A Gaussian Mixture Model Smoother for Continuous Nonlinear Stochastic Dynamical Systems: Applications

TAPOVAN LOLLA AND PIERRE F. J. LERMUSIAUX

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 18 February 2016, in final form 11 December 2016)

ABSTRACT

The nonlinear Gaussian Mixture Model Dynamically Orthogonal (GMM–DO) smoother for high-dimensional stochastic fields is exemplified and contrasted with other smoothers by applications to three dynamical systems, all of which admit far-from-Gaussian distributions. The capabilities of the smoother are first illustrated using a double-well stochastic diffusion experiment. Comparisons with the original and improved versions of the ensemble Kalman smoother explain the detailed mechanics of GMM–DO smoothing and show that its accuracy arises from the joint GMM distributions across successive observation times. Next, the smoother is validated using the advection of a passive stochastic tracer by a reversible shear flow. This example admits an exact smoothed solution, whose derivation is also provided. Results show that the GMM–DO smoother accurately captures the full smoothed distributions and not just the mean states. The final example showcases the smoother in more complex nonlinear fluid dynamics caused by a barotropic jet flowing through a sudden expansion and leading to variable jets and eddies. The accuracy of the GMM–DO smoother is compared to that of the Error Subspace Statistical Estimation smoother. It is shown that even when the dynamics result in only slightly multimodal joint distributions, Gaussian smoothing can lead to a severe loss of information. The three examples show that the backward inferences of the GMM–DO smoother are skillful and efficient. Accurate evaluation of Bayesian smoothers for nonlinear high-dimensional dynamical systems is challenging in itself. The present three examples—stochastic low dimension, reversible high dimension, and irreversible high dimension—provide complementary and effective benchmarks for such evaluation.

1. Introduction

Smoothing is a process of estimating the history of a system by integrating its dynamics with all available measurements across time, both past and future. In contrast, filtering is a process of estimating the current state of a dynamical system using only the past measurements. The theoretical basis of smoothing has been well established for some time (e.g., see Gelb 1974; Jazwinski 2007; Särkkä 2013) and has been used in multivariate geophysical applications (e.g., see Cohn et al. 1994; Bennett 1992; Wunsch 1996; Lermusiaux et al. 2002; Cosme et al. 2012; Lolla 2016). For linear systems with Gaussian measurement noise, the Kalman smoother is optimal in a Bayesian sense (Gelb 1974). However, ocean and atmospheric dynamics are highly nonlinear and can develop far-from-Gaussian distributions (Miller et al. 1999; Lermusiaux 1999a,b). Practical smoothing schemes for such systems are commonly limited to linearized and/or low-dimensional models (e.g., see Särkkä 2013; Lolla 2016), or to schemes that respect the nonlinearities in the dynamics, but are limited to Gaussian smoother updates, such as ensemble smoothers (Lermusiaux and Robinson 1999; Evensen and Van Leeuwen 2000; Bocquet 2005; Cosme et al. 2012). Hence, the utilization of efficient and accurate smoothing in high-dimensional nonlinear models, using fully Bayesian updates, is the challenge addressed here.

In the companion manuscript (Lolla and Lermusiaux 2017), we developed the theoretical basis for the present applications and derived the fundamental equations for the Gaussian Mixture Model Dynamically Orthogonal (GMM–DO) smoother, a methodology for retrospective Bayesian nonlinear inference of high-dimensional stochastic fields governed by nonlinear dynamics. In the present study, we exemplify the nonlinear GMM–DO smoother and contrast its performance with respect to that of other smoothers and their underlying filters by applications to three dynamical systems, all of which admit far-from-Gaussian statistics. These three applications also define complementary and effective benchmarks for

Corresponding author e-mail: Pierre F. J. Lermusiaux, pierrel@mit.edu

DOI: 10.1175/MWR-D-16-0065.1

TABLE 1. GMM–DO smoother: summary of equations and algorithm.

GMM–DO smoother

a. *Forward GMM–DO filter pass*: Solve the DO equations (A3) to predict the state pdf. At each observation time t_k, perform the analysis step of the GMM–DO filter (section 4). Save the following:
1) mean vectors $\overline{\mathbf{x}}_{k|1:k}$ (filtered) and $\overline{\mathbf{x}}_{k+1|1:k}$ (forecast) for $k = 1, 2, \ldots, K-1$;
2) sets of stochastic coefficients $\{\boldsymbol{\phi}^{(r)}_{k|1:k}\}_{r=1}^{N_r}$ (filtered) and $\{\boldsymbol{\phi}^{(r)}_{k+1|1:k}\}_{r=1}^{N_r}$ (forecast) for $k = 1, 2, \ldots, K-1$;
3) matrices of modes $\boldsymbol{\mathcal{X}}_k$ for $k = 1, 2, \ldots, K$;
4) the final-time filtered variables—the stochastic coefficients $\{\boldsymbol{\phi}^{(r)}_{K|1:K}\}_{r=1}^{N_r}$ and the mean vector $\overline{\mathbf{x}}_{K|1:K}$.

