SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 8, No. 3, pp. 1036–1060

A Bayesian Approach to Estimating Background Flows from a Passive Scalar*

Jeff Borggaard†, Nathan Glatt-Holtz‡, and Justin Krometis§

Abstract. We consider the statistical inverse problem of estimating a background flow field (e.g., of air or water) from the partial and noisy observation of a passive scalar (e.g., the concentration of a solute), a common experimental approach to visualizing complex fluid flows. Here the unknown is a vector field that is specified by a large or infinite number of degrees of freedom. Since the inverse problem is ill-posed, i.e., there may be many or no background flows that match a given set of observations, we regularize it by laying out a functional analytic and Bayesian framework for approaching this problem. In doing so, we leverage substantial recent advances in statistical inference and adjoint methods for infinite-dimensional problems. We then identify interesting example problems that exhibit posterior measures with simple and complex structure. We use these examples to conduct a large-scale benchmark of Markov chain Monte Carlo methods developed in recent years for infinite-dimensional settings. Our results indicate that these methods are capable of resolving complex multimodal posteriors in high dimensions.

Key words. Bayesian statistical inversion, Markov chain Monte Carlo (MCMC), passive scalars, fluid turbulence

AMS subject classifications. 65N21, 62F15, 76R05

DOI. 10.1137/19M1267544

1. Introduction.

A common approach to investigating complex fluid flows is through measurement of a substance moving within the fluid. For example, dye, smoke, or neutrally buoyant particles are injected into fluids to visualize vortices or other structures in turbulent flows [34, 35, 58, 51]. In this work we consider the inverse problem of estimating a background fluid flow from partial, noisy observations of a dye, pollutant, or other solute advecting and

*Received by the editors June 10, 2019; accepted for publication (in revised form) May 26, 2020; published electronically August 11, 2020.

https://doi.org/10.1137/19M1267544

Funding: The work of the first author was supported in part by the National Science Foundation (NSF) under grants DMS-1522616 and DMS-1819110, and by the National Institute for Occupational Safety and Health under grant 200-2014-59669. The work of the second author was supported in part by the NSF under grants DMS-1313272 and DMS-1816551, and by the Simons Foundation under grant 515990. This work was also supported by the Mathematical Sciences Research Institute, the Tulane University Math Department, and the International Centre for Mathematical Sciences (ICMS), where significant portions of this project were developed and carried out. The Society for Industrial and Applied Mathematics (SIAM) and NSF grant DMS-1613965, NSF grant 1700124, and the Virginia Tech Interdisciplinary Center for Applied Math (ICAM), respectively, provided funds for travel to the 2016 Gene Golub SIAM Summer School, 2016 FOMICS Winter School on Uncertainty Quantification, and the 2017 ICMS workshop Probabilistic Perspectives in Nonlinear Partial Differential Equations, where the third author presented the preliminary results of this work and received valuable feedback.

†Department of Mathematics, Virginia Tech, Blacksburg, VA 24061 (jborggaard@vt.edu, https://www.math.vt.edu/people/jborggaa/).

‡Department of Mathematics, Tulane University, New Orleans, LA 70118 (negh@tulane.edu, http://www.math.tulane.edu/%7Enegh/).

§Advanced Research Computing, Virginia Tech, Blacksburg, VA 24061 (jkrometis@vt.edu, https://www.arc.vt.edu/justin-krometis/).

diffusing within the fluid. The initial condition is assumed to be known, so the problem can be interpreted as a controlled experiment, where a substance is added at known locations and then observed as the system evolves to investigate the structure of the underlying flow.

The physical model considered is the two-dimensional advection-diffusion equation on the periodic domain $\mathbb{T}^2 = [0,1]^2$:

$$\frac{\partial}{\partial t}\theta(t,\mathbf{x}) = -\mathbf{v}(\mathbf{x})\cdot\nabla\theta(t,\mathbf{x}) + \kappa\Delta\theta(t,\mathbf{x}), \quad \theta(0,\mathbf{x}) = \theta_0(\mathbf{x}). \tag{1.1}$$

Here, the following hold:

- $\theta : \mathbb{R}^+ \times \mathbb{T}^2 \to \mathbb{R}$ is a *passive scalar*, typically the concentration of some solute of interest, which is spread by diffusion and the motion of a (time-stationary) fluid flow $\mathbf{v}$. This solute is "passive" in that it does not affect the motion of the underlying fluid.
- $\mathbf{v} : \mathbb{T}^2 \to \mathbb{R}^2$ is an *incompressible background flow*, i.e., $\mathbf{v}$ is constant in time and satisfies $\nabla\cdot\mathbf{v} = 0$.
- $\kappa > 0$ is the *diffusion coefficient*, which models the rate at which local concentrations of the solute spread out within the solvent in the absence of advection.

We obtain finite observations $\mathcal{Y} \in Y$ (e.g., $\mathbb{R}^N$ or $\mathbb{C}^N$) subject to additive noise η, i.e.,

$$\mathcal{Y} = \mathcal{G}(\mathbf{v}) + \eta, \quad \eta \sim \gamma_0, \tag{1.2}$$

for some measure γ_0 related to the precision of the observations. Here, $\mathcal{G} : H \to Y$ is the *parameter-to-observable*, or forward, map. This $\mathcal{G}$ associates the background flow $\mathbf{v}$, sitting in a suitable function space H, with a finite collection of measurements (observables) of the resulting $\theta = \theta(\mathbf{v})$. The observations may take a number of forms, such as

- spatial-temporal point observations: $\mathcal{G}_j(\mathbf{v}) = \theta(t_j, \mathbf{x}_j, \mathbf{v})$ for $t_j \in [0,T]$ and $\mathbf{x}_j \in [0,1]^2$;
- spectral components: $\mathcal{G}_j(\mathbf{v}) = \langle\theta(t_j,\cdot,\mathbf{v}), \mathbf{e}_{\mathbf{k}_j}\rangle_{L^2(\mathbb{T}^2)}$ for some basis $\{\mathbf{e}_\mathbf{k}\}$ of the scalar field θ;
- local averages: $\mathcal{G}_j(\mathbf{v}) = \frac{1}{|\mathcal{D}_j|}\int_{\mathcal{D}_j}\theta(t,\mathbf{x},\mathbf{v})d\mathbf{x}\,dt$, for subdomains $\mathcal{D}_j \subset [0,T]\times[0,1]^2$, where $|\mathcal{D}_j|$ denotes the volume of $\mathcal{D}_j$.
- other physical quantities of interest from θ, such as variance, dissipation rate, or structure functions.

This work will focus on point observations as the most obvious practical implementation. However, we note that the methodology outlined in this manuscript is quite general. Moreover, while we have assumed a divergence-free flow, point observations, and periodic boundary conditions, the framework herein could be adapted to other assumptions via a different definition of the forward map $\mathcal{G}$.

As we illustrate below, the proposed inverse problem is ill-posed, i.e., there may be many or no background flows $\mathbf{v}$ that match a given dataset $\mathcal{Y}$. To address this issue, we adopt a Bayesian approach, incorporating prior knowledge of background flows and descriptions of the observation error to develop probabilistic estimates of $\mathbf{v}$. Summaries of the Bayesian approach to inverse problems can be found in [23] and [33]. Moreover, since the target of the inversion, the background flow $\mathbf{v}$, is infinite-dimensional, this work will leverage the considerable amount of recent research in infinite-dimensional Bayesian inference, grounding much of our approach in the overview of the field provided in [20]. To compute observables, such as the mean,

variance, or (normalized) histogram of a given quantity on $\mathbf{v}$ or θ, we use recently developed Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithms that are well-defined in infinite dimensions. We focus on preconditioned Crank–Nicolson (pCN) [18] and Hamiltonian (or Hybrid) Monte Carlo (HMC) [4] samplers; some results for the independence sampler and Metropolis-adjusted Langevin (MALA) methods (see descriptions in, e.g., [20] and [3], respectively) are provided in the supplementary material.

This work makes two important contributions. We lay out a Bayesian framework for the estimation of divergence-free background flows from observations of scalar behavior, a common experimental approach to investigating complex fluid flows [34, 35, 58, 51]. In doing so, we leverage recent advances in the theory of Bayesian inference and adjoint methods in high dimensions. We then identify two interesting examples for which the resulting posterior measures have very different structures—one fairly simple and one highly multimodal. We use these two examples to conduct a systematic, large-scale numerical study to benchmark the convergence of the MCMC methods mentioned above for "easy" and "hard" problems. This is a companion paper to [10], where we investigate the behavior of the posterior measure as the number of point observations grows large (see also [40]), and to [9], where we identify a computationally efficient approach to computing the forward map.

The structure of the paper is as follows. Section 2 defines our parametrization of the space of divergence-free flows, describes why the inverse problem is ill-posed in the traditional sense, and presents the Bayesian approach to the inverse problem. Section 3 describes the numerical approach to computing the posterior measure: MCMC methods for sampling from the posterior, numerical methods for solving the advection-diffusion equation (1.1), and an adjoint method for computing the gradients required for some MCMC methods. Section 4 provides results of the inference and convergence of MCMC methods as applied to two example problems. For completeness, appendices in the supplementary material provide additional numerical results (sections SM1 and SM2) and a description of Bayesian inference in a very general setting (section SM3).

1.1. Literature review.

Bayesian inference and MCMC. Comprehensive overviews of modern Bayesian techniques, from the basics of probability theory to computational practicalities, can be found in [23] and [33]. In the last ten years, much attention has been paid to the development of the theory of Bayesian inference for infinite-dimensional problems (e.g., where the target of the inversion is a function). These advances are summarized in [20], building upon the work in [52]; we follow the former closely in section 2 and provide a somewhat more general derivation of Bayes' Theorem in section SM3. See also [13, 45] for studies of applications of Bayesian inversion to PDE-constrained inverse problems.

Similarly, while Metropolis–Hastings MCMC methods date back to the foundational works [42] and [30], substantial recent work has gone into extending these methods to problems where the space to be sampled is high- or infinite-dimensional [5, 6, 7]. The goal of these efforts has been to define sampling kernels that are both well-defined and yield robust convergence even as the number of dimensions to be sampled grows large. The extension of Metropolis–Hastings methods to generalized state spaces was described in [55]. The behavior of the traditional random walk approach as the dimension grows large was investigated in [41] for a broad class

of target measures. The pCN and MALA algorithms suitable for infinite-dimensional problems were laid out in [18]; the optimal choice of the step size parameter in the MALA algorithm was shown in [46]. HMC was similarly extended to infinite dimensions in [4], work that was later generalized in [3] and [24]. Dimension-independent convergence of some of the above methods has been investigated by showing that the kernels have spectral gaps [22, 29, 56], leveraging a generalized version of Harris's Theorem [26, 28, 27, 43] for Markov kernels. The work in section 4 represents one of the first attempts to benchmark these methods for an infinite-dimensional application; see also [2, 3, 14] for related numerical experiments.

In subsection 4.2 we present a multimodal posterior measure, which MCMC methods have difficulty resolving. This has been a known problem with MCMC almost since its inception, and a number of ideas have been proposed for improving sampling for these distributions. One example is tempering, in which a series of "less steep" distributions are used to try to increase the probability of jumps between modes; see, e.g., the description in [23, section 12.3] and associated references. A related method is equi-energy sampling [36], in which rings of parameter values associated with different energy levels are constructed and samples are allowed to jump within rings.

Advection-diffusion. The problem of observing scalar behavior to infer the underlying velocity field is a common experimental approach for investigating the structure of complex fluid flows. The textbooks [58] and [51] describe many such methods, including examples where dye, smoke, temperature, hydrogen bubbles, or photosensitive tracers are used. An overview of dye-based visualization techniques is provided in [34]. An application of dye to investigate two-dimensional turbulence is described in [57]; see the survey article [35] for additional examples.

To our knowledge, this work is the first to apply Bayesian inference to the problem of estimating a background fluid flow from measurements of a passive scalar. However, a number of works, such as [1], have used inversion techniques to determine a source (forcing) term in advection-diffusion problems. In those previous works, the background flow was assumed to be known and the initial condition assumed to be zero; their goal was to determine the function (in particular the location) from which the pollutant was dispersed. The source-identification work was extended to ensure robustness to uncertainties in the velocity field in [60].

More generally, the advection and diffusion of passive scalars has been studied extensively. Numerical difficulties in modeling the behavior of passive scalars for advection-dominated cases are described in [44] and [53]. Passive scalars exhibit similar behavior for turbulent and random flows, so the latter, simpler case may be used to model the former. One such model was introduced by Kraichnan [37, 38, 39]; the energy spectrum from this model motivates the construction of the prior measure in section 4.

