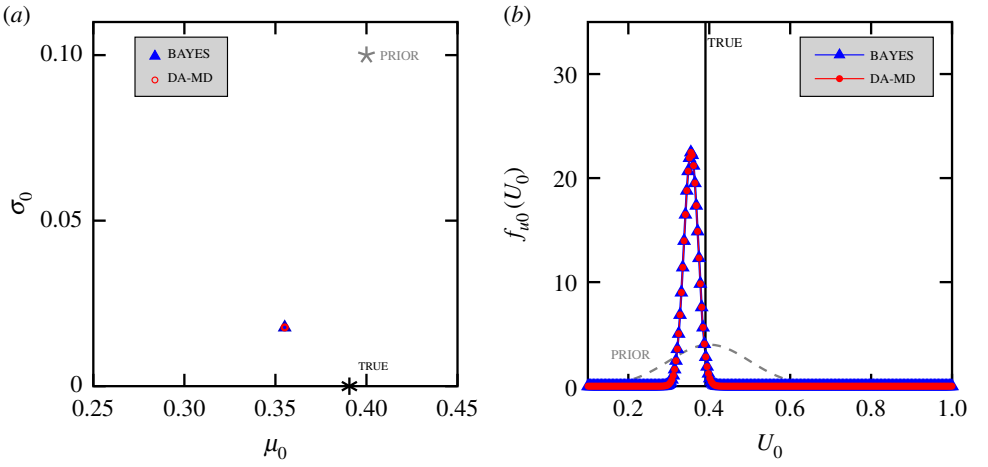

Figure 1. Prior and posterior distributions for the initial state u_0 on the statistical manifold defined by the coordinates $\{\mu_0, \sigma_0\}$ representing the mean and standard deviation of a Gaussian distribution (*a*), and in the value space (*b*). The black asterisk in (*a*) and the black vertical line in (*b*) represent the true value ($u_0^{\text{true}} = 0.391$), for which a Gaussian PDF degenerates into the Dirac distribution (delta function). The grey star (*a*) and the grey dashed line (*b*) represent a prior distribution ($\mu_0^{\text{prior}} = 0.4$, $\sigma_0^{\text{prior}} = 0.1$). The blue triangle (*a*) and line (*b*) identify the Bayesian solution, whereas the corresponding red symbols and lines identify the DA-MD solution. Parameters are set to $k = 1$, $\sigma_\varepsilon = 0.04$ and $N_{\text{meas}} = 20$. (Online version in colour.)

analytically computable Gaussian posteriors $\hat{f}_{u_0}(U_0|\mathbf{d}_{1:N_{\text{meas}}})$ and $\hat{f}_{u_b}(U_b|\mathbf{d}_{1:N_{\text{meas}}})$. In what follows, we compare those with the posterior parameters obtained via DA-MD, $\boldsymbol{\varphi}_0^{(N_{\text{meas}})}$ and $\boldsymbol{\varphi}_b^{(N_{\text{meas}})}$, respectively. These posterior DA-MD parameters uniquely define the coefficients of the CDF equation (4.2), which then serves as an updated predictive tool. Equation (4.2) has an analytical solution F_u although, in general, numerical minimization in (3.6) needs to be employed to compute its approximation F_{NN}.

Figure 1 exhibits the prior and posterior distributions for u_0 (those for u_b behave similarly) computed with the alternative data assimilation strategies. Figure 1*a* represents these distributions as coordinates (μ_0, σ_0) on the statistical manifold of Gaussian distributions, whereas figure 1*b* shows them as PDFs in the value space Ω_{U_0}. The Bayesian update and the DA-MD approach yield almost identical results after assimilation of the same set of measurements, sharpening the distribution of the parameters around the true value.

Similar to figure 1*a*, the prior and posterior CDFs of the state variable $u(x,t)$, both obeying the CDF equation (4.2), are represented as points on the statistical manifold $\mathcal{F}$ with coordinates $(x, t, \boldsymbol{\varphi}^{(0)})$ and $(x, t, \boldsymbol{\varphi}^{(N_{\text{meas}})})$, respectively. The amount of information used during the analysis and transferred from the measurements to the conditional predictions can be thought of as the distance between these points: the information gain from prior to posterior is quantified by the KL divergence between these distributions (§3b). For the same prior and the same observations, DA-MD and the Bayesian update yield almost identical KL discrepancies. Moreover, D_{KL} does not vary within the assimilation regions, i.e. it remains constant in the regions of the space–time domain where F_u depends on either $\boldsymbol{\varphi}_0$ or $\boldsymbol{\varphi}_b$. The KL divergence also allows one to compare the informational gain from different sets of observations: doubling the number of measurements from $N_{\text{meas}} = 20$ to $N_{\text{meas}} = 40$ yields, in the assimilation regions informed by either the initial or the boundary conditions, a gain in KL terms of 7% and 9%, respectively.

(b) Uncertain reaction rate

In the following two test cases, we treat the uncertain coefficient k in (4.1) first as a random constant and then as a random field. The auxiliary variables u_0 and u_b in (4.1*b*) are taken to be

deterministic, so that the CDF equation (2.3) is subject to initial and boundary conditions

$$F_u(U;x,0)=\mathcal{H}(U-u_0) \quad \text{and} \quad F_u(U;b,t)=\mathcal{H}(U-u_\mathrm{b}-s(t)).$$

(i) Random variable

The coefficients (2.4) in the CDF equation (2.3) take the form (appendix A)

$$\boldsymbol{\mathcal{Q}}=\begin{pmatrix} 1 \\ -\langle k\rangle U-\dfrac{\sigma_k^2 U}{\langle k\rangle}[1-\mathrm{e}^{\langle k\rangle t^*}] \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mathcal{D}}=\begin{pmatrix} 0 & 0 \\ 0 & -\dfrac{\sigma_k^2 U^2}{\langle k\rangle}[1-\mathrm{e}^{\langle k\rangle t^*}] \end{pmatrix}, \tag{4.4}$$

where $t^*(U,x,t)=\min\{t,x,\langle k\rangle^{-1}\ln(U_\mathrm{max}/U)\}$, and $\langle k\rangle$ and σ_k are the ensemble mean and standard deviation of k, respectively. The coordinates of the dynamic manifold $\mathcal{F}$ of the approximated CDFs F_u are $\tilde{\boldsymbol{\varphi}}=\{x,t,\langle k\rangle,\sigma_k\}$. The CDF equation is solved via finite volumes (FV) using the `Fipy` solver [38], setting the discretization elements to $\Delta t=0.01$, $\Delta x=L/200$ and $\Delta U=(U_\mathrm{max}-U_\mathrm{min})/128$, with domain size defined by $L=1$, $U_\mathrm{min}=0$ and $U_\mathrm{max}=1$.