b. *Joint subspaces GMM-fitting pass*: Form the realizations of $\boldsymbol{\Phi}_{k,k+1|k}(\omega)$, as per (A4): $\boldsymbol{\phi}^{(r)}_{k,k+1|1:k} = \begin{bmatrix} \boldsymbol{\phi}^{(r)}_{k|1:k} \\ \boldsymbol{\phi}^{(r)}_{k+1|1:k} \end{bmatrix}$.

Fit a GMM to each joint ensemble $\{\boldsymbol{\phi}^{(r)}_{k,k+1|1:k}\}_{r=1}^{N_r}$ using the EM–BIC procedure, to obtain the joint filtered GMMs (A5) for $k = 1, 2, \ldots, K-1$:

$$p_{\boldsymbol{\Phi}_k,\boldsymbol{\Phi}_{k+1}|\mathbf{Y}_{1:k}}(\boldsymbol{\phi}_k, \boldsymbol{\phi}_{k+1} \,|\, \mathbf{y}_{1:k}) = \sum_{j=1}^{M} \pi^j \times \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\phi}_k \\ \boldsymbol{\phi}_{k+1} \end{bmatrix}; \begin{bmatrix} \overline{\boldsymbol{\mu}}^j_{k|1:k} \\ \overline{\boldsymbol{\mu}}^j_{k+1|1:k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^j_{k,k|1:k} & \boldsymbol{\Sigma}^j_{k,k+1|1:k} \\ \boldsymbol{\Sigma}^j_{k+1,k|1:k} & \boldsymbol{\Sigma}^j_{k+1,k+1|1:k} \end{bmatrix} \right).$$

c. *Backward-smoothing pass*: Execute the following steps sequentially, starting from $k = K-1$ until $k = 1$:
1) For each $r = 1, 2, \ldots, N_r$
(i) determine the subspace conditional pdf $p_{\boldsymbol{\Phi}_k|\boldsymbol{\Phi}_{k+1},\mathbf{Y}_{1:k}}(\cdot \,|\, \boldsymbol{\phi}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k})$ from (A6)–(A7), where

$$p_{\boldsymbol{\Phi}_k|\boldsymbol{\Phi}_{k+1},\mathbf{Y}_{1:k}}(\boldsymbol{\phi}_k \,|\, \boldsymbol{\phi}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k}) = \sum_{j=1}^{M} \hat{\pi}^{j,(r)} \times \mathcal{N}(\boldsymbol{\phi}_k; \tilde{\boldsymbol{\mu}}^{j,(r)}_k, \tilde{\boldsymbol{\Sigma}}^j_k);$$

(ii) draw the sample $\tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K}$ from (A6), where $\tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K} \sim p_{\boldsymbol{\Phi}_k|\boldsymbol{\Phi}_{k+1},\mathbf{Y}_{1:k}}(\cdot \,|\, \boldsymbol{\phi}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k})$;
2) compute the smoothed mean $\overline{\mathbf{x}}_{k|1:K}$ from (A8), where $\overline{\mathbf{x}}_{k|1:K} = \overline{\mathbf{x}}_{k|1:k} + \boldsymbol{\mathcal{X}}_k \times \left(\dfrac{1}{N_r} \sum_{r=1}^{N_r} \tilde{\boldsymbol{\phi}}^{(r)} \right)$;
3) compute the zero-mean vectors of smoothed stochastic DO coefficients $\boldsymbol{\phi}^{(r)}_{k|1:K}$ from (A9), where

$$\boldsymbol{\phi}^{(r)}_{k|1:K} = \tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K} - \frac{1}{N_r} \sum_{r=1}^{N_r} \tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K}, \; r = 1, 2, \ldots, N_r;$$

4) decrement k by 1 and go to step 1).

the generic evaluation of Bayesian smoothers with higher-dimensional nonlinear dynamical models.

The questions that inspired this work include the following: What are the expected gains of a Bayesian nonlinear smoother over a Bayesian nonlinear filter? Can one obtain analytical solutions for the smoothed probability distributions of dynamic fields governed by nonlinear partial differential equations? How do GMM–DO smoother estimates compare to such analytical solutions? What is the performance of the GMM–DO smoother in more complex geophysical flows consisting of dynamic jets and eddies? How does the GMM–DO smoother compare to linear smoothers, Gaussian ensemble smoothers, or approximate Bayesian smoothers? Overall, a main objective here is to exemplify the GMM–DO smoother and its properties while addressing the above questions within the context of geophysical applications.