2. Mathematical framework and Bayesian inference.

In this section, we describe the mathematical framework of the inverse problem (1.2). We begin by defining the functional analytic setting for the problem, including how we represent divergence-free background flows. We then describe reasons why the inverse problem is ill-posed, i.e., why a given set of measurements $\mathcal{Y}$ cannot identify a unique background flow $\mathbf{v}$ that generated them. We close the section by defining the Bayesian approach to the inverse problem.

2.1. Representation of divergence-free background flows. The target of the inference is a divergence-free background flow $\mathbf{v}$, so we start by describing the space H of such flows that we will consider. For this purpose we begin by recalling the Sobolev spaces of (scalar-valued) periodic functions on the domain $\mathbb{T}^2 = [0,1]^2$,

(2.1)
$$H^s(\mathbb{T}^2) = \left\{ u : u = \sum_{\mathbf{k}\in\mathbb{Z}^2\setminus\{\mathbf{0}\}} c_{\mathbf{k}} e^{2\pi i \mathbf{k}\cdot\mathbf{x}},\ \overline{c_{\mathbf{k}}} = c_{-\mathbf{k}},\ \|u\|_{H^s} < \infty \right\},\ \|u\|_{H^s}^2 := \sum_{\mathbf{k}\in\mathbb{Z}^2} \|\mathbf{k}\|^{2s} |c_{\mathbf{k}}|^2,$$

defined for any $s \in \mathbb{R}$; see, e.g., [49, 54]. We will abuse notation and use the same notation for periodic divergence-free background flows by replacing the coefficients $c_{\mathbf{k}}$ in (2.1) with

(2.2)
$$c_{\mathbf{k}} = v_{\mathbf{k}} \frac{\mathbf{k}^\perp}{\|\mathbf{k}\|_2}, \quad \overline{v_{\mathbf{k}}} = -v_{-\mathbf{k}},$$

where for $\mathbf{k} = (k_1, k_2)$ we choose $\mathbf{k}^\perp = (-k_2, k_1)$ to ensure $\mathbf{k}\cdot\mathbf{k}^\perp = 0$. Throughout the remainder of the paper we fix our parameter space as follows.

Notation 2.1 (parameter space, H). *We consider background flows $\mathbf{v} \in H$, where $H = H^m(\mathbb{T}^2)$ (see (2.1)) for some $m > 1$, with coefficients $c_{\mathbf{k}}$ given by (2.2).*

Here the exponent m is chosen so that vector fields in H, as well as their corresponding solutions $\theta(\mathbf{v})$, exhibit continuity properties convenient for our analysis below (see Proposition 2.2). We take $L^p(\mathbb{T}^2)$ with $p \in [1,\infty]$ for the usual Lebesgue spaces and denote the space of continuous and pth integrable, X-valued functions by $C([0,T];X)$ and $L^p([0,T];X)$, respectively, for a given Banach space X. All of these spaces are endowed with their standard topologies unless otherwise specified.

In what follows we frequently consider Borel probability measures on H, denoted sometimes as $\Pr(H)$, in reference to the prior and posterior measures produced by Bayes' Theorem below. A natural approach to constructing certain classes of such infinite-dimensional probability measures is to decompose them into one-dimensional probability measures acting independently on individual components of a sequence of elements sitting in an underlying function space; see, e.g., [20, section 2]. Concretely in our setting, probability measures on the space of divergence-free vector fields can be defined by letting $v_{\mathbf{k}}$ be random fields, as long as $v_{\mathbf{k}}$ exhibits suitable decay to zero as $\|\mathbf{k}\| \to \infty$ commensurate with $\mathbf{v} \in H$ (see Notation 2.1). In particular we make use of this construction on the space of divergence-free vector fields to define prior distributions in the numerical examples in section 4.

2.2. Mathematical setting for the advection-diffusion equation. In this section, we provide a precise definition of solutions θ for the advection-diffusion problem (1.1). Crucially, the setting we choose yields a map from $\mathbf{v}$ to θ and then to observations of θ that is continuous.

Proposition 2.2 (well-posedness and continuity of the solution map for (1.1)).

(i) *Fix any $s \geq 0$ and $m \geq s$ with $m > 0$ and suppose that $\mathbf{v} \in H^m(\mathbb{T}^2)$ and $\theta_0 \in H^s(\mathbb{T}^2)$. Then there exists a unique $\theta = \theta(\mathbf{v}, \theta_0)$ such that*

$$\theta \in L^2_{loc}([0,\infty); H^{s+1}(\mathbb{T}^2)) \cap L^\infty([0,\infty); H^s(\mathbb{T}^2)) \quad \text{with} \quad \frac{\partial\theta}{\partial t} \in L^2_{loc}([0,\infty); H^{s-1}(\mathbb{T}^2))$$

so that in particular $\theta \in C([0,\infty); H^s(\mathbb{T}^2))$ *solves* (1.1) *at least weakly; namely,*

$$\left\langle \frac{\partial \theta}{\partial t}, \phi \right\rangle_{H^{-1}(\mathbb{T}^2)\times H^1(\mathbb{T}^2)} + \langle \mathbf{v}\cdot\nabla\theta, \phi\rangle_{L^2(\mathbb{T}^2)} + \kappa\langle \nabla\theta, \nabla\phi\rangle_{L^2(\mathbb{T}^2)} = 0 \tag{2.3}$$

for all $\phi \in H^1(\mathbb{T}^2)$ *and almost all time* $t \in [0,\infty)$.

(ii) *For any* $T > 0$ *the map that associates* $\mathbf{v} \in H^m(\mathbb{T}^2)$ *and* $\theta_0 \in H^s(\mathbb{T}^2)$ *to the corresponding* $\theta(\mathbf{v}, \theta_0)$ *is continuous relative to the standard topologies on* $H^m(\mathbb{T}^2)\times H^s(\mathbb{T}^2)$ *and* $C\left([0,T]\times H^s(\mathbb{T}^2)\right)$.

(iii) *For any* $T > 0, m \geq s > 1$ *the map which associates* $\mathbf{v} \in H^m(\mathbb{T}^2)$ *and* $\theta_0 \in H^s(\mathbb{T}^2)$ *to the corresponding* $\theta(\mathbf{v}, \theta_0)$ *is continuous relative to the standard topologies on* $H^m(\mathbb{T}^2)\times H^s(\mathbb{T}^2)$ *and* $C\left([0,T]\times \mathbb{T}^2\right)$.

A sketch of the proof is provided in [10].

Remark 2.3. Since the background flow $\mathbf{v}$ enters (1.1) through the $\mathbf{v}\cdot\nabla\theta$ term, the inverse problem of recovering $\mathbf{v}$ from $\theta(\mathbf{v})$ can be ill-posed. One important class of examples illustrating this difficulty arises when $\mathbf{v}\cdot\nabla\theta$ is zero everywhere, in which case the fluid flow does not have any influence on θ. Two such examples are as follows:

(i) *Ill-posedness: Laminar flow.* Let $\theta_0(\mathbf{x})$ be independent of y and $\mathbf{v}^\star = [0, f(x)]$. Then $\theta(\mathbf{v}^\star) = \theta(\mathbf{v})$ for any $\mathbf{v} = [0, g(x)]$.

(ii) *Ill-posedness: Radial symmetry.* Set $\theta_0(\mathbf{x}) \propto \sin(\pi x) + \sin(\pi y)$ and $\mathbf{v}^\star = [\cos(\pi x), -\cos(\pi y)]$. Then $\theta(\mathbf{v}^\star) = \theta(\mathbf{v})$ for any $\mathbf{v} = c\mathbf{v}^\star$, $c \in \mathbb{R}$.

In these cases, the even noiseless and complete spatial/temporal observations of θ cannot discriminate between a range of background flows, making it impossible to uniquely identify a true background flow $\mathbf{v}^\star$.

With this general result in hand we now fix some notation used for the remainder of the paper.

Definition 2.4 (solution operator $\mathcal{S}$, observation operator $\mathcal{O}$). *Fix* $\theta_0 \in H^s(\mathbb{T}^2)$ *and a final time* $T > 0$ *and consider the phase space* H *defined as in Notation* 2.1. *The forward map* $\mathcal{G}$ *as in* (1.2) *is interpreted as the composition* $\mathcal{G}(\mathbf{v}) = \mathcal{O}\circ\mathcal{S}(\mathbf{v})$, *where the following hold:*

1. *The* solution operator $\mathcal{S} : H \to C([0,T]; H^s(\mathbb{T}^2))$ *maps a given* $\mathbf{v}$ *to the corresponding solution* $\theta(\mathbf{v}, \theta_0)$ *of* (1.1) *(in the sense of Proposition* 2.2*).*
2. *The* observation operator $\mathcal{O} : C([0,T]; H^s(\mathbb{T}^2)) \to Y$ *measures some quantities (e.g., point measurements, spectral data, tracers) from* θ. *Here, in general,* Y *is a separable Hilbert space. However, since we are primarily focused on the setting of finite observations, we typically have* $Y = \mathbb{R}^N$.

To make a connection with the range of observations provided in the introduction, we detail the following possibilities for $\mathcal{O}$.

Example 2.5. We consider finite observations $\mathcal{O}(\theta) = (\mathcal{O}_1(\theta), \ldots, \mathcal{O}_N(\theta))$ that could be

1. spectral observations: $\mathcal{O}_j(\theta) = \int_{[0,1]^2} \theta(t_j, \mathbf{x}) e_j d\mathbf{x}$, with $\{e_j\}_{j\geq 0}$ an orthonormal basis for H^s and $t_j \in [0,T]$;
2. local averages: $\mathcal{O}_j(\theta) = \frac{1}{|\mathcal{D}_j|}\int_{\mathcal{D}_j} \theta \, d\mathbf{x}\, dt$, for any subdomains $\mathcal{D}_j \subset [0,T]\times[0,1]^2$, where $|\mathcal{D}_j|$ denotes the volume of $\mathcal{D}_j$;

3. spatial-temporal point observations: $\mathcal{O}_j(\theta) = \theta(t_j, \mathbf{x}_j)$ for any $t_j \in [0,T]$ and $\mathbf{x}_j \in [0,1]^2$ (note that point observations are well-defined by Proposition 2.2(iii)).

This paper focuses on the final case of point observations $\mathcal{G}_j(\mathbf{v}) = \theta(t_j, \mathbf{x}_j, \mathbf{v})$, $j = 1, \dots, N$, as the most obvious practical implementation for the advection-diffusion problem.

2.3. Ill-posedness. We note that the classical inverse problem of recovering $\mathbf{v}$ from data $\mathcal{Y}$ (see (1.2)) is highly ill-posed in a number of ways:

1. The data is incomplete, i.e., we do not observe θ everywhere. For this reason we are interested in forward maps $\mathcal{G}(\mathbf{v})$ that are noninvertible and hence that do not uniquely specify $\mathbf{v}$. One such example is provided in subsection 4.2.
2. Even if solutions θ of (1.1) are observed everywhere in space and time, there are initial conditions θ_0 such that any of a range of background flows $\mathbf{v}$ produce the same scalar field θ. Two such examples are provided in Remark 2.3 above.
3. Because of the observational noise η in (1.2), there may be no $\mathbf{v}$ such that $\mathcal{G}(\mathbf{v}) = \mathcal{Y}$ for given data $\mathcal{Y}$. For example, in the case of point observations $\mathcal{Y}_j = \theta(t_j, \mathbf{x}_j, \mathbf{v}^\star) + \eta_j$, some realizations of η_j may cause $\mathcal{Y}_j$ to exceed the maximum value (or be less than the minimum value) of θ_0. However, because $\nabla \cdot \mathbf{v} = 0$, (1.1) is a parabolic PDE that is subject to a maximum principle implying $\|\theta(t)\|_{L^\infty(\mathbb{T}^2)} \le \|\theta_0\|_{L^\infty(\mathbb{T}^2)}$ for all $t > 0$. Thus there would be no $\mathbf{v}$ such that $\mathcal{G}(\mathbf{v}) = \mathcal{Y}$.

These considerations are typical of ill-posed inverse problems more broadly. See, e.g., [33, section 2] or [52] for further commentary.

2.4. Bayesian inference. Following the Bayesian approach to inverse problems [20, 33], instead of seeking a single best match $\mathbf{v}^\star$, we take a statistical interpretation of $\mathbf{v}$ and η as random quantities that we refer to as the "prior" and the "observation noise." The solution of (1.2) is a probability measure, known as the "posterior," associated with the conditional random variable "$\mathbf{v}|\mathcal{Y}$." The concentration of the prior measure in the limit of a large number of observations, i.e., the question of consistency, is investigated in detail in [10]. A quite general formulation of Bayes' Theorem is provided in section SM3 of the supplementary material; in this section we follow closely the derivation in [20]. We begin by imposing the following typical assumption.

Assumption 2.6. The joint distribution of the observation noise and the prior take the form $(\mathbf{v}, \eta) \sim \mu_0 \otimes \gamma_0$ for $\mu_0 \in \Pr(H)$, $\gamma_0 \in \Pr(\mathbb{R}^N)$ so that $\mathbf{v}$ and η are statistically independent.