Minimization of (3.6) is done using the L-BFGS-B method implemented in `TensorFlow` [32] with a convergence threshold for the loss function value of 10^{-3}. The solution of the CDF equation (2.3), whose coefficients are given by (4.4), is represented by a fully connected NN with fixed architecture (9 layers, 20 nodes per hidden layer) and a sigmoidal activation function (hyperbolic tangent). Weights and biases of the NN are initialized at the beginning of the sequential procedure by approximating a solution of the CDF equation with prior statistical parameters $\boldsymbol{\varphi}^{(0)}$. Successive iterations are initialized with weights and biases from the previous step. This procedure considerably accelerates the identification of the target parameters. Zero residuals are enforced at $N_\mathrm{res}=792$ locations within the space–time domain, whereas initial and boundary conditions are imposed at $N_\mathrm{aux}=406$ locations. Furthermore, we enforce non-negativity of σ_k.

Remark 4.1. The FV approximation is used to construct the observational CDFs, whereas the NN approximation is used on a sparse set of points for numerical gradient-based minimization. The NN surrogate solution of the CDF equation (2.3) could also be used as a prior for the next assimilation step, with the advantage of being virtually free of artificial diffusion and with no theoretical limitation on the number of dimensions. This is not exploited further in this work, as research on the use of physics-informed NN to solve PDEs is not yet mature. Nevertheless, it has been shown to yield accurate identification of PDE parameters [31] and to reproduce qualitatively actual PDE solutions.

We compare the DA-MD estimate of the PDF of the model parameter k with the Bayesian posterior PDF of k. The latter is obtained analytically by assuming a Gaussian prior $f_k(K)$ and taking advantage of the analytical solution of (4.1):

$$\begin{aligned}\hat{f}_k(K|\mathbf{d}_{1:N_\mathrm{meas}}) &\propto f_L(\mathbf{d}_{1:N_\mathrm{meas}}|\mathbf{u}[(x,t)_{1:N_\mathrm{meas}};K])f_k(K)\\ &\approx \prod_{m=1}^{N_\mathrm{meas}} f_L(d_m|u\,[(x,t)_m;K])f_k(K).\end{aligned} \tag{4.5}$$

The Bayesian and DA-MD posterior (and prior) PDFs of the random reaction rate k are presented in figure 2. Figure 2*b* shows these densities in the value space Ω_K of $f_k(K)$, whereas figure 2*a* represents the state distributions as points on the dynamic manifold $\mathcal{F}$. The Bayesian update is optimal and analytical. Its sole source of error stems from the calculation of the normalization constant via numerical integration; as such it is treated as a benchmark in this comparison. On the contrary, DA-MD is based on a series of approximations (closures for the CDF equation, FV and NN solutions of the CDF equation, numerical minimization of the loss function). Nevertheless, DA-MD yields an updated posterior which is close to the Bayesian one. The DA-MD posterior is sharper than the Bayesian posterior; this might be due to the effect of numerical diffusion that artificially smears the CDF profiles computed as a solution of the CDF equation. Convergence of

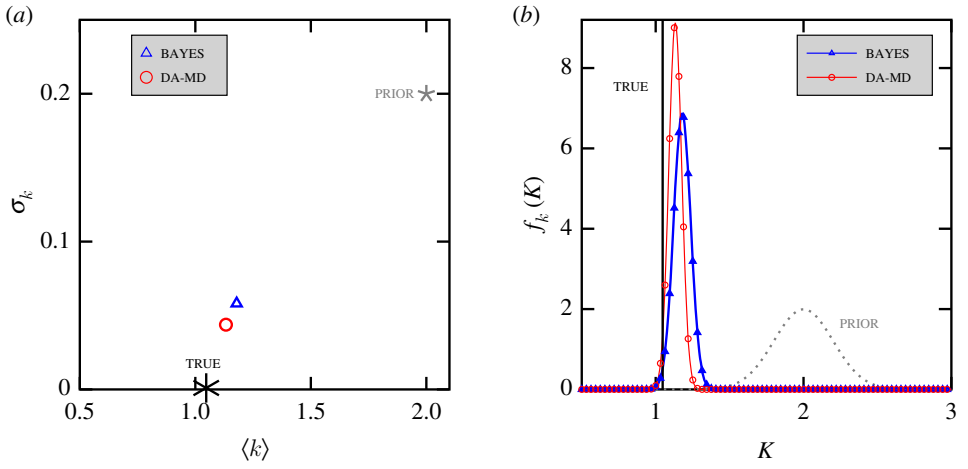

Figure 2. Prior and posterior PDFs of the random variable k shown (*a*) on the statistical manifold defined by coordinates $\{\langle k \rangle, \sigma_k\}$ representing the mean and standard deviation of k, and (*b*) in the value space Ω_K. The black asterisk in (*a*) and the black vertical line in (*b*) represent the true value ($k^{(\text{true})} = 1.047$), for which a Gaussian PDF degenerates into the Dirac distribution (delta function). The grey star (*a*) and the grey dashed line (*b*) represent a prior distribution ($\langle k \rangle^{\text{prior}} = 2$, $\sigma_k^{\text{prior}} = 0.2$). The blue triangle (*a*) and line (*b*) identify the Bayesian solution, whereas the corresponding red circles and lines identify the DA-MD solution. Parameters are set to $u_0 = 0.4$, $u_b = 0.5$, $\sigma_\varepsilon = 0.02$, $N_{\text{meas}} = 20$. (Online version in colour.)

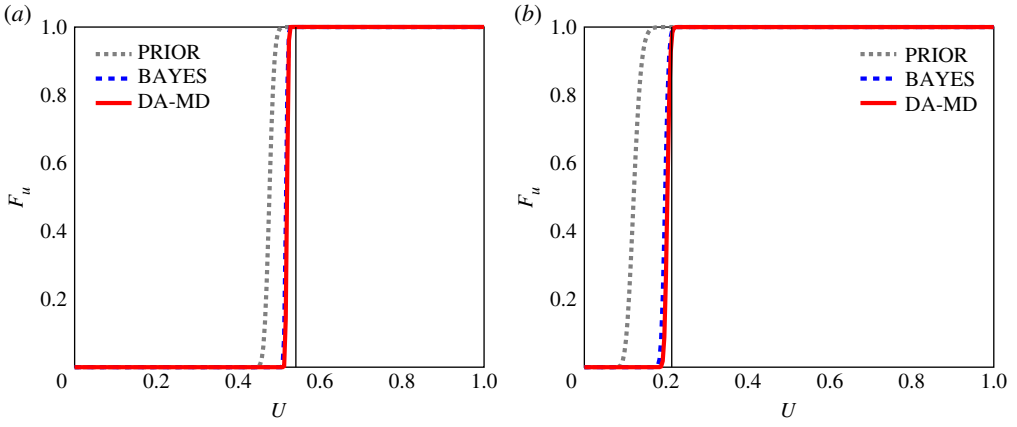

Figure 3. Prior and posterior CDF profiles $F_u(U)$ for the random k case at the final assimilation time t_M at two spatial locations: $x = 0.1$ (*a*) and $x = 0.8$ (*b*). The vertical black line represents the true solution, for which the CDF degenerates into a Heaviside step function. The dotted grey line represents the prior distribution with parameters $(\langle k \rangle, \sigma_k) = \{2, 0.2\}$; the dashed blue line and the solid red line represent the posterior distribution computed with updated Bayesian and DA-MD parameters, respectively. The updated parameters $\langle k \rangle$ and σ_k are those represented in figure 2. The remaining parameters are set to $k^{(\text{true})} = 1.047$, $u_0 = 0.4$, $u_b = 0.5$, $\sigma_\varepsilon = 0.02$ and $N_{\text{meas}} = 20$, $t_M = 0.6$. (Online version in colour.)