In its Rauch–Tung–Striebel (RTS) style forward–backward form, the equations of the GMM–DO smoother consist of three steps, summarized in Table 1. The first step is the *forward-filtering pass*, in which the Bayesian GMM–DO filter (Sondergaard and Lermusiaux 2013a) is used to assimilate observations sequentially over time. The prior probabilities are predicted forward in time using the DO differential equations (Sapsis and Lermusiaux 2009), which optimally reduce the dimensionality to a dominant, time-evolving stochastic subspace (Feppon and Lermusiaux 2017). Non-Gaussian features of the state variable are preserved during each assimilation cycle by fitting a GMM in the stochastic subspace and then analytically carrying out Bayes's law within the subspace. The second step of the GMM–DO smoother is the *joint-subspaces GMM fitting pass*. In this step, joint GMM distributions are fit to variables across all pairs of successive observation times. It is these joint-GMM fits that allow for analytical, nonlinear back propagation of information arising from future observations. The final step of the GMM–DO smoother is the *backward-smoothing pass*, in which the conditional probabilities given by the GMM–DO filter are updated by solving a Bayesian smoothing recursion equation backward in time, within the joint stochastic subspaces. The smoothed conditional probabilities in the smoothing recursion are evaluated analytically, using the joint-GMM fits from the GMM-fitting pass. Hence, the GMM–DO smoother recursively estimates the smoothed probabilities backward in time, without linearizing the dynamics, and preserves the non-Gaussian features of the stochastic field across time.

To illustrate and examine the performance of the GMM–DO smoother, we employ the following three test cases:

1) The first example is a 1D double-well stochastic diffusion experiment (section 2), in which we compare

the GMM–DO smoother to the ensemble Kalman smoother (EnKS). Standardizing the filtering skill across the two smoothers, we also compare the GMM–DO smoother to a modified Gaussian smoother that uses the GMM–DO filter in the filtering pass and the EnKS during the backward-smoothing pass. This low-dimensional example provides insight into how the GMM–DO smoother operates in a nonlinear, non-Gaussian setting and enables a direct comparison with the GMM–DO filter results (Sondergaard and Lermusiaux 2013b). It also sets the stage for the next two higher-dimensional examples.

2) In the second test case (section 3), the GMM–DO smoother is validated based on a passive tracer advected in a 2D reversible-shear flow field. This new smoothing test case is innovative in itself because it is one of the rare high-dimensional examples where the exact smoothed probability density can be derived and evaluated. This derivation is explained and the results are utilized for quantitative benchmarking and validation of the GMM–DO smoother. Additionally, we compare the GMM–DO smoother to the GMM–DO ESSE, a composite smoother that uses the GMM–DO filter in the forward pass and the Gaussian ESSE smoother (Lermusiaux and Robinson 1999; Lermusiaux 1999a) in the backward pass.
3) In the final test case (section 4), we exemplify the GMM–DO smoother using a more complex, simulated ocean flow exiting a strait–estuary with dynamic jets and eddies. As in 2), the GMM–DO smoother is compared to the GMM–DO ESSE smoother.

For each test case, we study and quantify specific properties of the GMM–DO smoother and its posterior probabilities. We also stress its equation-based and dynamic characteristics. Together, these three test cases may be used as complementary benchmarks for Bayesian smoothing schemes and also for new Bayesian filters.

In sections 2–4, we showcase the results of the GMM–DO smoother and compare them to other smoothers. The conclusions are provided in section 5. The smoothing problem and the main equations are summarized in the appendix. The notation is listed in Table 2. The GMM–DO smoother algorithm in RTS form is synthesized in Table 1.

2. Double-well stochastic diffusion experiment

In the double-well stochastic diffusion experiment, the objective is to track the position of a particle forced by a location-dependent "pseudo-gravity" and an external additive white noise. This test case admits bimodal stationary distributions and, hence, lends itself well to the assessment of schemes that aim to capture non-Gaussian features. It has been used to evaluate various filtering schemes (Sondergaard and Lermusiaux 2013b, and references therein); here, we utilize it to study different smoothers. We first illustrate the GMM–DO smoother and contrast its performance with that of the GMM–DO filter and the EnKS. Next, we compare the GMM–DO smoother with a modified EnKS, which uses the GMM–DO filter for forward filtering and the EnKS for backward smoothing.