Under Assumption 2.6, the "likelihood" $\mathbb{Q}_\mathbf{v}$, heuristically $\mathcal{Y}|\mathbf{v}$, is as follows.

Lemma 2.7 (likelihood $\mathbb{Q}_\mathbf{v}$). *For any deterministic background flow $\mathbf{v} \in H$ and observation noise $\eta \sim \gamma_0$, the* likelihood $\mathbb{Q}_\mathbf{v}$ *satisfies $\mathcal{G}(\mathbf{v}) + \eta \sim \mathbb{Q}_\mathbf{v}$ so that for any $A \in \mathcal{B}(\mathbb{R}^N)$,*

$$\mathbb{Q}_\mathbf{v}(A) = \gamma_0(\{y - \mathcal{G}(\mathbf{v}) : y \in A\}). \tag{2.4}$$

With the form of the likelihood measure $\mathbb{Q}_\mathbf{v}$ in hand we introduce the following notational convention used several times below.

Notation 2.8 (true background flow, $\mathbf{v}^\star$). *We frequently fix a "true" background flow by $\mathbf{v}^\star \in H$. For the given $\mathbf{v}^\star$, the observed data $\mathcal{Y} = \mathcal{G}(\mathbf{v}^\star) + \eta$ can be viewed as draws from the distribution $\mathbb{Q}_{\mathbf{v}^\star}$ (though $\mathbf{v}^\star$ is not necessarily the only $\mathbf{v}$ that could produce such data).*

As in [20], we make the following assumption.

Assumption 2.9. The likelihood $\mathbb{Q}_{\mathbf{v}}$ (see Lemma 2.7) is absolutely continuous with respect to the noise measure γ_0 for all $\mathbf{v} \in H$.

We note that this assumption holds when γ_0 is any continuously distributed measure, such as a (nondegenerate) Gaussian, that has the whole space $\mathbb{R}^N$ as its support. (We also note in Proposition SM3.5 in the supplementary material that γ_0 is not the only suitable choice of reference measure.) Then we define the following.

Definition 2.10 (potential, Φ). *When Assumption 2.9 holds, the* potential *or* negative log-likelihood $\Phi : H \times \mathbb{R}^N \to \mathbb{R}$ *is defined as*

$$\Phi(\mathbf{v}; \mathcal{Y}) = -\log\left(\frac{d\mathbb{Q}_{\mathbf{v}}}{d\gamma_0}(\mathcal{Y})\right), \tag{2.5}$$

where $\frac{d\mathbb{Q}_{\mathbf{v}}}{d\gamma_0}$ *is the Radon–Nikodym derivative of* $\mathbb{Q}_{\mathbf{v}}$ *with respect to* γ_0.

Example 2.11 (Gaussian noise). If the observation noise is a centered Gaussian, i.e., $\gamma_0 = N(0, \mathcal{C}_\eta)$, then by (2.4) we have (up to a factor independent of $\mathbf{v}$)

$$\Phi(\mathbf{v}; \mathcal{Y}) = \frac{1}{2}\left\|\mathcal{C}_\eta^{-1/2}\left(\mathcal{Y} - \mathcal{G}(\mathbf{v})\right)\right\|^2. \tag{2.6}$$

Finally, we have the following adaptation of Bayes' Theorem to the advection-diffusion problem.

Theorem 2.12 (Bayes' Theorem [20]). *Let* $\mathbb{Q}_{\mathbf{v}}$ *and* Φ *be defined as in Lemma 2.7 and Definition 2.10, respectively, and let* $\mathbb{Q}_{\mathbf{v}}$ *satisfy Assumption 2.9. Suppose that* Φ *is measurable in* $\mathbf{v}$ *and* $\mathcal{Y}$ *and that*

$$Z = \int \exp\left(-\Phi(\mathbf{v}; \mathcal{Y})\right) \mu_0(d\mathbf{v}) > 0. \tag{2.7}$$

Then the measure $\mu_{\mathcal{Y}}$ *associated with the random variable* $\mathbf{v}|\mathcal{Y}$ *is absolutely continuous with respect to* μ_0*, with Radon–Nikodym derivative*

$$\frac{d\mu_{\mathcal{Y}}}{d\mu_0}(\mathbf{v}) = \frac{1}{Z}\exp\left(-\Phi(\mathbf{v}; \mathcal{Y})\right). \tag{2.8}$$

3. Computational approach and challenges.

In this section, we describe the numerical methods used to approximate the posterior measure μ. We begin by introducing Markov chain Monte Carlo (MCMC) methods used to generate samples from μ (subsection 3.1). Subsection 3.2 describes how we discretize and solve the advection-diffusion equation (1.1) to compute the potential Φ (see Definition 2.10). Finally, in subsection 3.3 we define an adjoint method for efficient computation of the Fréchet derivative $D\Phi$, which is required for implementation of some of the more advanced MCMC algorithms described in subsection 3.1.

3.1. Sampling from μ via Markov chain Monte Carlo. To sample from the posterior measure $\mu_{\mathcal{Y}}$ (see Theorem 2.12), we use two MCMC methods recently developed for or extended to infinite-dimensional problems: (1) preconditioned Crank–Nicolson (pCN) [18], a generalization of the classical random walk algorithm that requires one forward evaluation (PDE solve) per iteration and represents the "inexpensive" end of the computational spectrum (see Algorithm 3.1); and (2) Hamiltonian Monte Carlo (HMC) [11, 3, 21], a "computationally expensive" method that requires multiple PDE solves and gradient computations per iteration (see Algorithm 3.2). In the supplementary material section SM1, we additionally present some results for the independence sampler and Metropolis-adjusted Langevin algorithm (MALA) (see, e.g., descriptions in [20] and [3], respectively). See also [40, Chapter 5] for a detailed description of these four methods and [32] for an algorithm to recursively select parameters for HMC.

Algorithm 3.1. Preconditioned Crank–Nicolson (pCN).

1: Given free parameter $\beta \in (0,1]$ and initial sample $\mathbf{v}^{(k)}$
2: Propose $\tilde{\mathbf{v}} = \sqrt{1-\beta^2}\mathbf{v}^{(k)} + \beta\xi^{(k)}$, $\xi^{(k)} \sim N(0,\mathcal{C})$
3: Set $\mathbf{v}^{(k+1)} = \tilde{\mathbf{v}}$ with probability $\min\left\{1, \exp\left(\Phi\left(\mathbf{v}^{(k)}\right) - \Phi\left(\tilde{\mathbf{v}}\right)\right)\right\}$, otherwise $\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)}$

Algorithm 3.2. Hamiltonian MCMC (HMC).

1: Given free parameters $\tau \geq \epsilon > 0$ and initial sample $\mathbf{v}^{(k)}$. Set $L = \frac{\tau}{\epsilon}$.
2: Set $(\mathbf{q}_0, \mathbf{w}_0) = (\mathbf{v}^{(k)}, \mathbf{w})$, where $\mathbf{w} \sim \mu_0 = N(0,\mathcal{C})$
3: **for** $i = 1$ to L **do**
4: Integrate Hamiltonian dynamics $(\mathbf{q}_{i-1}, \mathbf{w}_{i-1}) \mapsto (\mathbf{q}_i, \mathbf{w}_i)$ via (see [3, equation (13)])

$$\begin{aligned}
\mathbf{w}_- &= \mathbf{w}_{i-1} - \frac{\epsilon}{2}\mathcal{C}D\Phi(\mathbf{q}_{i-1}),\\
\mathbf{q}_i &= \quad (\cos\epsilon)\mathbf{q}_{i-1} + (\sin\epsilon)\mathbf{w}_-,\\
\mathbf{w}_+ &= -(\sin\epsilon)\mathbf{q}_{i-1} + (\cos\epsilon)\mathbf{w}_-,\\
\mathbf{w}_i &= \mathbf{w}_+ - \frac{\epsilon}{2}\mathcal{C}D\Phi(\mathbf{q}_i)
\end{aligned}$$

5: **end for**
6: Compute $\Delta\mathcal{H} = \mathcal{H}(\mathbf{q}_L, \mathbf{w}_L) - \mathcal{H}(\mathbf{v}^{(k)}, \mathbf{w})$
7: Set $\mathbf{v}^{(k+1)} = \mathbf{q}_L$ with probability $\min\{1, \exp[-\Delta\mathcal{H}]\}$, otherwise $\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)}$

3.2. Evaluation of $\mathcal{G}$. Computing the potential $\Phi(\mathbf{v})$ (see Definition 2.10) as in Algorithm 3.1 or 3.2 requires evaluating $\mathcal{G}(\mathbf{v})$, i.e., computing (e.g., point) observations for $\theta(\mathbf{v})$. This requires numerically solving (1.1) using a PDE solver. We do so using a spectral method [15, 25], expanding $\mathbf{v}$ in a Fourier basis as in (2.1), (2.2), and θ similarly as $\theta(t,\mathbf{x}) = \sum_{\mathbf{k}} \theta_{\mathbf{k}}(t) e^{2\pi i \mathbf{k}\cdot\mathbf{x}}$. We apply a Galerkin projection, writing the coefficients $\theta_{\mathbf{k}}$ as a system of ODEs that reduces to

$$\frac{d}{dt}\vec{\theta}(t) = A\vec{\theta}(t), \quad \text{where } (A)_{lm} = -iv_{\mathbf{k}'}\left(\mathbf{k}'^{\perp}\cdot\mathbf{k}_m\right) - \kappa\left\|\mathbf{k}_l\right\|^2\delta_{lm}, \quad \text{where } \mathbf{k}' = \mathbf{k}_l - \mathbf{k}_m. \tag{3.1}$$

This system is then integrated using the implicit midpoint (Crank–Nicolson) method to approximate $\vec{\theta}(t)$. Point observations are calculated as $\theta(t_j, \mathbf{x}_j) = \sum_{\mathbf{k}} \theta_{\mathbf{k}}(t_j) e^{2\pi i \mathbf{k}\cdot\mathbf{x}_j}$, with evaluation at time t_j interpolated if t_j does not fall on a timestep of the integration method. When κ is small, an accurate representation of θ requires a large number of components due to the frequency cascade $\mathbf{k}'$ in (3.1) (see [53]). This challenge motivates our concurrent work in [9], in which we introduce a particle method for efficient evaluation of $\mathcal{G}(\mathbf{v})$, allowing the computation of large numbers of samples for low-κ problems.

3.3. An adjoint method for evaluating the gradient of Φ. HMC (Algorithm 3.2) requires evaluating the Fréchet derivative of the potential Φ (see Definition 2.10) with respect to changes in $\mathbf{v}$, a direct approach to which would require many PDE solves. Here we introduce an adjoint approach that requires a single PDE solve per gradient computation. Adjoint methods are well developed for applications in optimization and uncertainty quantification; we note in particular [1], in which a similar adjoint equation was derived for a different application involving advection and diffusion. More generally, we refer the reader to [31] for an abstract overview of the adjoint approach to computing gradients for PDE-constrained optimization and to, e.g., [12, 16, 47, 59] for applications of adjoint methods to inverse problems and data assimilation. For our application, the structure of the PDE turns out to be such that the adjoint solution can be computed using (a forced version of) the same solver used to solve the forward equation (1.1), as we now describe.

Theorem 3.1 (adjoint method for evaluating $D\Phi$). *For a given background flow $\mathbf{v} \in H$, let $\theta(\mathbf{v}) \in H^m\left([0,T]; H^s(\mathbb{T}^2)\right)$ be a weak solution of the advection-diffusion equation (1.1) in the sense that $\theta(0,\mathbf{x}) = \theta_0$ a.e. and*

$$\left\langle \rho, \frac{d}{dt}\theta + \mathbf{v}\cdot\nabla\theta - \kappa\Delta\theta \right\rangle_{H^{1-m,2-s}\times H^{m-1,s-2}} = 0 \tag{3.2}$$

for all $\rho \in H^{1-m}\left([0,T]; H^{2-s}(\mathbb{T}^2)\right)$. Suppose that Φ (see Definition 2.10) and $\mathcal{O} : H^m\left([0,T]; H^s(\mathbb{T}^2)\right) \to \mathbb{R}^N$ (see Definition 2.4) are continuously differentiable in $\mathcal{G}$ and θ, respectively. Suppose there exists a $\rho_0 \in H^{1-m}\left([0,T]; H^{2-s}(\mathbb{T}^2)\right)$ with $\rho_0(0,\mathbf{x}) = 0$ a.e. that solves the forced adjoint equation

$$\left\langle \frac{d}{dt}\rho_0 - \mathbf{v}\cdot\nabla\rho_0 - \kappa\Delta\rho_0, \phi \right\rangle_{H^{-m,-s}\times H^{m,s}} = -\frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\mathcal{O}[\tilde{\phi}] \tag{3.3}$$

for all $\phi \in H^m\left([0,T]; H^s(\mathbb{T}^2)\right)$ with $\phi(T,\mathbf{x}) = 0$ a.e., where $\tilde{\phi}(t,\mathbf{x}) \coloneqq \phi(T-t,\mathbf{x})$. Then the Fréchet derivative of Φ at $\mathbf{v}$ in the direction $\hat{\mathbf{v}}$ is given by

$$D_{\hat{\mathbf{v}}}\Phi(\mathbf{v}) = \langle \tilde{\rho}_0, \hat{\mathbf{v}}\cdot\nabla\theta\rangle_{H^{1-m,2-s}\times H^{m-1,s-2}}, \tag{3.4}$$

where $\tilde{\rho}_0(t,\mathbf{x}) \coloneqq \rho_0(T-t,\mathbf{x})$.