the DA-MD is slow, but its computational time is not expected to scale with the dimensionality of the problem (e.g. when dealing with random parameter fields). This flexibility represents a major advantage of the proposed procedure versus Bayesian inference, and it is explored in a more challenging scenario in the following section.

The prior and posterior CDFs of u, $F_u(U;\cdot)$, at the final assimilation time t_M are shown in figure 3. The posterior CDFs, for both the Bayesian and DA-MD assimilation, provide a state prediction that is closer than the prior CDF to its true value thanks to a more accurate parameter

identification (shown in figure 2). The value of measurements is evaluated in terms of their impact on the shape of the CDF at the measurement locations, and quantified by the KL divergence from the posterior to the prior. In this example, all locations exhibit the same information gain quantified by the KL divergence going from the posterior to the prior. That is because of the analytical one-to-one relation between k and $u(\mathbf{x}, t)$.

(ii) Random field

Keeping all other conditions and settings unchanged, we now consider a spatially varying uncertain parameter $k(x)$. It is treated as a second-order stationary (statistically homogeneous) multivariate Gaussian random field with constant mean $\langle k\rangle^{(\mathrm{true})}$ and standard deviation $\sigma_k^{(\mathrm{true})}$; its two-point autocovariance function $C_k^{(\mathrm{true})}(x-x')$ has either zero correlation length (i.e. uncorrelated random field or white noise),

$$C_k^{(\mathrm{true})}(x-x') = \sigma_k^2 \delta(x-x'),$$

or a finite correlation length $\lambda_k^{(\mathrm{true})}$,

$$C_k^{(\mathrm{true})}(x-x') = \sigma_k^2 \exp\left(-|x-x'|/\lambda_k^{(\mathrm{true})}\right).$$

One realization with the chosen statistical parameters represents the reference random field $k^{(\mathrm{true})}(x)$, which was used to construct synthetic observations via the FV solution of (2.1) with (4.1).

The coefficients (2.4) in the CDF equation (2.3) now take the form (appendix A)

$$\mathcal{Q} = \begin{pmatrix} 1 \\ -\langle k\rangle U - \dfrac{\sigma_k^2 U}{2} \end{pmatrix} \quad \text{and} \quad \mathcal{D} = \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{\sigma_k^2 U^2}{2} \end{pmatrix}, \tag{4.6}$$

if $k(x)$ is white noise, or

$$\mathcal{Q} = \begin{pmatrix} 1 \\ -\langle k\rangle U - \dfrac{\sigma_k^2 U}{\alpha}[\mathrm{e}^{\alpha t^*} - 1] \end{pmatrix} \quad \text{and} \quad \mathcal{D} = \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{\sigma_k^2 U^2}{\alpha}[\mathrm{e}^{\alpha t^*} - 1] \end{pmatrix} \tag{4.7}$$

with $\alpha = \langle k\rangle - 1/\lambda_k$ and $t^*(U, x, t) = \min\{t, x, \langle k\rangle^{-1} \ln(U_{\max}/U)\}$, if $k(x)$ has the exponential correlation C_k. The corresponding dynamic manifolds have either the coordinates $\tilde{\boldsymbol{\varphi}} = \{x, t, \langle k\rangle, \sigma_k\}$ or the coordinates $\tilde{\boldsymbol{\varphi}} = \{x, t, \langle k\rangle, \sigma_k, \lambda_k\}$, respectively.

Unlike Bayesian update, which identifies the k values at each spatial location with a consequent dramatic increase of the dimensionality of the target joint posterior PDF, DA-MD focuses on a finite set of parameters $\boldsymbol{\varphi}$ (the mean $\langle k\rangle$, the standard deviation σ_k and, in the correlated case, the correlation length λ_k); its computational cost is comparable to that for the constant random parameter case. We compare the updated DA-MD parameters with an approximation of the Bayesian posterior, since the number of random parameters and the nonlinearity of the problem prevent analytical treatment.

We employ a standard EnKF [9,10] for the update of the random field $k(x)$, discretized into N_x point values, with ensemble size N_{ens}. EnKF relies on multiple solutions of the physical model, which typically require special numerical treatment because of the high spatio-temporal variability of the model parameters. The choice of a spatial resolution poses another difficulty because the correlation length of the target random field is *a priori* unknown. This increases the numerical complexity of EnKF, to the advantage of MD. To focus on the data assimilation aspect of the problem, we solve both the physical model and the CDF equation on the same grid and with the same FV numerical solver, thus not taking advantage of MD's lower numerical complexity. Update is done sequentially for DA-MD, and recursively for EnKF [39, and references therein], i.e. at each assimilation step the ensemble members are forecast from the initial time to the current assimilation time. In both cases (EnKF and DA-MD), we assimilate the same N_{meas} measurements collected at two spatial locations, $x = 0.1$ and $x = 0.8$, in ten separate temporal instances, $t = \{0.15, 0.2, \ldots, 0.6\}$.