a. Experimental setup

The position of the particle, denoted by $X(t;\omega)\in\mathbb{R}$, is governed by the nondimensional scalar stochastic differential equation:

$$dX = f(X)dt + \kappa d\Gamma(t;\omega),$$
$$f(X) = 4X - 4X^3, \quad \Gamma \sim \mathcal{N}(\cdot;0,1), \tag{1}$$

where f acts as the pseudo-gravitational force and the diffusion coefficient κ (≥ 0) controls the strength of the stochastic excitation. We are given infrequent access to direct, but noisy, measurements of the particle's location, modeled as

$$Y = X + \Upsilon, \quad \Upsilon \sim \mathcal{N}(v;0,\sigma_0^2), \tag{2}$$

that is, $p_{Y|X}(y\,|\,x) = \mathcal{N}(y;x,\sigma_0^2)$, and Υ is the measurement noise. Based on these observations, we wish to infer the entire history of the particle's location. We are thus faced with a smoothing task.

The noise-free dynamics of the particle [i.e., (1) with $\kappa = 0$] has three fixed points, each corresponding to a root of the pseudo-gravity force f. The fixed points at $x = -1$ and $x = 1$ are stable, while that at $x = 0$ is unstable. Therefore, for $\kappa = 0$, the particle will eventually settle at either $x = -1$ or $x = -1$, assuming it is not released at $x = 0$. For $\kappa \neq 0$, the stochastic process (1) forms an ergodic Markov chain and possesses a bimodal stationary distribution, well approximated by a GMM of complexity 2 (Sondergaard and Lermusiaux 2013b). The modes of this distribution are located at $x = \pm 1$. The average time of transition between the *positive* and *negative* wells (i.e., the regions $x > 0$ and $x < 0$, respectively) depends on κ. As κ is increased, the particle, on average, spends less time in any well and transitions are more frequent (Eyink and Kim 2006). A crucial aspect of the estimation is whether a given method can successfully track such transitions.

b. Test procedure

In this example, the DO methodology (A3) of dimension one captures the full state dynamics since the state space is one-dimensional (X is a single scalar). Our DO numerical solution is then equivalent

TABLE 2. Table of notation.