Proof. Application of the chain rule yields

$$D_{\hat{\mathbf{v}}}\Phi(\mathbf{v}) = \frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot D_{\hat{\mathbf{v}}}\mathcal{G}(\mathbf{v}) = \frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\mathcal{O}\left[D_{\hat{\mathbf{v}}}\theta(\mathbf{v})\right]. \tag{3.5}$$

Denote $D_{\hat{\mathbf{v}}}\theta(\mathbf{v})$ by $\psi(\mathbf{v},\hat{\mathbf{v}})$. Then by applying (3.2) to $\theta(\mathbf{v}+\epsilon\hat{\mathbf{v}})$ and $\theta(\mathbf{v})$, subtracting, taking the $\epsilon \to 0$ limit, and using the definition of the Fréchet derivative, we see that $\psi \in H^m\left([0,T];H^s(\mathbb{T}^2)\right)$ satisfies

$$\left\langle \rho, \frac{d}{dt}\psi + \mathbf{v}\cdot\nabla\psi - \kappa\Delta\psi + \hat{\mathbf{v}}\cdot\nabla\theta \right\rangle_{H^{1-m,2-s}\times H^{m-1,s-2}} = 0, \tag{3.6}$$

with $\psi(0,\mathbf{x}) = 0$ a.e., for all $\rho \in H^{1-m}\left([0,T];H^{2-s}(\mathbb{T}^2)\right)$. Also, changing variables from t to $T-t$ in (3.3) yields the following relationship for $\tilde{\rho}_0(t) = \rho_0(T-t)$:

$$\left\langle \frac{d}{dt}\tilde{\rho}_0 + \mathbf{v}\cdot\nabla\tilde{\rho}_0 + \kappa\Delta\tilde{\rho}_0, \tilde{\phi} \right\rangle_{H^{-m,-s}\times H^{m,s}} = \frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\mathcal{O}[\phi], \tag{3.7}$$

with $\tilde{\rho}_0(T,\mathbf{x}) = \tilde{\phi}(0,\mathbf{x}) = 0$ a.e. Then applying (3.5), (3.7), and (3.6) in succession yields

$$\begin{aligned} D_{\hat{\mathbf{v}}}\Phi(\mathbf{v}) = \frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\mathcal{O}[\psi] &= \left\langle \frac{d}{dt}\tilde{\rho}_0 + \mathbf{v}\cdot\nabla\tilde{\rho}_0 + \kappa\Delta\tilde{\rho}_0, \psi \right\rangle_{H^{-m,-s}\times H^{m,s}} \\ &= \left\langle \tilde{\rho}_0, -\frac{d}{dt}\psi - \mathbf{v}\cdot\nabla\psi + \kappa\Delta\psi \right\rangle_{H^{1-m,2-s}\times H^{m-1,s-2}} \\ &= \langle \tilde{\rho}_0, \hat{\mathbf{v}}\cdot\nabla\theta\rangle_{H^{1-m,2-s}\times H^{m-1,s-2}}, \end{aligned}$$

which is the desired result. ∎

Remark 3.2. Note that $\left[\frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\mathcal{O}\right] \in H^{-m}\left([0,T];H^{-s}(\mathbb{T}^2)\right)$, so solving (3.3) amounts to finding the weak solution in $H^{-m}\left([0,T];H^{-s}(\mathbb{T}^2)\right)$ of

$$\frac{d}{dt}\rho_0 - \mathbf{v}\cdot\nabla\rho_0 - \kappa\Delta\rho_0 = \frac{\partial\Phi}{\partial\mathcal{G}}(\mathbf{v})\cdot\tilde{\mathcal{O}}, \quad \rho_0(0,\mathbf{x}) = 0 \text{ a.e.}, \tag{3.8}$$

where $\tilde{\mathcal{O}}[\phi(t,\mathbf{x})] \coloneqq \mathcal{O}[\phi(T-t,\mathbf{x})]$.

To compute the full gradient $D\Phi$ (the derivative with respect to an array of bases $\{\mathbf{e_k}\}$), we compute the integration (3.4) for $\hat{\mathbf{v}} = \mathbf{e_k}$ for each $\mathbf{k}$. The resulting algorithm is summarized in Algorithm 3.3. Note that solving (3.6) and substituting into (3.5) would also yield the derivative of Φ. However, this approach would require a PDE solve for each direction $\hat{\mathbf{v}}$ in which we want to take the derivative. In particular, if we want the full gradient, we have to do many PDE solves. By contrast, Algorithm 3.3 requires only one additional PDE solve per gradient calculation. Moreover, note that (3.8) is equivalent to (1.1) with zero initial condition, a reversed vector field, and a forcing term. Thus, the same PDE solver can be used for both the forward and adjoint solves with minimal modification. For the numerical experiments in section 4, the adjoint equation (3.3) was solved using the same spectral method described in subsection 3.2, where the forcing term was similarly expanded in the Fourier basis $e^{2\pi i\mathbf{k}\cdot\mathbf{x}}$ via Galerkin projection. The integration (3.4) was computed directly from the spectral representation of θ and ρ_0, yielding

$$D_{\mathbf{e_k}}\Phi(\mathbf{v}) = \sum_j 2\pi i\left(\mathbf{k}^\perp\cdot\mathbf{k}_j\right)\int_0^T \tilde{\rho}_{0_l}(t)\theta_j(t), \quad \text{where} \quad l \ni \mathbf{k}_l = -\mathbf{k}-\mathbf{k}_j, \tag{3.9}$$

where the time integration was computed via trapezoidal rule.

Algorithm 3.3. Adjoint method for computing $D\Phi$.

Given $\mathbf{v}$ and basis $\mathbf{e_k}$
Solve (1.1) for $\theta(t, \mathbf{x}, \mathbf{v})$
Solve (3.3) for $\rho_0(t, \mathbf{x}, \mathbf{v})$
for each $\mathbf{k}$ **do**
    Compute $D_{\mathbf{k}}\Phi(\mathbf{v})$ via (3.4) with $\hat{\mathbf{v}} = \mathbf{e_k}$
end for

Example 3.3 (point observations, Gaussian noise). Let the observation operator be point observations $\mathcal{O}_j[\theta] = \theta(t_j, \mathbf{x}_j)$. These observations are well-defined for $\theta \in H^m\left([0,T]; H^s(\mathbb{T}^2)\right)$ with $m > \frac{1}{2}, s > 1$ (see Proposition 2.2(iii)). Let $\eta_j \sim N(0, \sigma_\eta^2)$ for $j = 1, \dots, N$ so that Φ is given by (2.6). Then solving (3.3) amounts to finding the weak solution of

$$\frac{\partial}{\partial t}\rho_0 - \mathbf{v}\cdot\nabla\rho_0 - \kappa\Delta\rho_0 = \sum_j \frac{1}{\sigma_\eta^2}\left(\mathcal{Y}_j - \theta(t_j, \mathbf{x}_j, \mathbf{v})\right)\delta(T - t_j - t, \mathbf{x} - \mathbf{x}_j),$$

where $\rho_0(0, \mathbf{x}) = 0$ a.e. and $\delta(t - t_0, \mathbf{x} - \mathbf{x}_0)$ is a Dirac distribution centered at $(t_0, \mathbf{x}_0)$.

Example 3.4 (integral observations). Let observations be given by $\mathcal{O}_j[\theta] = \langle f_j, \theta\rangle_{L^2([0,T]\times\mathbb{T}^2)}$ for some $f_j \in H^1([0,T]; H^2(\mathbb{T}^2))$. Let $\theta \in L^2([0,T]; L^2(\mathbb{T}^2))$ (i.e., $m = s = 0$ in Theorem 3.1). Let $\eta_j \sim N(0, \sigma_\eta^2)$ for $j = 1, \dots, N$ so that Φ is given by (2.6). Then solving (3.3) amounts to finding $\rho_0 \in H^1([0,T], H^2(\mathbb{T}^2))$ with $\rho_0(0, \mathbf{x}) = 0$ such that

$$\frac{\partial}{\partial t}\rho_0 - \mathbf{v}\cdot\nabla\rho_0 - \kappa\Delta\rho_0 = \sum_j \frac{1}{\sigma_\eta^2}\left(\mathcal{Y}_j - \mathcal{O}_j[\theta(\mathbf{v})]\right) f_j(T - t, \mathbf{x}).$$

4. Numerical experiments: Posterior complexity and MCMC convergence.

In this section, we describe applications of the above methods to two sample problems. We begin with an example (subsection 4.1) that yields a posterior measure with a relatively simple structure. This provides a baseline for measuring convergence of the pCN and HMC samplers (see subsection 3.1), which for our purposes represent the "inexpensive" and "expensive" ends of the computational spectrum, respectively. Subsection 4.2 then presents a second example for which the posterior measure exhibits a complicated, multimodal structure. This example is more challenging for MCMC methods to sample from, and thus a good test of the advantages offered by more "expensive" methods like HMC. The supplementary material presents analogous results for the independence and MALA samplers (section SM1) and additional observables of interest to the passive scalar community (section SM2).

In each example, we generate data $\mathcal{Y}$ by running a high-resolution simulation of the system for a given true vector field $\mathbf{v}^\star$ and applying the observation operator $\mathcal{O}$. The PDE solver, adjoint solver, and MCMC methods were implemented in the Julia numerical computing language [8]. Thousands, or in some cases millions, of samples were generated using the computational resources at Virginia Tech.[1]

[1] http://www.arc.vt.edu

4.1. Example 1: Single-welled posterior. In this subsection, we construct an example that yields a posterior distribution with a simple, single-welled structure. The problem parameters for this example are enumerated in Table 1.[2] The true flow $\mathbf{v}^\star$ is shown in Figure 1.

Table 1
Problem parameters for Example 1.

Parameter	Value	Parameter	Value
Observation operator, $\mathcal{O}$	Point observations at 1,024 uniformly random (t, x, y)	Data, $\mathcal{Y}$	$\mathcal{G}(\mathbf{v}^\star)$
Prior, μ_0	Kraichnan (4.2)	Noise, γ_0	$N(0, \sigma_\eta^2 I)$, $\sigma_\eta = 2^{-6}$
True flow, $\mathbf{v}^\star$ (Figure 1)	Randomly drawn from Kraichnan prior, $\|\mathbf{k}\|_2 \leq 32$	Sampling space, H_N	$\|\mathbf{k}\|_2 \leq 8$ (197 components)
Diffusion, κ	0.282, for water in air [19]	θ_0	$\frac{1}{2} - \frac{1}{4}\cos(2\pi x) - \frac{1}{4}\cos(2\pi y)$

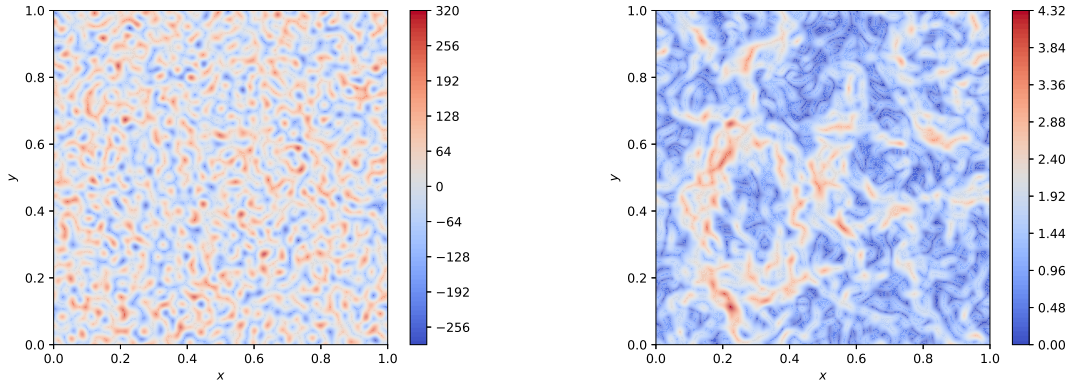

Figure 1. $\mathbf{v}^\star$ *for Example* 1. *Left: Vorticity* $\nabla \times \mathbf{v}^\star$. *Right:* $\|\mathbf{v}^\star\|$.