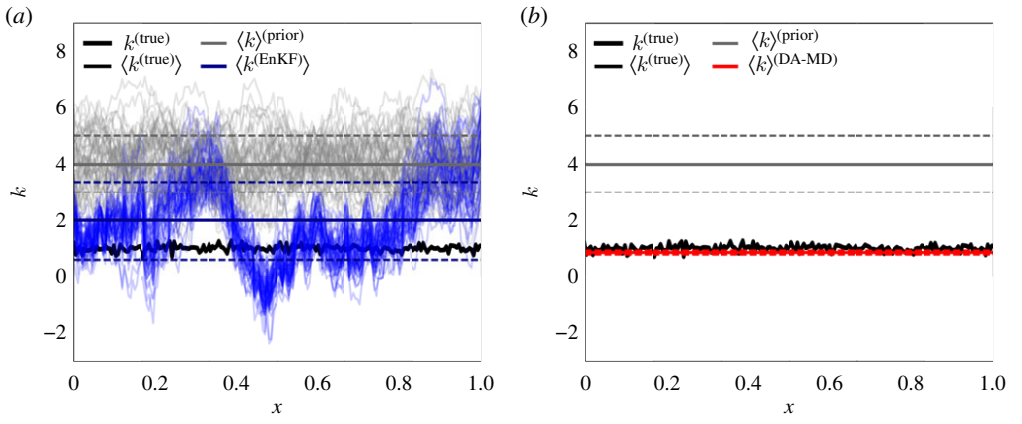

Figure 4. Parameter identification for the uncorrelated $k(x)$ field via EnKF (*a*) and DA-MD (*b*). Both panels contain the true field, $k^{(\mathrm{true})}(x)$, in black, and the prior field moments (grey lines). Both the prior and posterior random fields are defined by their mean value $\langle k\rangle$ (solid line), and a buffer region with half-width equal to the standard deviation (dashed lines). For the EnKF (*a*), both the prior and posterior ensemble members are represented. Posterior values are $\langle k\rangle^{(\mathrm{DA\text{-}MD})} = 0.86$, $\sigma_k^{(\mathrm{DA\text{-}MD})} = 0.07$, $\langle \bar{k}^{(\mathrm{EnKF})}\rangle = 1.7$, $\bar{\sigma}_k^{(\mathrm{EnKF})} = 1.19$, $\bar{k}^{(\mathrm{true})} = 1.01$, $\bar{\sigma}_k^{(\mathrm{true})} = 0.1$, where $\bar{\cdot}$ represents the spatial average. Parameters are set to $u_0 = 0.4$, $u_b = 0.5$, $N_{\mathrm{ens}} = 50$, $N_{\mathrm{meas}} = 20$, $\sigma_\varepsilon = 0.02$, $N_x = 200$, $\boldsymbol{\varphi}^{(0)} = \{\langle k\rangle, \sigma_k\}^{(\mathrm{prior})} = (4, 1)$. (Online version in colour.)

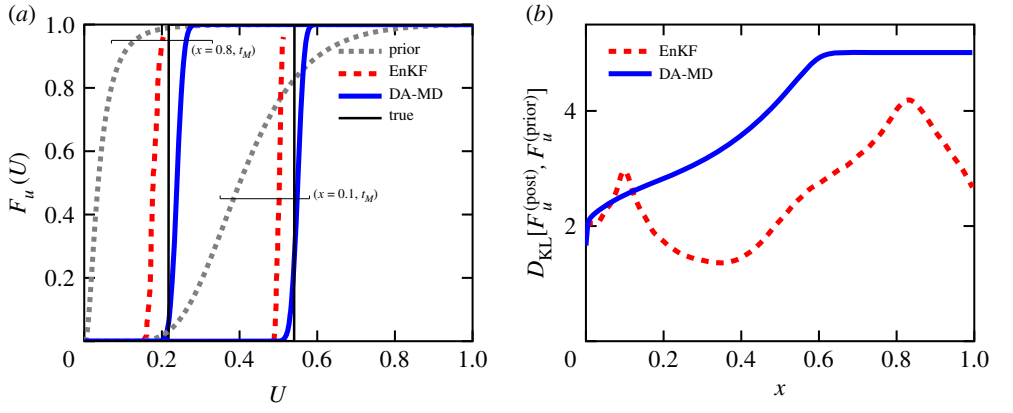

Figure 5. Prior and posterior CDFs (*a*) and corresponding KL divergence D_{KL} (*b*) for the uncorrelated $k(x)$ field. The CDF profiles (*a*) are computed at $(x, t) = (0.1, t_M)$ and $(x, t) = (0.8, t_M)$ as a solution of the CDF equation with prior $\boldsymbol{\varphi}^{(0)}$ and posterior $\boldsymbol{\varphi}^{(N_{\mathrm{meas}})}$ parameters (dotted grey and solid blue lines, respectively). The CDFs from EnKF (dashed red line) are computed as an empirical distribution of the ensemble members. The true solution is plotted as a Heaviside function centred on the true value $u^{(\mathrm{true})}(x, t)$ (black thin line), $\mathcal{H}(U - u^{(\mathrm{true})}(x, t))$. The selected coordinates for the profiles ($x = 0.1$ and $x = 0.8$) correspond to measurement locations. For both DA-MD and EnKF, the KL divergence D_{KL} between the posterior distribution and the prior distribution is computed as a function of x at time t_M. Parameters are set to $u_0 = 0.4$, $u_b = 0.5$, $N_{\mathrm{ens}} = 50$, $N_{\mathrm{meas}} = 20$, $\sigma_\varepsilon = 0.02$, $N_x = 200$, $\Delta x = 1.6 \times 10^{-3}$, $\Delta U = 8.3 \times 10^{-4}$, $\Delta t = 10^{-3}$, $\boldsymbol{\varphi}^{(0)} = \{\langle k\rangle, \sigma_k\}^{(\mathrm{prior})} = \{4, 1\}$, $t_M = 0.6$. (Online version in colour.)

Figures 4 and 6 exhibit the EnKF and DA-MD posterior random fields for the uncorrelated and correlated cases, respectively. When the true field $k(x)^{(\mathrm{true})}$ is white noise, DA-MD accurately identifies the updated mean $\langle k\rangle^{(\mathrm{DA\text{-}MD})}$, but underestimates the value of $\sigma_k^{(\mathrm{DA\text{-}MD})}$. The latter might be due to the impact of artificial diffusion on the solution of the CDF equation used as a prior in the DA-MD procedure. EnKF yields a wider posterior estimate for k, with spatial averages for the mean $\langle k\rangle^{(\mathrm{EnKF})}$ and the standard deviation $\sigma_k^{(\mathrm{EnKF})}$ that are further away from the spatial averages of the moments of the true field (values in the caption).