Symbol	Type	Description
Scalars		
t		Time
ω		Experiment number
i	$\in \mathbb{N}$	Stochastic subspace index
j	$\in \mathbb{N}$	Mixture component index
n	$\in \mathbb{N}$	Dimension of spatial coordinate
$N_{\mathbf{x}}$	$\in \mathbb{N}$	Dimension of discrete state vector $\mathbf{X}(t; \omega)$
N_{obs}	$\in \mathbb{N}$	Dimension of observation vector $\mathbf{Y}(t; \omega)$
s	$\in \mathbb{N}$	Dimension of stochastic space
N_r	$\in \mathbb{N}$	No. of (Monte Carlo) realizations
M	$\in \mathbb{N}$	No. of mixture components
r	$\in\{1,2,\ldots,N_r\}$	Realization index
K	$\in \mathbb{N}$	No. of smoothing indices
k, l	$\in\{1,2,\ldots,K\}$	Smoothing indices
t_k	$\in \mathbb{R}$	Time at smoothing index k
π_k^j	$\in \mathbb{R}^+$	jth component weight of the prior forecast GMM pdf, $p_{\mathbf{\Phi}_k\|\mathbf{Y}_{1:k-1}}$
$\hat{\pi}_k^j$	$\in \mathbb{R}^+$	jth component weight of the posterior filtered GMM pdf, $p_{\mathbf{\Phi}_k\|\mathbf{Y}_{1:k}}$
$X(\mathbf{r}, t; \omega)$	$\in \mathbb{R}$	Continuous stochastic field
$\bar{x}(\mathbf{r}, t)$	$\in \mathbb{R}$	Continuous mean field [mean of $X(\mathbf{r}, t; \omega)$]
$\tilde{x}_i(\mathbf{r}, t)$	$\in \mathbb{R}$	Continuous DO mode i: orthonormal basis for stochastic subspace
$\Phi_i(t; \omega)$	$\in \mathbb{R}$	Stochastic coefficient i
$\phi_i^{(r)}(t)$	$\in \mathbb{R}$	Realization No. r of stochastic coefficient i
Vectors		
$\mathbf{r}$	$\in \mathbb{R}^n$	Spatial coordinate
$\mathbf{\Phi}(t; \omega)$	$\in \mathbb{R}^s$	Vector of stochastic coefficients, $[\Phi_1(t; \omega), \Phi_2(t; \omega), \ldots, \Phi_s(t; \omega)]^{\mathrm{T}}$
$\mathbf{\Phi}_k(\omega)$	$\in \mathbb{R}^s$	Vector of stochastic coefficients at time t_k
$\boldsymbol{\phi}^{(r)}(t)$	$\in \mathbb{R}^s$	Realization r of the vector of stochastic coefficients, $[\phi_1^{(r)}(t), \phi_2^{(r)}(t), \ldots, \phi_s^{(r)}(t)]^{\mathrm{T}}$
$\boldsymbol{\phi}_k^{(r)}$	$\in \mathbb{R}^s$	Realization r of the random vector of stochastic coefficients at time t_k
$\mathbf{X}(t; \omega)$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	State vector [spatially discretized $X(\mathbf{r}, t; \omega)$]
$\bar{\mathbf{x}}(t)$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	Discrete mean field [mean of $\mathbf{X}(t; \omega)$]
$\tilde{\mathbf{x}}_i(t)$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	Discrete DO mode i, forming the orthonormal basis for stochastic subspace
$\bar{\mathbf{x}}^j(t)$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	Mean vector of mixture component j in state space
$\mathbf{x}^{(r)}(t)$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	Realization number r in state space
$\mathbf{x}_k^{(r)}$	$\in \mathbb{R}^{N_{\mathbf{x}}}$	Realization number r in state space at time t_k
$\bar{\boldsymbol{\mu}}^j(t)$	$\in \mathbb{R}^s$	Mean vector of mixture component j in stochastic subspace
$\mathbf{Y}$	$\in \mathbb{R}^{N_{\text{obs}}}$	Observation vector
$\mathbf{y}$	$\in \mathbb{R}^{N_{\text{obs}}}$	Realization of observation vector $\mathbf{Y}$
$\boldsymbol{\Upsilon}$	$\in \mathbb{R}^{N_{\text{obs}}}$	Observation noise
υ	$\in \mathbb{R}^{N_{\text{obs}}}$	Realization of observation noise
$\mathbf{\Psi}_{k\|1:l}$		Any random vector $\mathbf{\Psi}(t_k; \omega)$ conditioned on observations $\mathbf{Y}_{1:l}$
$\mathbf{\Psi}_{k,k+1\|1:l}$		Vector formed by augmenting $\mathbf{\Psi}_{k+1\|1:l}$ to $\mathbf{\Psi}_{k\|1:l}$
Matrices		
$\boldsymbol{\mathcal{X}}$	$\in \mathbb{R}^{N_{\mathbf{x}} \times s}$	Matrix of orthonormal DO basis vectors $[\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_s]$
$\mathbf{P}$	$\in \mathbb{R}^{N_{\mathbf{x}} \times N_{\mathbf{x}}}$	Covariance matrix in state space
$\mathbf{\Sigma}_j$	$\in \mathbb{R}^{s \times s}$	Covariance matrix of mixture component j in the stochastic subspace
$\mathbf{P}_j$	$\in \mathbb{R}^{N_{\mathbf{x}} \times N_{\mathbf{x}}}$	Covariance matrix of mixture component j in the state space
$\mathbf{R}$	$\in \mathbb{R}^{N_{\text{obs}} \times N_{\text{obs}}}$	Covariance matrix of observation noise
$\mathbf{H}$	$\in \mathbb{R}^{N_{\text{obs}} \times N_{\mathbf{x}}}$	Observation matrix
Densities		
$p_{\mathbf{\Psi}_k\|\mathbf{Y}_{1:l}}$		The pdf of the vector $\mathbf{\Psi}(t_k; \omega)$ conditioned on observations $\mathbf{Y}_{1:l}$—filtering ($k = l$), forecast ($k > l$), and smoothing ($k < l$)
$p_{\mathbf{\Psi}_k,\mathbf{\Psi}_m\|\mathbf{Y}_{1:l}}$		Joint pdf of vectors $\mathbf{\Psi}_k$ and $\mathbf{\Psi}_m$ conditioned on $\mathbf{Y}_{1:l}$
$p_{\mathbf{\Psi}_k\|\mathbf{\Psi}_{k+1},\mathbf{Y}_{1:k}}$		The pdf of $\mathbf{\Psi}_k$, conditioned on both $\mathbf{\Psi}_{k+1}$ and $\mathbf{Y}_{1:k}$

to using a standard Monte Carlo method to solve (1) and directly obtain the realizations of $X(t;\omega)$. We nonetheless continue using the terms GMM–DO filter and GMM–DO smoother to describe the results, even though there is in fact no DO dimension reduction.

For each realization (indexed by r), an Euler–Maruyama scheme (Higham 2001) is used to discretize (1) as follows:

$$x_{k+1}^{(r)} = x_k^{(r)} + f(x_k^{(r)})\Delta t + \kappa v_k^{(r)}\sqrt{\Delta t},$$
$$r = 1, 2, \ldots, N_r, \quad k = 0, 1, \ldots, K-1, \tag{3}$$

where $x_k^{(r)}$ denotes the rth Monte Carlo realization of X at time step k, Δt is the time-step size, and $v_k^{(r)}$ is white in time and is drawn from a standard normal distribution.