For the prior measure, we leverage the Kraichnan model [37, 38] of turbulent advection via a Gaussian random velocity field with energy spectrum (see [17, equation (28)])

$$E(k) = E_0 \sum_{i=0}^{N} \left(\frac{k}{k_i}\right)^4 \exp\left[-\frac{3}{2}\left(\frac{k}{k_i}\right)^2\right] k_i^{-\xi}, \tag{4.1}$$

where $k_i = \sqrt{2}^i$ is the characteristic wave number of the ith subfield, N is the number of subfields, and E_0 controls the overall energy. The resulting spectrum exhibits $E(k) \propto k^{-\xi}$ for $1 < k < k_N = 2^{N/2}$ and exponential decay for $k > k_N$. We then choose prior

$$\mu_0 = N(0, \tilde{E}), \quad \tilde{E}_{lm} = \frac{1}{2\pi \|\mathbf{k}_l\|_2} E(\|\mathbf{k}_l\|_2)\delta_{lm}, \tag{4.2}$$

[2]Note that $\mathbf{v}^\star$ and observation locations/times were chosen randomly only to generate the scenario and data. This random selection ensures a sufficiently general problem. The data and observation operator then of course remain fixed during MCMC sampling.

where E is as defined in (4.1), with $\xi = \frac{3}{2}$ motivated by [17, 39]. Then for $\mathbf{v} = \sum_{\mathbf{k}} v_k \sim \mu_0$, the expected energy associated with wave numbers of norm k (integrating across the shell $S_k = \{\mathbf{k} : \|\mathbf{k}\|_2 = k\}$) is $E(k)$. Laminar ($k = 0$) components represent something of a degenerate case; we simply assume that these two components are known to be zero.

Note: The Kraichnan model of mixing typically involves a velocity field with energy spectrum (4.1) but that is white (δ-correlated) in time [50]. Here $\mathbf{v}$ is a background flow, i.e., constant in time; we simply use the Kraichnan model as motivation for the energy decay modeled in the prior.

4.1.1. Posterior structure. As described above, the output of a Bayesian inference is the posterior $\mu_{\mathcal{Y}}$, a probability measure on the space of divergence-free vector fields H or, in practice, on a finite-dimensional approximation H_N given by the truncated expansion of the basis described in subsection 2.1. To approximate the exact posterior, we assembled a list of 10 million samples by running a series of 40 pCN chains of 250,000 samples each, with every chain beginning with an initial sample randomly chosen from the prior.[3] Figure 2 shows the structure of the computed posterior. The left-hand plot shows mean, variance, skew, and excess kurtosis (kurtosis minus 3) of the posterior by Fourier component of $\mathbf{v}$, where $\mathbf{v}$, incorporating the discretization (2.1) and reality condition (2.2), is constructed from the components as (laminar components v_0, v_1 are assumed to be zero, as noted above)

$$\begin{aligned}\mathbf{v}(\mathbf{x}) = {} & v_2\,(0, \cos(2\pi y)) + v_3\,(0, -\sin(2\pi y)) + v_4\,(\cos(2\pi x), 0) + v_5\,(-\sin(2\pi x), 0) \\ & + v_6\,(0, \cos(4\pi y)) + v_7\,(0, -\sin(4\pi y)) + v_8\,(\cos(2\pi x), \cos(2\pi y)) \\ & + v_9\,(-\sin(2\pi x), -\sin(2\pi y)) + \cdots .\end{aligned} \tag{4.3}$$

Because of the influence of the prior measure, the mean and covariance of the posterior for higher-order components tend to zero. Skew and excess kurtosis are near zero (up to computational resolution) for all components, indicating that the marginal distribution for each component is approximately Gaussian. The right-hand plot in Figure 2 presents one- and two-dimensional histograms of the first eight components of $\mathbf{v}$. Note that the histograms (and other plots omitted for brevity) all show a contiguous mass of probability, indicating that one "class" of vector field matches both the prior and the data.

4.1.2. MCMC sampling. To test the behavior of "inexpensive" and "expensive" MCMC methods, both pCN and HMC (see subsection 3.1) were applied to Example 1. The pCN parameter $\beta = 0.15$ was chosen to match the optimal acceptance rate of 23% from [48]. For HMC, $\epsilon = 0.125$ and $\tau = 1$ were chosen because these values showed a good balance between the desire for high acceptance rate, large jumps between samples, and low computational cost in numerical experiments. The resulting acceptance rates were 23.9% for pCN and 81.0% for HMC. Figure 3 shows the trace and autocorrelation of the potential Φ (see Definition 2.10). When pCN is applied to this example, we see "random walk" behavior—the samples move about the posterior, but are correlated with each other. For HMC, the random walk effect is reduced and samples exhibit independence from each other after orders of magnitude fewer

[3] pCN was chosen here because it provided a computationally inexpensive approximation to what proved to be a posterior with simple structure.

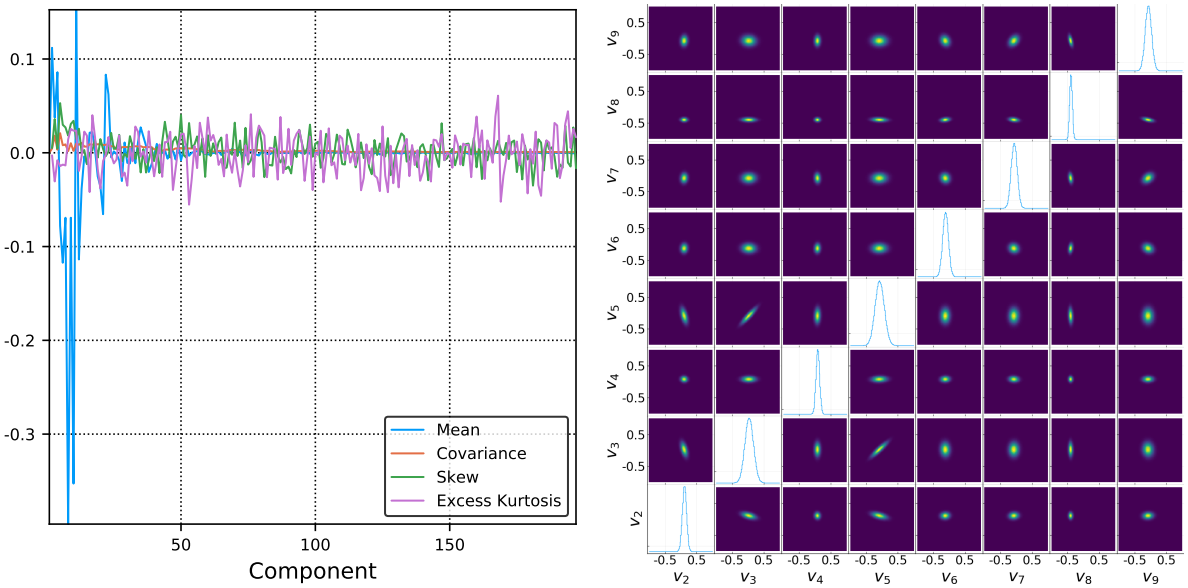

Figure 2. *Structure of posterior. Left: Mean, covariance, skew, and excess kurtosis of posterior measure, by component of* $\mathbf{v}$. *Right: Posterior* (μ_Y) *one-dimensional (diagonal) and two-dimensional (off-diagonal) marginal distributions for the first eight components of* $\mathbf{v}$ *(out of* 197*).*

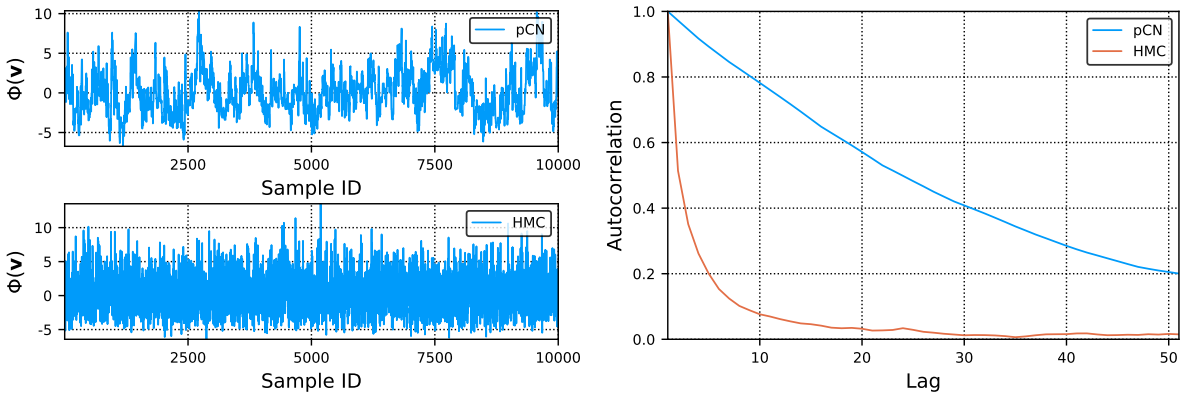

Figure 3. *Trace (left) and autocorrelation (right) of the potential* Φ.

iterations than for pCN. The HMC chain explores the posterior more quickly as a result. This is explored in the next section.

4.1.3. Convergence of measures. The difference between the "true" (Figure 2) and computed marginal distributions can be evaluated via the total variation distance.[4] Convergence of the MCMC chains to the true marginal distributions are shown in Figure 4. The figure

[4]Note that the real desire would be to measure convergence in the full 197-dimensional sample space. In this paper, we will use the total variation norm to measure convergence between one- and two-dimensional distributions, a necessary and easy to picture—but not sufficient—requirement for convergence in the full-dimensional space.

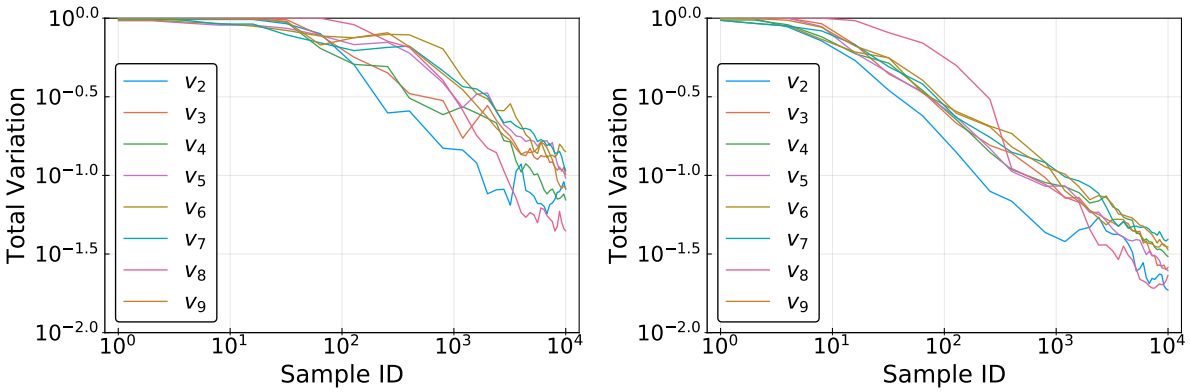

Figure 4. *Total variation norm between computed and "true" marginal probability density function for* $v_2, \ldots, v_9$ *for* 10,000 *samples. Left: pCN. Right: HMC.*

shows that HMC achieves a close approximation to the posterior marginal distributions within a few hundred iterations, while similar convergence takes about an order of magnitude longer for pCN.

4.1.4. Equal runtime comparison. Recall from subsection 4.1.2 that the parameters used for HMC were $\epsilon = 0.125$ and $\tau = 1.0$, meaning that $\frac{\tau}{\epsilon} = 8$ PDE and adjoint solves (see Algorithm 3.2) were required per HMC sample. Because of these solves and the additional costs required for the gradient computation (see Algorithm 3.3) and time integration, each HMC sample took the time of approximately 39 pCN samples to compute. Thus, we can reweight pCN samples by 39 to get a comparison of the sampling accuracy per computational unit. Figure 5 shows the convergence of total variation norm for chains of runtime equal to 10,000 samples of HMC; the results can be compared with Figure 4. We see that chains of equal runtime are largely equivalent between the two methods when applied to Example 1; the faster convergence of HMC is essentially balanced by the larger amount of computation required to generate the samples.