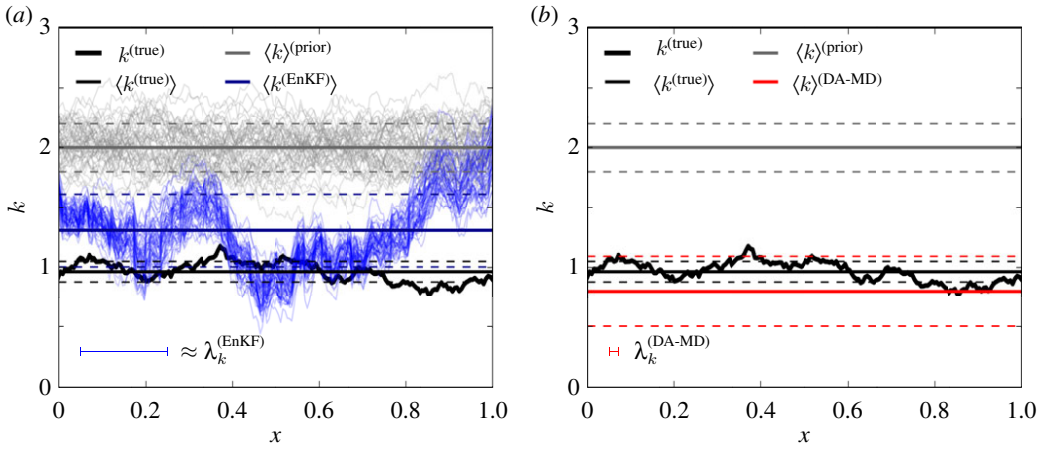

Figure 6. Parameter identification for the correlated field $k(x)$ via EnKF (*a*) and DA-MD (*b*). Both panels contain the true field, $k^{(\mathrm{true})}(x)$, in black, and the prior field (grey lines). Both the prior and posterior fields are defined by their mean value $\langle k\rangle$ (solid line), and a buffer region with half-width equal to the standard deviation value (dashed lines). An estimate of the posterior correlation length is in the bottom left corner of both panels. For EnKF (*a*), both the prior and posterior ensemble members are also represented. Posterior values are $\langle k\rangle^{(\mathrm{DA\text{-}MD})} = 0.80$, $\sigma_k^{(\mathrm{DA\text{-}MD})} = 0.30$, $\lambda_k^{(\mathrm{DA\text{-}MD})} = 0.013$, $\langle \bar{k}^{(\mathrm{EnKF})}\rangle = 1.23$, $\bar{\sigma}_k^{(\mathrm{EnKF})} = 0.33$, $\bar{k}^{(\mathrm{true})} = 0.96$, $\bar{\sigma}_k^{(\mathrm{true})} = 0.09$, $\lambda_k^{\mathrm{true}} = 0.3$, where $\bar{\cdot}$ represents the spatial average. Parameters are set to $u_0 = 0.4$, $u_\mathrm{b} = 0.5$, $N_\mathrm{ens} = 50$, $N_\mathrm{meas} = 20$, $\sigma_\varepsilon = 0.01$, $N_x = 200$, $\boldsymbol{\varphi}^{(0)} = \{\langle k\rangle, \sigma_k, \lambda_k\}^{(\mathrm{prior})} = \{2, 0.2, 0.2\}$. (Online version in colour.)

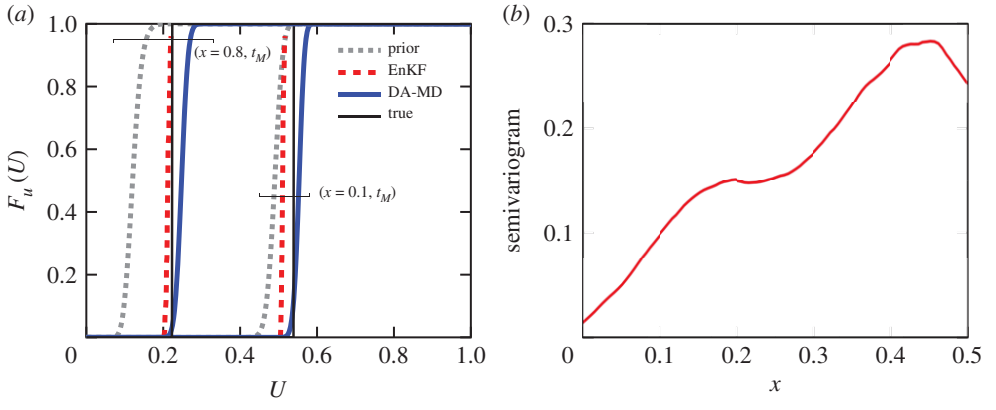

Figure 7. (*a*) Prior and posterior CDFs of the correlated field $k(s)$. The CDF profiles are computed at $(x, t) = (0.1, t_M)$ and at $(x, t) = (0.8, t_M)$ as a solution of the CDF equation with prior $\boldsymbol{\varphi}^{(0)}$ and posterior $\boldsymbol{\varphi}^{(N_\mathrm{meas})}$ parameters (dotted grey and solid blue lines, respectively). The CDFs from EnKF (dashed red line) are computed as an empirical distribution of the ensemble members. The true solution is plotted as a Heaviside function centred on the true value $u^{(\mathrm{true})}(x, t)$ (black thin line), $\mathcal{H}(U - u^{(\mathrm{true})}(x, t))$. (*b*) Semivariogram for the EnKF posterior ensemble members. Parameters are set to $u_0 = 0.4$, $u_\mathrm{b} = 0.5$, $N_\mathrm{ens} = 50$, $N_\mathrm{meas} = 20$, $\sigma_\varepsilon = 0.01$, $N_x = 200$, $\Delta U = 3.75 \times 10^{-4}$, $\Delta x = 1.6 \times 10^{-3}$, $\Delta U = 8.3 \times 10^{-4}$, $\Delta t = 10^{-3}$, $\boldsymbol{\varphi}^{(0)} = \{\langle k\rangle, \sigma_k, \lambda_k\}^{(\mathrm{prior})} = (2, 0.2, 0.2)$, $t_M = 0.6$. (Online version in colour.)

Despite these differences in reconstruction of the statistical properties of the posterior $k(x)$, both assimilation techniques yield a posterior prediction of $F_u(U)$ that approaches the true value of the solution (figure 5*a*). The information gain from the measurements is quantified in terms of the KL divergence for both DA-MD and EnKF (figure 5*b*) at time t_M. MD densities (both the prior and the posterior) are calculated via finite differences from the solution of the CDF equation, whereas EnKF densities are computed via kernel density estimation with Gaussian kernel and

Scott's bandwidth, using the ensemble members as data points. Our results suggest that DA-MD extracts more information than EnKF from the same set of measurements in the current configuration at almost all values of x, as is also reflected in an accurate characterization of the posterior k field. Observations collected at $x > t_M$ (the region where characteristic lines originate from the initial conditions) are more informative for DA-MD assimilation. The KL divergence for EnKF highlights the locations of more informative measurements, displaying two distinctive peaks.

Figure 6 exhibits the results of a similar analysis for the correlated field $k(x)^{(\text{true})}$. DA-MD posterior estimates of the mean and standard deviation of k are closer to the averaged statistical properties of the true field than EnKF estimates are (values are in the figure caption). DA-MD underestimates the spatial correlation length λ_k, whereas the identification of λ_k via EnKF is inconclusive as the semivariogram for $k(x)$ does not develop a sill. We identify an intermediate plateau and assume the corresponding lag value as the updated correlation length for the field. The semivariogram is computed using the posterior ensemble member values, and is shown in figure 7*b*. The corresponding state CDFs $F_u(U; x, t)$ are plotted in figure 7*a* in two representative sections that correspond to measurement locations. Both DA-MD and EnKF yield a posterior state CDF F_u considerably closer to the true value than the prior distribution.