In the results to follow, we set the number of realizations as $N_r = 10^4$. We focus on the case where $\kappa = 0.5$ and $\sigma_0^2 = 0.25$. We then briefly summarize the effect of κ and σ_0^2. For a fair comparison, all the model runs are deterministically initialized (at time step $k = 0$) as $x_0^{(r)} = 1$, $r = 1, 2, \ldots, N_r$; that is, all Monte Carlo realizations are initially located at $x = 1$. Starting at (nondimensional) time $t = 0$, the forward-filter pass of the GMM–DO smoother is carried out until the final time $t = 40$, with a time step $\Delta t = 0.01$. Seven observations, equally spaced over time, arrive between $t = 4$ and 40. These observations are noisy measurements drawn from a "true" solution that transitions from the positive well to the negative well at time 20. This transition is central in the present test.

c. Results and discussions

Recall that at the final assimilation time, the estimates of the GMM–DO smoother and EnKS are equal to those of their respective filters. Figure 1a compares the results of the GMM–DO filter and the ensemble Kalman filter (EnKF). Superimposed on the true solution, we show the temporal mean (thick lines) and standard deviation envelope (thin lines) for each filter. Also depicted are the noisy observations and their error bars. We show only the mean and standard deviation of the GMM–DO filter so as to enable a direct comparison to the EnKF. Nonetheless, as we shall see shortly, the pdf of the particle location is far from Gaussian, which the GMM–DO equations capture. During the first three assimilation cycles, we find that the GMM–DO filter and EnKF successfully track the particle, as evident from their temporal mean and standard deviation envelopes. At $t = 20$, the true particle transitions into the negative well. Prior to assimilating the fourth observation [marked as (i) in Fig. 1a], the GMM–DO filter assigns a bimodal prior since several realizations are in the negative well. After assimilation, the GMM–DO filter assigns a larger weight to the negative Gaussian component, thereby increasing the overall variance and decreasing its mean estimate. However, because of the large observation error, neither filter means transitions at this time. Following the fifth observation, the GMM–DO filter transitions into the negative well by further increasing the weight of the negative Gaussian. This transition is confirmed as the filter reduces the variance of its estimate after the sixth observation. However, the EnKF fails to recognize the transition on account of its strong Gaussian prior (centered at the positive well) and the measurement noise, which is too large to force individual particles into the negative well.

Figure 1b depicts the results of the GMM–DO smoother and the EnKS. The backward-smoothing pass of the GMM–DO smoother is initialized as the solution of the GMM–DO filter at the final time, $t = 40$. We find that the smoother and filter solutions almost coincide between $t = 34$ and 40. The GMM–DO filter had detected the transition of the particle at $t = 28$ [marked as (ii) in Fig. 1a], albeit with low confidence. The final two observations strongly suggest that the particle remains in the negative well; the smoother thus accounts for this future information to estimate the past particle location between $t = 28$ and 34 with greater confidence than the GMM–DO filter. Arguably, the most remarkable improvement in the estimate occurs between $t = 16$ and 28, an interval that straddles the period where the particle diffuses into the negative well. Figure 1b shows that the GMM–DO smoother accurately captures this transition within the standard deviation envelope. This improvement arises because it balances the information from the past observations (which repeatedly suggest that the particle is located in the positive well) with those in the future (which indicate that the transition has taken place). The GMM–DO filter, in contrast, only detected the transition at $t = 28$, as it is oblivious to the future observations. We also confirm that the observations that arrive at after the particle's transition provide little information about its history prior to the transition. As a consequence, the GMM–DO smoother solution within the time interval $t = 0$–16 (i.e., prior to the transition) is largely similar to that of the filter.

The EnKS does not detect the transition, although its temporal mean increasingly aligns toward the center. It is clear from Fig. 1a that the primary reason for this is that the EnKS smoothing pass is initialized from the underlying EnKF solution, most of whose realizations are located in the positive well at $t = 40$. To compare the EnKS and GMM–DO smoother on an equal footing, we evaluate the performance of the GMM–DO–Filter–EnKS , a modified EnKS method