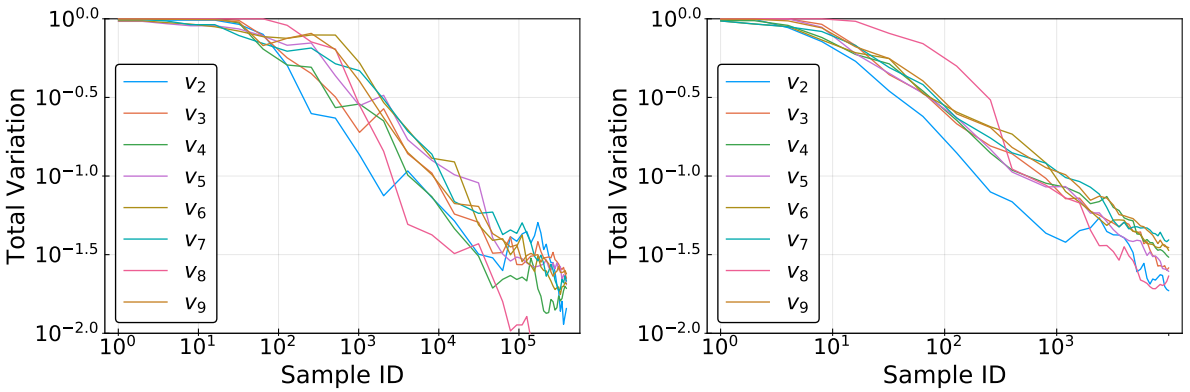

Figure 5. *Total variation norm between computed and "true" marginal probability density function for* $v_2, \ldots, v_9$ *for runtime equivalent to* 10,000 *HMC samples. Left: pCN. Right: HMC.*

4.2. Example 2: Multimodal posterior. In this section we present an example where the prior and data interact to produce a posterior with multiple regions of mass; posteriors of this kind are difficult for MCMC methods to resolve because chains have trouble jumping between the wells. We take the initial condition $\theta_0(\mathbf{x}) = \frac{1}{2} - \frac{1}{4}\cos 2\pi x - \frac{1}{4}\cos 2\pi y$ and true background flow $\mathbf{v}^\star = [8\cos 2\pi y, 8\cos 2\pi x]$. Symmetry guarantees that for $\mathbf{x}_1 = [0,0]$ and $\mathbf{x}_2 = [\frac{1}{2}, \frac{1}{2}]$ we have $\theta(\mathbf{v}^\star, t, \mathbf{x}_i) = \theta(-\mathbf{v}^\star, t, \mathbf{x}_i)$, $i = 1, 2$. (In fact, there are more points for which this is true; however, two points suffice for the purposes of this example.) We therefore let the data $\mathcal{Y}$ be point measurements $\theta(t, \mathbf{x})$ from $t = 0.001$ to 0.050 in intervals of 0.001 at each of $\mathbf{x}_1$ and $\mathbf{x}_2$. Then we have $\Phi(\mathbf{v}^\star) = \Phi(-\mathbf{v}^\star)$, i.e., both $\mathbf{v}^\star$ and $-\mathbf{v}^\star$ match the data equally well. Finally, we use the mean-zero Kraichnan prior (4.2), which assigns the same probability to both $\mathbf{v}^\star$ and $-\mathbf{v}^\star$. The problem parameters for this example are listed in Table 2.

Since both $\mathbf{v}^\star$ and $-\mathbf{v}^\star$ are given the same probability by both the prior and the data, they will be equally likely according to the posterior. We show in the next section that the symmetry in the problem setup results in multiple distinct probability masses in the posterior.

Table 2
Problem parameters for Example 2.

Parameter	Value	Parameter	Value
Observation operator, $\mathcal{O}$	Point observations at $\mathbf{x}_1 = [0,0]$ and $\mathbf{x}_2 = [\frac{1}{2}, \frac{1}{2}]$	Data, $\mathcal{Y}$	$\mathcal{G}(\mathbf{v}^\star)$
Prior, μ_0	Kraichnan (4.2)	Noise, γ_0	$N(0, \sigma_\eta^2 I)$, $\sigma_\eta = 2^{-3}$
True flow, $\mathbf{v}^\star$	$[8\cos 2\pi y, 8\cos 2\pi x]$	Sampling space, H_N	$\|\mathbf{k}\|_2 \leq 8$ (197 components)
Diffusion, κ	3×10^{-5} [17, Table I]	θ_0	$\frac{1}{2} - \frac{1}{4}\cos(2\pi x) - \frac{1}{4}\cos(2\pi y)$

4.2.1. Posterior structure. As in Example 1, we approximate the exact posterior via a large number of samples; in this case we use 500,000 samples generated from 100 HMC chains of 5,000 samples apiece, each beginning with an initial sample randomly chosen from the prior measure. We chose HMC chains because they provided better convergence to the posterior than the other methods, as we describe below. Figure 6 shows the resulting posterior structure. The left plot shows the computed mean, variance, skew, and excess kurtosis of the posterior, by Fourier component of $\mathbf{v}$. We note that, due to the influence of the prior measure, the mean and covariance of the posterior tend to zero for higher-order components. Also, the deviations of excess kurtosis from zero indicate the presence of highly non-Gaussian marginal distributions for some components. The plot on the right presents one- and two-dimensional histograms for the first few components of $\mathbf{v}$ (see the expansion in (4.3) for interpretation of the components). Note that the symmetry of the problem results in multiple large modes both v_2 and v_4, as well as in several smaller bumps in the distributions of the other components.

Moreover, in contrast to Example 1, the two-dimensional histograms—the approximate posterior joint probability density of pairs of vector field components—show that the vector field components are highly correlated with each other (see, e.g., the "X" shape between v_3 and v_8). It is worth noting that the posterior contains these correlation structures even though the prior assumes independence of the components.

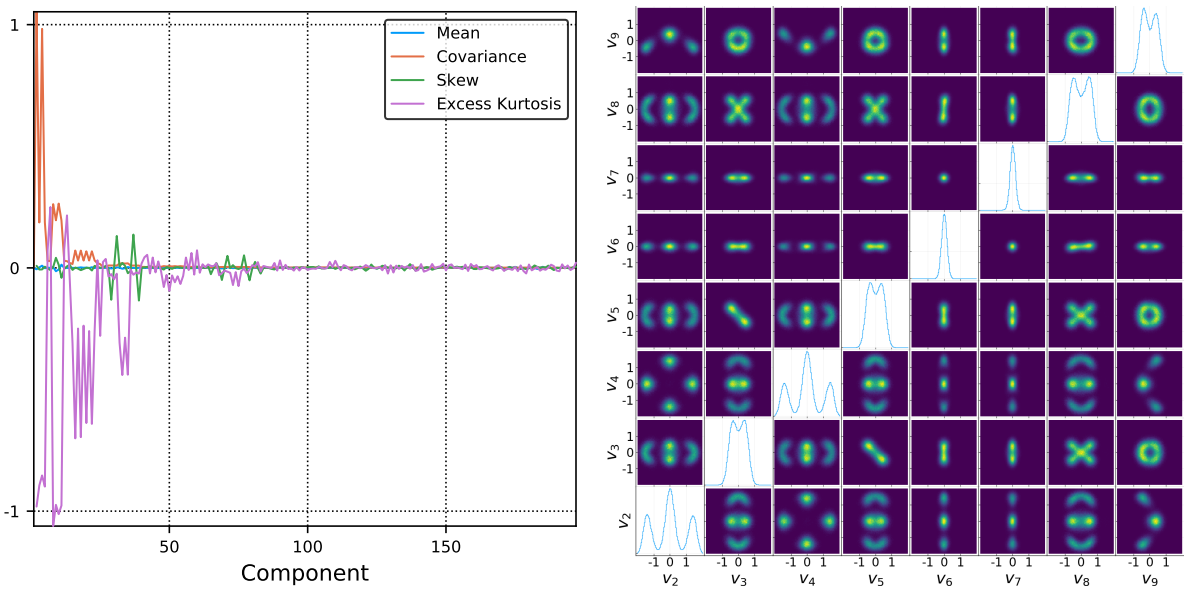

Figure 6. *Structure of posterior. Left: Mean, covariance, skew, and excess kurtosis of posterior measure, by component of* $\mathbf{v}$. *Right: Posterior* ($\mu_{\mathcal{Y}}$) *one-dimensional (diagonal) and two-dimensional (off-diagonal) marginal distributions for the first eight components of* $\mathbf{v}$ *(out of* 197*).*

Finally, it is worth noting that not all observables of the posterior exhibit complicated structures. Figure 7 shows the computed posterior one- and two-dimensional histograms of background flow vorticity at nine observation locations. The one-dimensional histograms are simple—i.e., nearly Gaussian—at each point. However, the two-dimensional histograms (except at the center point, $\mathbf{x}_5$) exhibit multiple modes of different shapes.

4.2.2. MCMC sampling. The multimodal structure of the posterior is typical of distributions that are difficult for MCMC methods to resolve efficiently, as the chains have difficulty moving across the regions of low probability between the regions of mass. We now use this structure to test the viability of pCN and HMC (see subsection 3.1) in resolving complicated posteriors. The tests use parameter values of $\beta = 0.2$ for pCN (again corresponding to the optimal acceptance rate of 23% from [48]) and $\epsilon = 0.125$ and $\tau = 4$ for HMC, which in numerical experiments showed good convergence behavior.

Figure 8 shows the trace and autocorrelation of the potential Φ (see Definition 2.10) for MCMC sampling of Example 2. As in Example 1, we see "random walk" behavior for pCN, whereas for HMC many fewer iterations are required to achieve statistical independence between samples. Unlike Example 1, however, the HMC chains for Example 2 exhibit negative autocorrelation between consecutive samples. The authors plan to investigate this phenomenon, which may be related to the multimodal structure of the problem, in later work.

We can also see the contrast between pCN and HMC in the traces of vector field components shown in Figure 9. The pCN samples move within a relatively limited range (a single probability mass), while the HMC samples occasionally jump between the different probability regions. In parameter testing, we observed that the frequency of these jumps increased

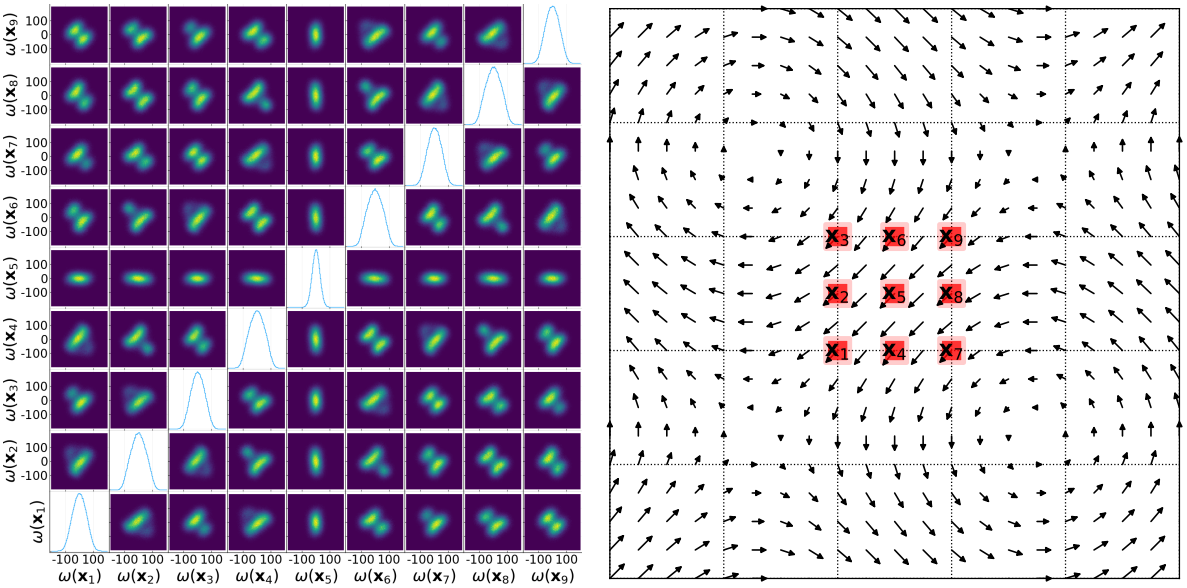

Figure 7. *Posterior one- and two-dimensional histograms of vorticity (left) at nine observation points (shown against* $\mathbf{v}^\star$*, right).*

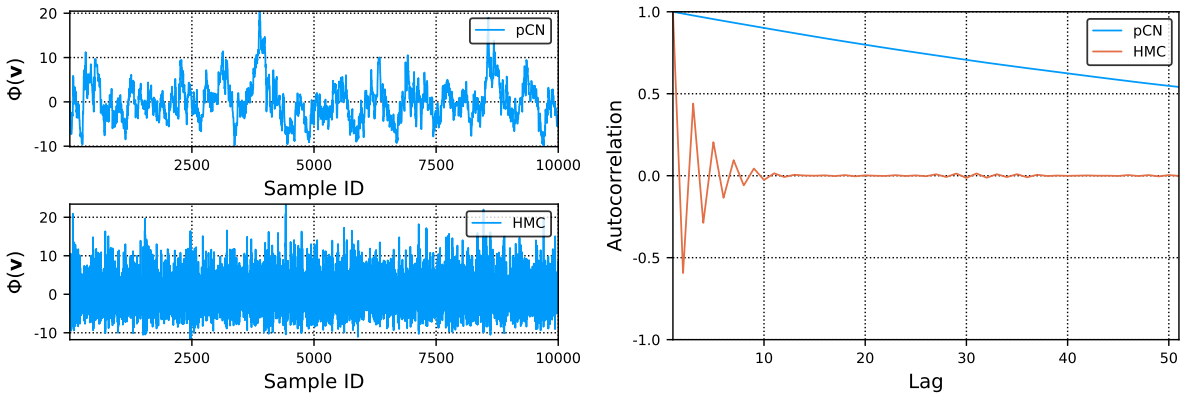

Figure 8. *Trace (left) and autocorrelation (right) of the potential* Φ*.*

roughly linearly with τ ($\tau = 4$ produced twice as many jumps as $\tau = 2$, for example) because longer integration times allowed the Hamiltonian system to evolve further, overcoming the areas of low probability that separate the regions of mass.