5. Summary and future work

We proposed a novel methodology for parameter estimation that leverages the MD for both the forecast and analysis steps. Reduction of uncertainty in model parameters is recast into a problem of identification of closure parameters for the CDF equation, expressing the space–time evolution of uncertainty for the model output. Specifically, we identify the parameters in the CDF equation (2.3) that yield an estimate in the measurement locations as close as possible to the state distribution. This is expressed by an observational Bayesian posterior in that specific location, which is obtained by combining the data model and the physically based prior. The procedure is done sequentially, progressively updating the parameters of the CDF equation as more measurements are assimilated. We demonstrated that our method reproduces Bayesian posteriors in scenarios where Bayesian inference can be performed analytically, and ameliorates parameter identification when compared to EnKF (as an approximation of Bayesian update) in cases where Bayesian inference is elusive.

This work opens multiple possible research a venues. In particular, we plan to (i) explore the construction of novel data-driven closure approximations for MD; (ii) investigate the use of novel ML techniques for more efficient optimization and/or solution of PDEs; and (iii) introduce multi-point statistics.

Data accessibility. There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

Authors' contributions. F.B. conceived the study, developed the analysis, carried out the simulations and drafted the manuscript; D.M.T. participated in the design of the study, participated in the analysis and critically revised the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This work was supported in part by Air Force Office of Scientific Research under award no. FA9550-18-1-0474, and by a gift from TOTAL.

Appendix A. Derivation of cumulative distribution function equations

MD commences by defining a so-called raw CDF $\pi(U; \mathbf{x}, t) \equiv \mathcal{H}(U - u(\mathbf{x}, t))$, where $\mathcal{H}(\cdot)$ is the Heaviside function. Let $f_u(U; \mathbf{x}, t)$ denote the single-point PDF of $u(\mathbf{x}, t)$. Then it follows from the

definition of the ensemble mean $\mathbb{E}[\cdot] \equiv \langle \cdot \rangle$ that

$$\mathbb{E}[\pi(U;\mathbf{x},t)] = \int_{U_{\min}}^{U_{\max}} \mathcal{H}(U-\mathcal{U}) f_u(\mathcal{U};\mathbf{x},t)\mathrm{d}\mathcal{U} = \int_{U_{\min}}^{U} \mathcal{H}(U-\mathcal{U}) f_u(\mathcal{U};\mathbf{x},t)\mathrm{d}\mathcal{U}$$
$$= F_u(U;\mathbf{x},t). \tag{A 1}$$

Other useful properties of π are

$$\frac{\partial \pi}{\partial t} = \frac{\partial \pi}{\partial u}\frac{\partial u}{\partial t} = -\frac{\partial \pi}{\partial U}\frac{\partial u}{\partial t} \quad \text{and} \quad \nabla \pi = -\frac{\partial \pi}{\partial U}\nabla u. \tag{A 2}$$

Accounting for these properties, multiplication of (2.1) by $-\partial_U \pi$ yields

$$\frac{\partial \pi}{\partial t} + \dot{\mathbf{q}}(U)\cdot\nabla\pi + r(U)\frac{\partial \pi}{\partial U} = 0, \tag{A 3}$$

where $\dot{\mathbf{q}} = \mathrm{d}\mathbf{q}(U)/\mathrm{d}U$. This equation is exact as long as solutions of (2.1), $u(\mathbf{x},t)$, are smooth (do not develop shocks) for each realization of random parameters $\tilde{\boldsymbol{\theta}}$. It is subject to initial and boundary conditions derived from the initial and boundary conditions of the physical problem, and to $\pi(U = U_{\min};\mathbf{x},t) = 0$ and $\pi(U = U_{\max};\mathbf{x},t) = 1$.

In the absence of uncertainty, (A 3) is deterministic and equivalent to (2.1); the model output $u(\mathbf{x},t)$ can be recovered from $\pi(U,\mathbf{x},t)$ by integration. In the presence of uncertainty affecting the parameters and the auxiliary inputs, it follows from (A 1) that the ensemble average of (A 3) is

$$\frac{\partial F_u}{\partial t} + \langle \dot{\mathbf{q}}(U;\boldsymbol{\theta}_q)\cdot\nabla\pi\rangle + \left\langle r(U;\boldsymbol{\theta}_r)\frac{\partial \pi}{\partial U}\right\rangle. \tag{A 4}$$

If the model parameters $\boldsymbol{\theta}$ are deterministic, then so is the evolution dynamics, and uncertainty in predictions of $u(\mathbf{x},t)$ is solely due to uncertainty in the initial and the boundary conditions. In that case, (A 4) gives an exact CDF equation,

$$\frac{\partial F_u}{\partial t} + \dot{\mathbf{q}}(U;\boldsymbol{\theta}_q)\cdot\nabla F_u + r(U;\boldsymbol{\theta}_r)\frac{\partial F_u}{\partial U}. \tag{A 5}$$

Otherwise, closure approximations are necessary to obtain a workable expression for the undefined terms in (A 4). These expressions depend on the closure strategy and on the functional form of $\mathbf{q}$ and r.

To be specific, we set $\mathbf{q}(u) = \mathbf{v}(\mathbf{x})u$ and $r(u) = kr_\alpha(u;\alpha,u_{\mathrm{eq}}) = k\alpha(u_{\mathrm{eq}}^\alpha - u^\alpha)$. Here, $\mathbf{v}(\mathbf{x})$ is the divergence-free velocity, $\nabla\cdot\mathbf{v} = 0$, of steady incompressible flow; and $\alpha \in \mathbb{N}^+$ is the order of an equilibrium reaction with reaction rate $k(\mathbf{x})$, which drives the system towards its equilibrium state u_{eq}. An analogous system was studied in detail in [2,6]. In what follows, we summarize the closure approximations developed in these works for the case of deterministic $\mathbf{v}(\mathbf{x})$ and random $k(\mathbf{x})$.

We use the Reynold decomposition to represent random quantities as the sum of their respective means and zero-mean fluctuations around these means,

$$k = \langle k\rangle + k' \quad \text{and} \quad \pi = \mathrm{F} + \pi'. \tag{A 6}$$

A first-order (in the variance σ_k^2 of stationary random fluctuations k') approximation of (A 4) takes the form of (2.3) with the coefficients [2]

$$\left.\begin{aligned} &\mathcal{Q}_i = v_i(\mathbf{x}), \quad i = 1,\dots,d,\\ &\mathcal{Q}_{d+1} \approx \langle k\rangle r_\alpha(U) + \int_0^t\int_{\tilde{\Omega}} G(\mathbf{x},U,\mathbf{y},V,t-\tau)C_k(\mathbf{x},\mathbf{y})\frac{\mathrm{d}r_\alpha(U)}{\mathrm{d}U}\mathrm{d}\mathbf{y}\mathrm{d}V\mathrm{d}\tau\\ \text{and}\quad &\mathcal{D}_{ij} \approx \delta_{i,d+1}\delta_{j,d+1}r_\alpha(U)\int_0^t\int_{\tilde{\Omega}} G(\mathbf{x},U,\mathbf{y},V,t-\tau)C_k(\mathbf{x},\mathbf{y})r_\alpha(V)\mathrm{d}\mathbf{y}\mathrm{d}V\mathrm{d}\tau, \quad i,j = 1,\dots,d+1. \end{aligned}\right\} \tag{A 7}$$