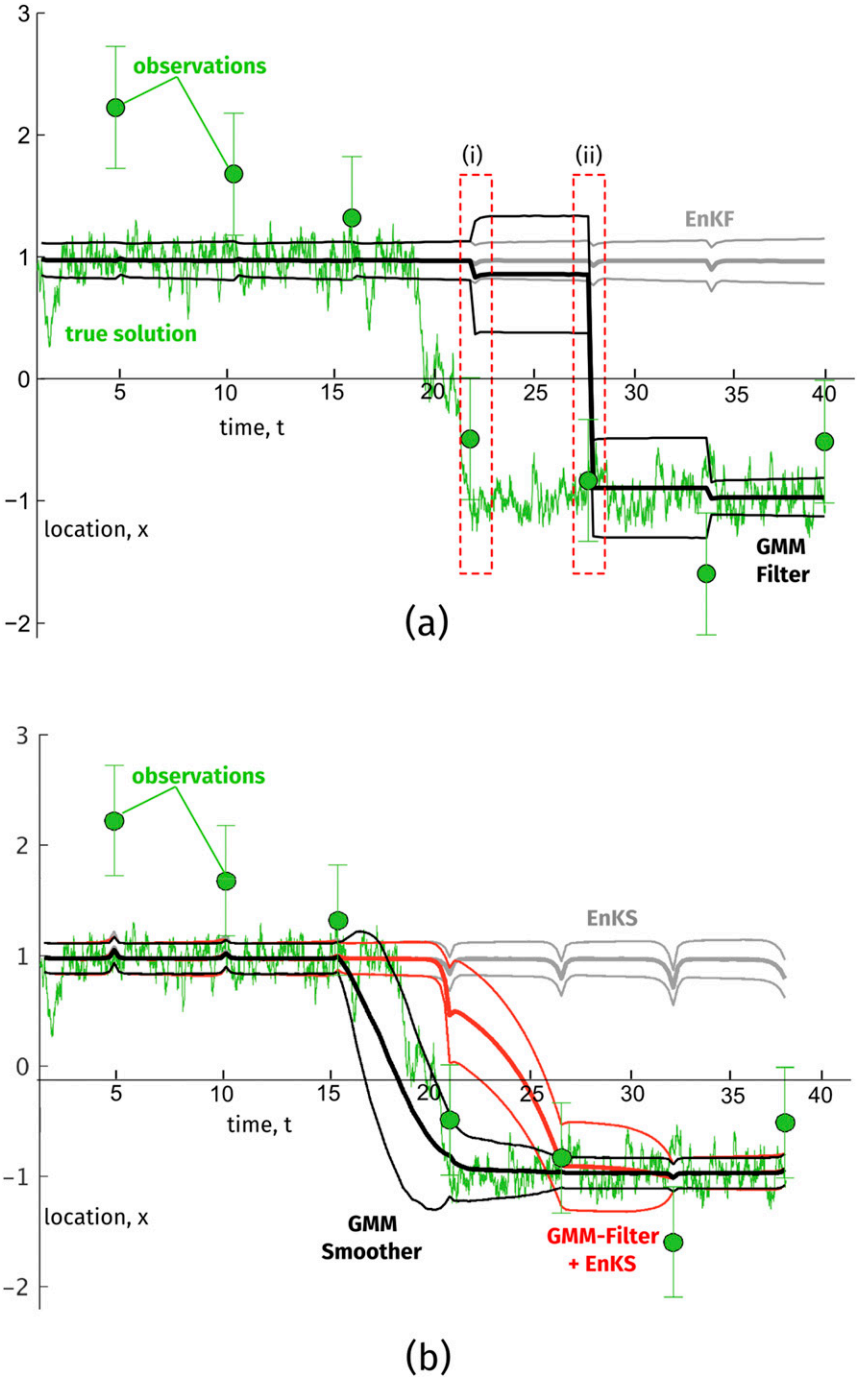

FIG. 1. The double-well stochastic diffusion experiment. (a) A comparison of the GMM–DO filter and EnKF. The time series of the particle's true location is shown in green. The mean estimates of the GMM–DO filter and the EnKF along with their standard deviation envelopes are depicted in black and gray, respectively. The instances highlighted as (i) and (ii) are discussed in section 2c. Even though only the mean and standard deviation of the GMM–DO filter are plotted, the particle's location follows a far from Gaussian distribution, as depicted in Fig. 2. (b) A comparison of the performance of the smoothers: GMM–DO smoother (black), EnKS (gray), and GMM–DO–Filter–EnKS (red). The GMM–DO smoother successfully detects the transition of the particle within its standard deviation envelope. Although having performed better than the EnKS, the GMM–DO–Filter–EnKS does not satisfactorily detect the particle's transition. The mean and standard deviation envelopes are depicted so that the GMM–DO smoother and the two EnKSs can be directly compared.