4.2.3. Convergence of measures. We now consider how the computed probability density functions (normalized histograms) compare for each of the MCMC methods. Figure 10 shows the computed one- and two-dimensional distributions for the first few vector field components after 150,000 samples. This figure can be compared with the "true" distributions shown in Figure 6. The histograms for pCN show only some of the many distinct probability regions in the posterior, as the chain failed to jump across the regions of low probability. The distributions for v_2 and v_4 in particular show only one or two of the three modes shown

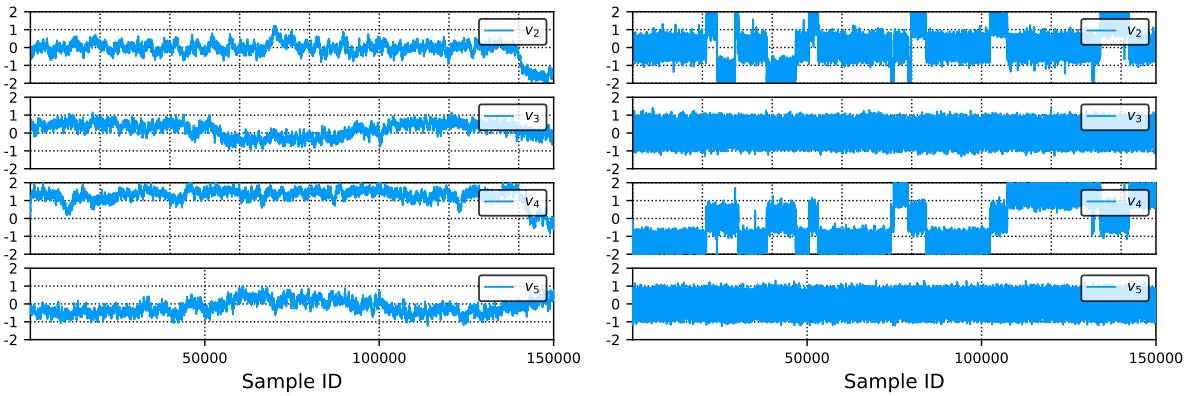

Figure 9. *Trace of $v_2, \ldots, v_5$ by sample number, pCN (left) and HMC (right).*

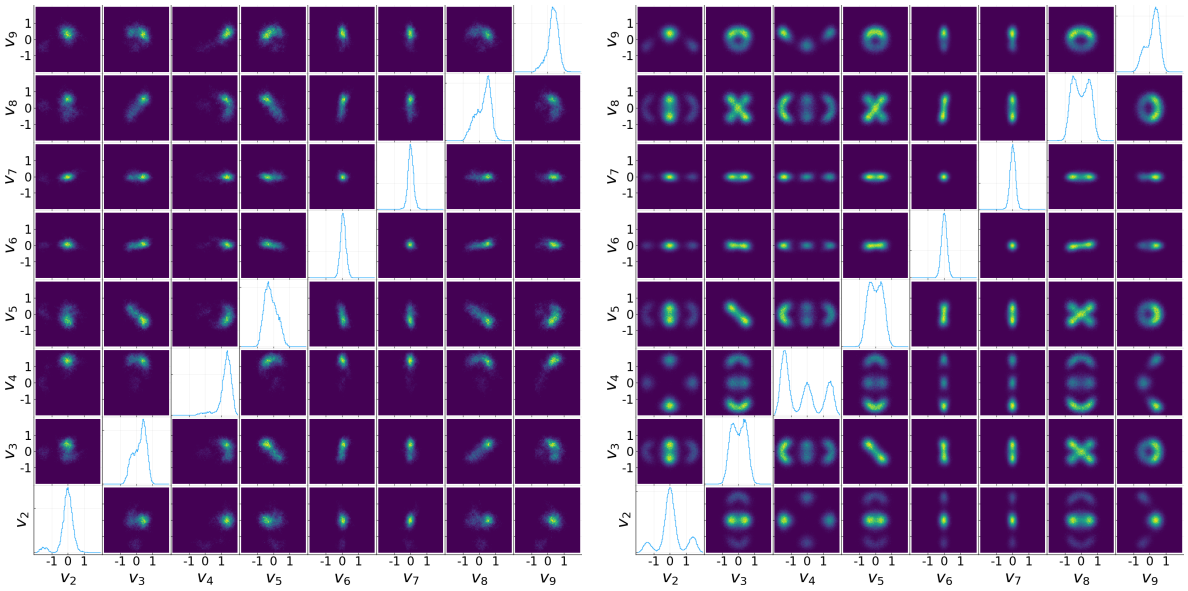

Figure 10. *Computed one- and two-dimensional marginal distributions for each of the first eight vector field components (out of* 197*) for* 150,000 *samples, pCN (left) and HMC (right).*

in Figure 6. The histograms for HMC, by contrast, resolve all major features in the posterior, though some of the features still exhibited imbalance when the chain terminated. The asymmetry in v_4 and v_9 indicates that the chain has perhaps not fully converged yet.

As in Example 1, we can get a feel for convergence of the methods by computing the total variation distance between the "true" μ and N-sample $\mu^{(N)}$ marginal distributions. The results are shown in Figure 11. We note the "sawtooth" behavior for HMC, as the number of samples in each mode of, for example, v_4 slowly balances.

4.2.4. Equal runtime comparison. It is worth noting that, due to the selected values of the parameters ϵ and τ, the HMC method used above required 32 PDE and 32 adjoint solves per sample, making it quite expensive relative to a sample of pCN. In our implementation, we

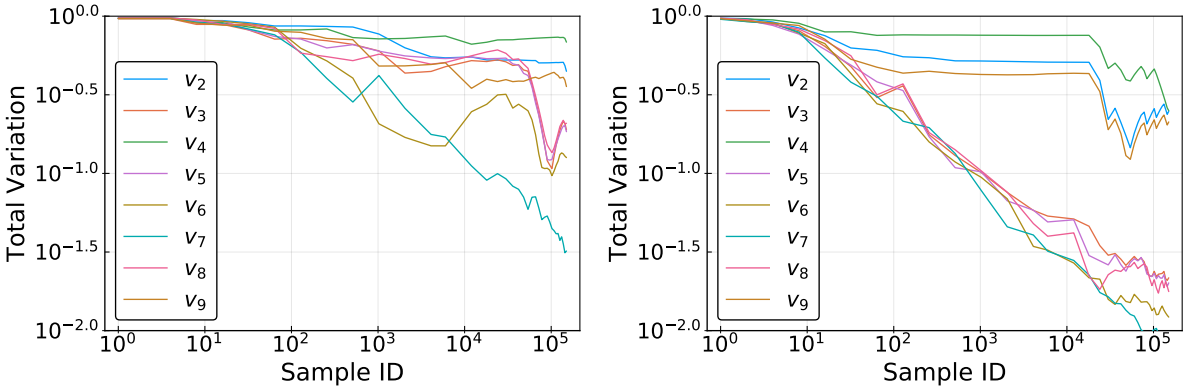

Figure 11. *Total variation norm between computed (*150,000 *samples) and "true" marginal probability density function for* $v_2, \ldots, v_9$*, pCN (left) and HMC (right).*

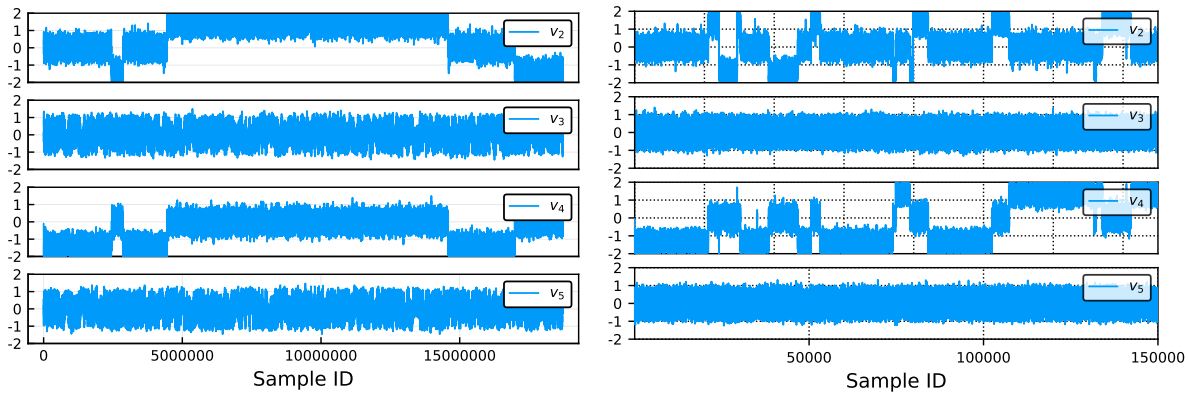

Figure 12. *Trace of* $v_2, \ldots, v_5$ *by sample number for runtime equivalent to* 150,000 *HMC samples, pCN (left) and HMC (right).*

were able to compute 125 pCN samples for each HMC sample. As a result, almost 19 million pCN samples could be computed in the time required to generate 150,000 HMC samples. For comparison in terms of equal computational cost, we now present the results of a single 19 million–sample pCN chain with the 150,000-sample HMC run shown earlier. Figure 12 compares the trace of the first few vector field components by sample number. We see that pCN does eventually achieve the jumps between states that HMC shows; however, the jumps are much less frequent for pCN, even when weighted by runtime, than for HMC (approximately 5 jumps vs. 20 jumps, respectively).

Figure 13 shows the computed one- and two-dimensional histograms for the 19 million–sample pCN chain, which can be compared with the HMC figure in Figure 10. Figure 14 compares the evolution of the total variation norm between the computed and "true" two-dimensional distributions for pCN (blue) and HMC (orange) chains of equal runtime. (Convergence of the (v_2, v_9) correlation structure, for example, is shown in the bottom right subplot.) These two plots are more equivocal between the two methods. pCN produces better convergence for histograms with one probability mode (e.g., the pair (v_6, v_7)); for these components,

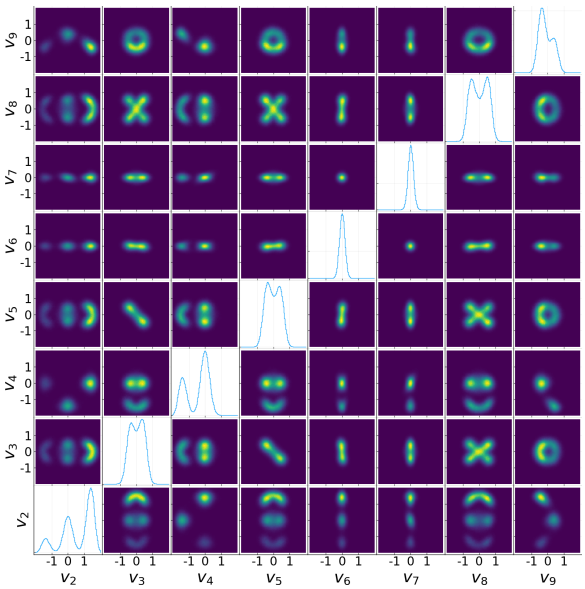

Figure 13. *Computed one- and two-dimensional marginal distributions for each of the first eight vector field components (out of* 197*) for* 19 *million samples of pCN (same runtime as* 150,000 *samples of HMC).*

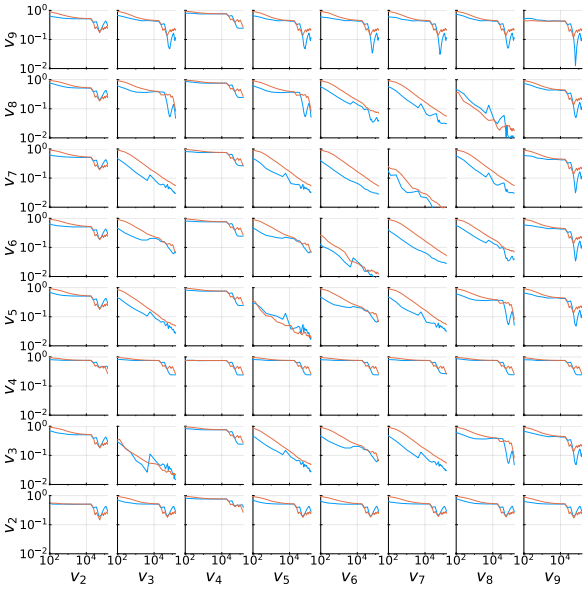

Figure 14. *Total variation norm between computed and "true" two-dimensional probability density for pairs of vector field components for pCN (blue) and HMC (orange) of equal computational time.*

the extra computations involved in HMC (which was tuned for larger jumps) do not appear to provide any benefit. However, the two methods exhibited very similar convergence for multimodal distributions (e.g., those associated with v_2 or v_4). HMC also reached all of the modes in the distribution, while pCN generated no samples in the mode near $v_4 \approx 1.5$. Overall, it appears that pCN did a better job (on an equal-runtime basis) of sampling within modes, while HMC did a better job of finding modes.