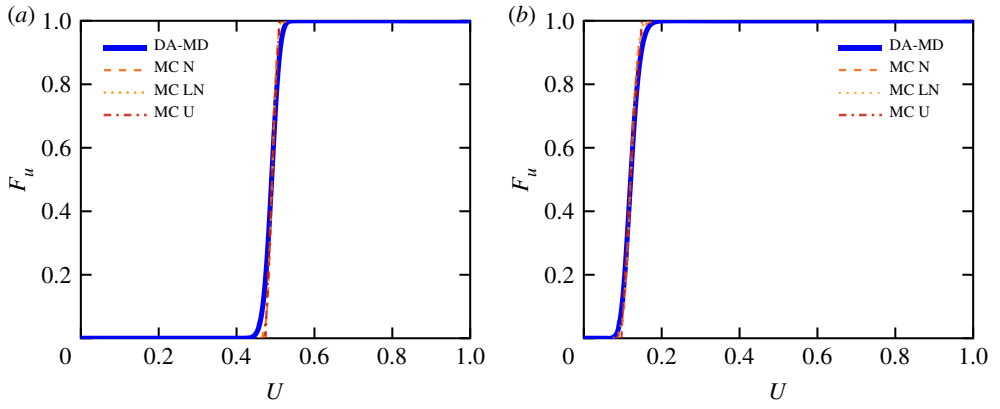

Figure 8. Comparison between the FV approximation of the prior CDF and its MC counterpart for the random k scenario. Both techniques use the same mean and variance for k, $\langle k\rangle = 2$, $\sigma_k = 0.2$. MCS are repeated for different distributions of k sharing the same mean and variance: normal, lognormal and uniform distributions, respectively. Parameters are set to: $N_{\mathrm{MC}} = 1000$, $\Delta t = 0.001$, $\Delta x = 1.6 \times 10^{-4}$ and $\Delta U = 8.3 \times 10^{-4}$. (Online version in colour.)

Here, $\delta_{i,d+1}$ is the Kronecker delta, $C_k(\mathbf{x},\mathbf{y}) = \langle k'(\mathbf{x}')k'(\mathbf{x})\rangle$ is the covariance function of $k'(\mathbf{x})$, and $G(\mathbf{x}, U, \mathbf{y}, V, t-\tau)$ is the mean-field Green's function that is defined as a solution of

$$\frac{\partial G}{\partial \tau} + \mathbf{v}\cdot\nabla' G + \langle k\rangle \frac{\mathrm{d}r_\alpha G}{\mathrm{d}U} = -\delta(\mathbf{x}-\mathbf{y})\delta(U-V)\delta(t-\tau), \quad \tau < t \tag{A 8}$$

with homogeneous initial (at $\tau = 0$) and boundary conditions on $\partial\tilde{\Omega}$. The closure approximations are thus expressed in terms of the mean and two-point covariance of the random input $k(\mathbf{x})$.

The derivation of (2.3) and (A 7) is based on the following assumptions: ∇F varies slowly in space and time to justify the use of a local model, the random inputs are mutually uncorrelated, and the variance σ_k^2 is sufficiently small to warrant its use as a perturbation parameter.

Our numerical experiments consider one-dimensional ($d = 1$) advection in a deterministic velocity field with $v = 1$ and linear reaction ($\alpha = 1$) with second-order stationary reaction rate $k(\mathbf{x})$ with constant mean $\langle k\rangle$ and variance σ_k^2 and covariance function $C_k(x-y)$. The flow takes place in the semi-infinite domain Ω, so that $\tilde{\Omega} = [0,\infty)\times[U_{\min}, U_{\max}]$. The deterministic equilibrium state is set to $u_{\mathrm{eq}} = 0$. Under these conditions, (A 7) reduces to

$$\left.\begin{aligned} &\mathcal{D}_{11} = 0, \quad \mathcal{D}_{12} = \mathcal{D}_{21} = 0, \quad \mathcal{D}_{22} = U^2 \int_0^{t^*} \mathrm{e}^{\langle k\rangle\tau} C_k(v\tau)\mathrm{d}\tau \\ &\mathcal{Q}_1 = v, \quad \mathcal{Q}_2(\mathbf{x}, U, t) = -U\langle k\rangle + U\int_0^{t^*} \mathrm{e}^{\langle k\rangle\tau} C_k(v\tau)\mathrm{d}\tau, \end{aligned}\right\} \tag{A 9}$$

where $t^* = \min\{t, \langle k\rangle^{-1}\log(U_{\max}/U), x/v\}$. We consider three models of spatial correlation of $k(\mathbf{x})$. The first takes $k(\mathbf{x})$ to be perfectly correlated, so that $C_k(x-y) = \sigma_k^2$; then (A 9) simplifies to (4.4). The second considers the opposite case, i.e. uncorrelated random field with $C_k(x-y) = \sigma_k^2\delta(x-y)$, which yields (4.6). Finally, the third one deals with the exponential covariance function $C_k(x-y) = \sigma_k^2\exp(-|x-y|/\lambda_k)$, where λ_k is the correlation length of $k(\mathbf{x})$, with closure parameters (4.7).

The CDF equation (2.3), whose coefficients are defined by (A 9), depends only on the low moments of $k(\mathbf{x})$, i.e. on $\langle k\rangle$, σ_k^2 and C_k, rather than on its full PDF. We study the sensitivity of our closure to a choice of the functional form of the single-point PDF $f_k(K;\mathbf{x})$ of $k(\mathbf{x})$ for the perfectly correlated case. This is done by comparing a numerical (finite-volume) solution of (2.3) with the results of MCS. The latter consist of post-processing of $N_{\mathrm{MC}} = 1000$ analytical solutions of the physical model (4.1), whose parameters are drawn, alternatively, from the Gaussian, lognormal and uniform PDFs $f_k(K;\mathbf{x})$, with negligible discrepancy in CDF terms (figure 8). As uncertainty is reduced via data assimilation, the discrepancy between posteriors obtained with different

assumed PDF forms of k reduces, and the impact of closure approximations on the CDF equation decreases.