4.3. Summary of numerical experiments. To summarize the results of Examples 1 and 2, we see trade-offs between the MCMC methods:

- pCN provides an inexpensive method to generate samples and explore local regions of a probability measure, with a free parameter β that can be tuned to the problem.
- HMC samples are more computationally expensive to generate; for the HMC test cases reported here, each HMC sample took 39–125 times as much time as one pCN sample, though in general this ratio will be dictated by the cost of the gradient computation and the choice of number of integration steps $\frac{\tau}{\epsilon}$.
- In our numerical experiments, for posterior distributions with simple structure (e.g., Example 1 or some parts of Example 2), pCN exhibits similar (as measured by equal number of samples) or better (as measured by equal runtime) convergence to HMC.
- In our numerical experiments, for posterior distributions with more complicated structure (some components in Example 2), pCN still appeared to do a better job (for equal runtime) of sampling within probability modes, while HMC appeared to do a better job of jumping between states to find new modes. The overall impact on performance

is difficult to discern and will likely depend heavily on the desired observables.

- Finally, we note that implementation of HMC is much more involved than pCN. With a working PDE solver, pCN can be implemented in a matter of minutes or hours. Developing a gradient solver and implementing the HMC leapfrog integration (and debugging both)—if even possible—can require on the order of days or months of development time, depending on the complexity of the PDE solver.

Acknowledgments. We would like to thank G. Didier, J. Foldes, S. McKinley, C. Pop, G. Richards, G. Simpson, A. Stuart, and J. Whitehead for many useful discussions and thoughtful feedback on this work. The authors acknowledge Advanced Research Computing at Virginia Tech[5] for providing computational resources and technical support that have contributed to the results reported within this paper.

REFERENCES

[1] V. Akçelik, G. Biros, O. Ghattas, K. R. Long, and B. van Bloemen Waanders, *A variational finite element method for source inversion for convective–diffusive transport*, Finite Elements Anal. Des., 39 (2003), pp. 683–705.

[2] A. Attia, R. Ştefănescu, and A. Sandu, *The reduced-order hybrid Monte Carlo sampling smoother*, Internat. J. Numer. Methods Fluids, 83 (2017), pp. 28–51.

[3] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, and A. M. Stuart, *Geometric MCMC for infinite-dimensional inverse problems*, J. Comput. Phys., 335 (2017), pp. 327–351.

[4] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart, *Hybrid Monte Carlo on Hilbert spaces*, Stochastic Process. Appl., 121 (2011), pp. 2201–2230.

[5] A. Beskos, G. Roberts, and A. Stuart, *Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions*, Ann. Appl. Probab., 19 (2009), pp. 863–898.

[6] A. Beskos, G. Roberts, A. Stuart, and J. Voss, *MCMC methods for diffusion bridges*, Stoch. Dyn., 8 (2008), pp. 319–350.

[7] A. Beskos and A. Stuart, *MCMC methods for sampling function space*, in Invited Lectures, Sixth International Congress on Industrial and Applied Mathematics, ICIAM07, R. Jeltsch and G. Wanner, eds., European Mathematical Society, 2009, pp. 337–364.

[8] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, *Julia: A fresh approach to numerical computing*, SIAM Rev., 59 (2017), pp. 65–98, https://doi.org/10.1137/141000671.

[9] J. Borggaard, N. Glatt-Holtz, and J. Krometis, *GPU-accelerated particle methods for evaluation of sparse observations for inverse problems constrained by diffusion PDEs*, J. Comput. Phys., 391 (2019), pp. 142–154, https://doi.org/10.1016/j.jcp.2019.04.034.

[10] J. Borggaard, N. Glatt-Holtz, and J. Krometis, *On Bayesian consistency for flows observed through a passive scalar*, Ann. App. Probab., 30 (2020), pp. 1762–1783, https://doi.org/10.1214/19-AAP1542.

[11] N. Bou-Rabee and J. M. Sanz-Serna, *Geometric integrators and the Hamiltonian Monte Carlo method*, Acta Numer., 27 (2018), pp. 113–206.

[12] T. Bui-Thanh and O. Ghattas, *Analysis of the Hessian for inverse scattering problems:* I. *Inverse shape scattering of acoustic waves*, Inverse Problems, 28 (2012), 055001.

[13] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional Bayesian inverse problems, Part* I: *The linearized case, with application to global seismic inversion*, SIAM J. Sci. Comput., 35 (2013), pp. A2494–A2523, https://doi.org/10.1137/12089586X.

[14] T. Bui-Thanh and M. Girolami, *Solving large-scale PDE-constrained Bayesian inverse problems with Riemann manifold Hamiltonian Monte Carlo*, Inverse Problems, 30 (2014), 114014.

[5] http://www.arc.vt.edu

[15] C. Canuto, M. Hussaini, A. Quarteroni, and T. Zang, *Spectral Methods: Fundamentals in Single Domains*, Scientific Computation, Springer, 2007.
[16] Y. Cao, S. Li, L. Petzold, and R. Serban, *Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution*, SIAM J. Sci. Comput., 24 (2003), pp. 1076–1089, https://doi.org/10.1137/S1064827501380630.
[17] S. Chen and R. H. Kraichnan, *Simulations of a randomly advected passive scalar field*, Phys. Fluids, 10 (1998), pp. 2867–2884.
[18] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White, *MCMC methods for functions: Modifying old algorithms to make them faster*, Statist. Sci., 28 (2013), pp. 424–446.
[19] E. L. Cussler, *Diffusion: Mass Transfer in Fluid Systems*, Cambridge University Press, 2009.
[20] M. Dashti and A. M. Stuart, *The Bayesian approach to inverse problems*, in Handbook Uncertainty Quantification, Springer, 2017, pp. 311–428.
[21] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B, 195 (1987), pp. 216–222.
[22] A. Eberle, *Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions*, Ann. Appl. Probab., 24 (2014), pp. 337–377.
[23] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., CRC Press, 2014.
[24] M. Girolami and B. Calderhead, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, J. R. Stat. Soc. Ser. B Stat. Methodol., 73 (2011), pp. 123–214.
[25] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications*, CBMS-NSF Reg. Conf. Ser. in Appl. Math. 26, SIAM, 1977, https://doi.org/10.1137/1.9781611970425.
[26] M. Hairer and J. C. Mattingly, *Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations*, Ann. Probab., 36 (2008), pp. 2050–2091.
[27] M. Hairer and J. C. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Springer, 2011, pp. 109–117.
[28] M. Hairer, J. C. Mattingly, and M. Scheutzow, *Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations*, Probab. Theory Related Fields, 149 (2011), pp. 223–259.
[29] M. Hairer, A. M. Stuart, and S. J. Vollmer, *Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions*, Ann. Appl. Probab., 24 (2014), pp. 2455–2490.
[30] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109.
[31] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Math. Model. Theory Appl. 23, Springer, 2009.
[32] M. D. Hoffman and A. Gelman, *The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo*, J. Mach. Learn. Res., 15 (2014), pp. 1593–1623.
[33] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160, Springer Science & Business Media, 2005.
[34] G. K. Karch, F. Sadlo, D. Weiskopf, C. D. Hansen, G. S. Li, and T. Ertl, *Dye-based flow visualization*, Comput. Sci. Engrg., 14 (2012), pp. 80–86, https://doi.org/10.1109/MCSE.2012.118.
[35] H. Kellay and W. I. Goldburg, *Two-dimensional turbulence: A review of some recent experiments*, Rep. Progr. Phys., 65 (2002), pp. 845–894.
[36] S. Kou, Q. Zhou, and W. H. Wong, *Equi-energy sampler with applications in statistical inference and statistical mechanics*, Ann. Statist., 34 (2006), pp. 1581–1619.
[37] R. H. Kraichnan, *Inertial ranges in two-dimensional turbulence*, Phys. Fluids, 10 (1967), pp. 1417–1423.
[38] R. H. Kraichnan, *Small-scale structure of a scalar field convected by turbulence*, Phys. Fluids, 11 (1968), pp. 945–953.
[39] R. H. Kraichnan, *Stochastic modeling of isotropic turbulence*, in New Perspectives in Turbulence, Springer, 1991, pp. 1–54.
[40] J. Krometis, *A Bayesian Approach to Estimating Background Flows from a Passive Scalar*, Ph.D. thesis, Virginia Polytechnic Institute and State University, 2018.
[41] J. C. Mattingly, N. S. Pillai, and A. M. Stuart, *Diffusion limits of the random walk Metropolis*

algorithm in high dimensions, Ann. Appl. Probab., 22 (2012), pp. 881–930.
[42] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.
[43] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Comm. Control Engrg. Ser., Springer-Verlag, 1993.
[44] K. W. Morton, *Numerical Solution of Convection-Diffusion Problems*, Chapman & Hall, 1996.
[45] N. Petra, J. Martin, G. Stadler, and O. Ghattas, *A computational framework for infinite-dimensional Bayesian inverse problems, Part* II: *Stochastic Newton MCMC with application to ice sheet flow inverse problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1525–A1555, https://doi.org/10.1137/130934805.
[46] N. S. Pillai, A. M. Stuart, and A. H. Thiéry, *Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions*, Ann. Appl. Probab., 22 (2012), pp. 2320–2356.
[47] V. Rao and A. Sandu, *A posteriori error estimates for the solution of variational inverse problems*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 737–761, https://doi.org/10.1137/140990036.
[48] G. O. Roberts and J. S. Rosenthal, *Optimal scaling for various Metropolis-Hastings algorithms*, Statist. Sci., 16 (2001), pp. 351–367.
[49] J. C. Robinson, *Infinite-dimensional Dynamical Systems: An Introduction to Dissipative Parabolic PDEs and the Theory of Global Attractors*, Cambridge University Press, 2001.
[50] B. I. Shraiman and E. D. Siggia, *Scalar turbulence*, Nature, 405 (2000), pp. 639–646.
[51] A. J. Smits, *Flow Visualization: Techniques and Examples*, World Scientific, 2012.
[52] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
[53] M. Stynes, *Numerical Methods for Convection-Diffusion Problems or the* 30 *Years War*, preprint, https://arxiv.org/abs/1306.5172, 2013.
[54] R. Temam, *Navier–Stokes Equations and Nonlinear Functional Analysis*, 2nd ed., CBMS-NSF Regional Conf. Ser. in Appl. Math. 66, SIAM, 1995, https://doi.org/10.1137/1.9781611970050.
[55] L. Tierney, *A note on Metropolis-Hastings kernels for general state spaces*, Ann. Appl. Probab., 8 (1998), pp. 1–9.
[56] S. J. Vollmer, *Dimension-independent MCMC sampling for inverse problems with non-Gaussian priors*, SIAM/ASA J. Uncertain. Quantif., 3 (2015), pp. 535–561, https://doi.org/10.1137/130929904.
[57] B. S. Williams, D. Marteau, and J. P. Gollub, *Mixing of a passive scalar in magnetically forced two-dimensional turbulence*, Phys. Fluids, 9 (1997), pp. 2061–2080.
[58] M. Wolfgang, *Flow Visualization*, Academic Press, 1987.
[59] J. Worthen, G. Stadler, N. Petra, M. Gurnis, and O. Ghattas, *Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow*, Phys. Earth Planetary Interiors, 234 (2014), pp. 23–34.
[60] S. Zhuk, T. T. Tchrakian, S. Moore, R. Ordóñez-Hurtado, and R. Shorten, *On source-term parameter estimation for linear advection-diffusion equations with uncertain coefficients*, SIAM J. Sci. Comput., 38 (2016), pp. A2334–A2356, https://doi.org/10.1137/15M1034829.