References

1. Tartakovsky DM, Gremaud PA. 2015 Method of distributions for uncertainty quantification. In *Handbook of uncertainty quantification* (eds R Ghanem, D Higdon, H Owhadi), pp. 763–783. New York, NY: Springer.
2. Boso F, Broyda SV, Tartakovsky DM. 2014 Cumulative distribution function solutions of advection-reaction equations with uncertain parameters. *Proc. R. Soc. A* **470**, 20140189. (doi:10.1098/rspa.2014.0189)
3. Boso F, Tartakovsky DM. 2016 The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resour. Res.* **52**, 4700–4712. (doi:10.1002/2016WR018745)
4. Alawadhi AA, Boso F, Tartakovsky DM. 2018 Method of distributions for water-hammer equations with uncertain parameters. *Water Resour. Res.* **54**, 9398–9411. (doi:10.1029/2018WR023383)
5. Ghanem R, Red-Horse J. 2015 Polynomial chaos: modeling, estimation, and approximation. In *Handbook of uncertainty quantification* (eds R Ghanem, D Higdon, H Owhadi), pp. 1–31. New York, NY: Springer.
6. Venturi D, Tartakovsky DM, Tartakovsky AM, Karniadakis GE. 2013 Exact PDF equations and closure approximations for advective-reactive transport. *J. Comput. Phys.* **243**, 323–343. (doi:10.1016/j.jcp.2013.03.001)
7. Yang HJ, Boso F, Tchelepi HA, Tartakovsky DM. 2019 Probabilistic forecast of single-phase flow in porous media with uncertain properties. *Water Resour. Res.* **55**, 8631–8645. (doi:10.1029/2019WR026090)
8. Tartakovsky DM, Dentz M, Lichtner PC. 2009 Probability density functions for advective-reactive transport in porous media with uncertain reaction rates. *Water Resour. Res.* **45**, W07414. (doi:10.1029/2008WR007383)
9. Wikle CK, Berliner LM. 2007 A Bayesian tutorial for data assimilation. *Physica D* **230**, 1–16. (doi:10.1016/j.physd.2006.09.017)
10. Evensen G. 2009 *Data assimilation: the ensemble Kalman filter*. New York, NY: Springer.
11. Myung IJ. 2003 Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **47**, 90–100. (doi:10.1016/S0022-2496(02)00028-7)
12. Cousineau D, Helie S. 2013 Improving maximum likelihood estimation using prior probabilities: a tutorial on maximum a posteriori estimation and an examination of the Weibull distribution. *Tutor. Quant. Methods Psychol.* **9**, 61–71. (doi:10.20982/tqmp.09.2.p061)
13. Katzfuss M, Stroud JR, Wikle CK. 2016 Understanding the ensemble Kalman filter. *Am. Stat.* **70**, 350–357. (doi:10.1080/00031305.2016.1141709)
14. Brooks S, Gelman A, Jones G, Meng XL (eds). 2011 *Handbook of Markov chain Monte Carlo*. Boca Raton, FL: CRC Press.
15. Speekenbrink M. 2016 A tutorial on particle filters. *J. Math. Psychol.* **73**, 140–152. (doi:10.1016/j.jmp.2016.05.006)
16. Blei DM, Kucukelbir A, McAuliffe JD. 2017 Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877. (doi:10.1080/01621459.2017.1285773)
17. Herzog R, Kunisch K. 2010 Algorithms for PDE-constrained optimization. *GAMM-Mitt.* **33**, 163–176. (doi:10.1002/gamm.201010013)
18. Raissi M, Perdikaris P, Karniadakis GE. 2017 Machine learning of linear differential equations using Gaussian processes. *J. Comput. Phys.* **348**, 683–693. (doi:10.1016/j.jcp.2017.07.050)
19. Zhu Y, Zabaras N, Koutsourelakis PS, Perdikaris P. 2019 Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.* **394**, 56–81. (doi:10.1016/j.jcp.2019.05.024)
20. Zhang D, Lu L, Guo L, Karniadakis GE. 2019 Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *J. Comput. Phys.* **397**, 108850. (doi:10.1016/j.jcp.2019.07.048)
21. Wang P, Tartakovsky DM. 2013 CDF solutions of Buckley-Leverett equation with uncertain parameters. *Multiscale Model. Simul.* **11**, 118–133. (doi:10.1137/120865574)
22. Boso F, Tartakovsky DM. 2020 Data-informed method of distributions for hyperbolic conservation laws. *SIAM J. Sci. Comput.* **42**, A559–A583. (doi:10.1137/19M1260773)

23. Perthame B. 2002 *Kinetic formulation of conservation laws*, vol. 21. London, UK: Oxford University Press.
24. Cafaro C, Mancini S. 2011 Quantifying the complexity of geodesic paths on curved statistical manifolds through information geometric entropies and Jacobi fields. *Physica D* **240**, 607–618. (doi:10.1016/j.physd.2010.11.013)
25. Wang P, Tartakovsky DM. 2012 Uncertainty quantification in kinematic wave models. *J. Comput. Phys.* **231**, 7868–7880. (doi:10.1016/j.jcp.2012.07.030)
26. Giffin A, Caticha A. 2007 Updating probabilities with data and moments. *AIP Conf. Proc.* **954**, 74–84. (doi:10.1063/1.2821302)
27. Bellemare MG, Danihelka I, Dabney W, Mohamed S, Lakshminarayanan B, Hoyer S, Munos R. 2018 The Cramer distance as a solution to biased wasserstein gradients. (http://arxiv.org/abs/1705.10743)
28. Pinsker MS. 1964 *Information and information stability of random variables and processes*. San Francisco, CA: Holden-Day.
29. Topsoe F. 2000 Some inequalities for information divergence and related measures of discrimination. *IEEE Trans. Inf. Theory* **46**, 1602–1609. (doi:10.1109/18.850703)
30. Stein EM, Shakarchi R. 2011 *Functional analysis: introduction to further topics in analysis*, vol. 4. Princeton, NJ: Princeton University Press.
31. Raissi M, Perdikaris P, Karniadakis GE. 2019 Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707. (doi:10.1016/j.jcp.2018.10.045)
32. Abadi M *et al.* 2015 TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
33. Barron AR. 1993 Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Trans. Inf. Theory* **39**, 930–945. (doi:10.1109/18.256500)
34. Bölcskei H, Grohs P, Kutyniok G, Petersen P. 2019 Optimal approximation with sparsely connected deep neural networks. *SIAM J. Math. Data Sci.* **1**, 8–45. (doi:10.1137/18M118709X)
35. Amari SI. 2016 *Information geometry and its applications*. New York, NY: Springer.
36. Kullback S. 1997 *Information theory and statistics*. New York, NY: Wiley.
37. Cafaro C, Alsing PM. 2020 Information geometry aspects of minimum entropy production paths from quantum mechanical evolutions. *Phys. Rev. E* **101**, 022110. (doi:10.1103/PhysRevE.101.022110)
38. Guyer JE, Wheeler D, Warren JA. 2009 FiPy: partial differential equations with Python. *Comput. Sci. Eng.* **11**, 6–15. (doi:10.1109/MCSE.2009.52)
39. Crestani E, Camporese M, Baú D, Salandin P. 2013 Ensemble Kalman filter versus ensemble smoother for assessing hydraulic conductivity via tracer test data assimilation. *Hydrol. Earth Syst. Sci.* **17**, 1517–1531. (doi:10.5194/hess-17-1517-2013)