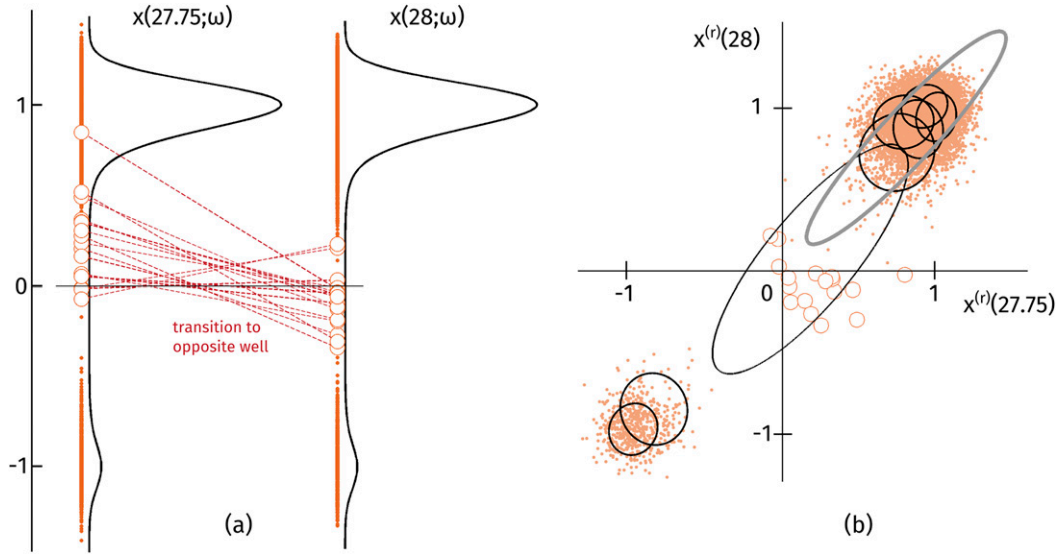

FIG. 2. A double-well stochastic diffusion experiment. Realizations and joint distributions of $X(27.75; \omega)$ and $X(28; \omega)$. Specifically, their (a) single-time marginal and (b) two-time joint distribution. The samples are shown in orange, while their GMM and Gaussian fits are, respectively, depicted in black and gray. The particles that transition across wells are depicted in larger, unshaded orange circles. Also shown in (a) is the handful of particle realizations that have transitioned across the two wells between times and $t = 27.75$ and 28.

in which the filter pass is carried out using the GMM–DO filter, in place of the EnKF. This ensures that the filtering skill is common across both methods; the two methods then differ only in their joint GMM-fitting and backward-smoothing passes. We recall that in the GMM-fitting pass, the GMM–DO smoother fits a GMM to the joint intermediate realizations using the EM–BIC criterion, whereas the GMM–DO–Filter–EnKS fits a Gaussian distribution (GMM of complexity $M = 1$).

Figure 1b also compares the results of the GMM–DO smoother and this GMM–DO–Filter–EnKS. The solutions of the two smoothers coincide between $t =$ 34 and 40. This is because in this interval a majority of realizations of the underlying GMM–DO filter are located in the negative well. In the GMM-fitting pass, it therefore suffices to use a Gaussian distribution to fit the realizations of the particle location at any two time instances in this interval (i.e., it suffices to use $M = 1$). A similar argument holds for the time interval $t = 0$–16. Between $t = 28$ and 34, the GMM–DO–Filter–EnKS increases its variance and its improvement over the filter is marginal. The most notable departure between the GMM–DO–Filter–EnKS and GMM–DO smoother is observed between $t = 16$ and 28, which is the interval of the particle's transition. The GMM–DO–Filter–EnKS, despite yielding a solution closer to the truth than the filter, is outperformed by the GMM–DO smoother. We now explain this superior performance by studying the role played by the GMM fitting in the backward-smoothing pass.

The panels in Fig. 2 show the ensemble members of the GMM–DO filter at $t = 27.75$ and 28. The samples from the joint GMM are shown in Fig. 2b and the corresponding scalar marginals in Fig. 2a. The planar GMM fit of the joint samples is indicated by black-colored ellipses in Fig. 2b and the corresponding scalar marginals of the GMM fit are shown in Fig. 2a. These figures clearly depict the bimodal nature of the particle location, which is captured by the joint-GMM fit. Only a small fraction of samples (about 10^{-4}), depicted as round unshaded markers, diffuse across into the opposite well between $t = 27.75$ and 28. This indicates a strong correlation between realizations in any well; that is, given that a particle is located in a certain well at $t = 27.75$, it is very likely to remain in the same well at $t = 28$, and vice versa. Figure 3 illustrates how the smoothed distributions evolve from $t = 28$ to 27.75 during the backward-smoothing passes of the GMM–DO smoother (continuous black curve) and GMM–DO–Filter–EnKS (continuous gray curve). We notice that at $t = 28$, a majority of samples of both smoothers are located in the negative well, although the samples of the GMM–DO–Filter–EnKS are more spread out. The GMM–DO smoother, owing to the joint-GMM fit of the filtered distributions, estimates with high confidence that the particle is located in the

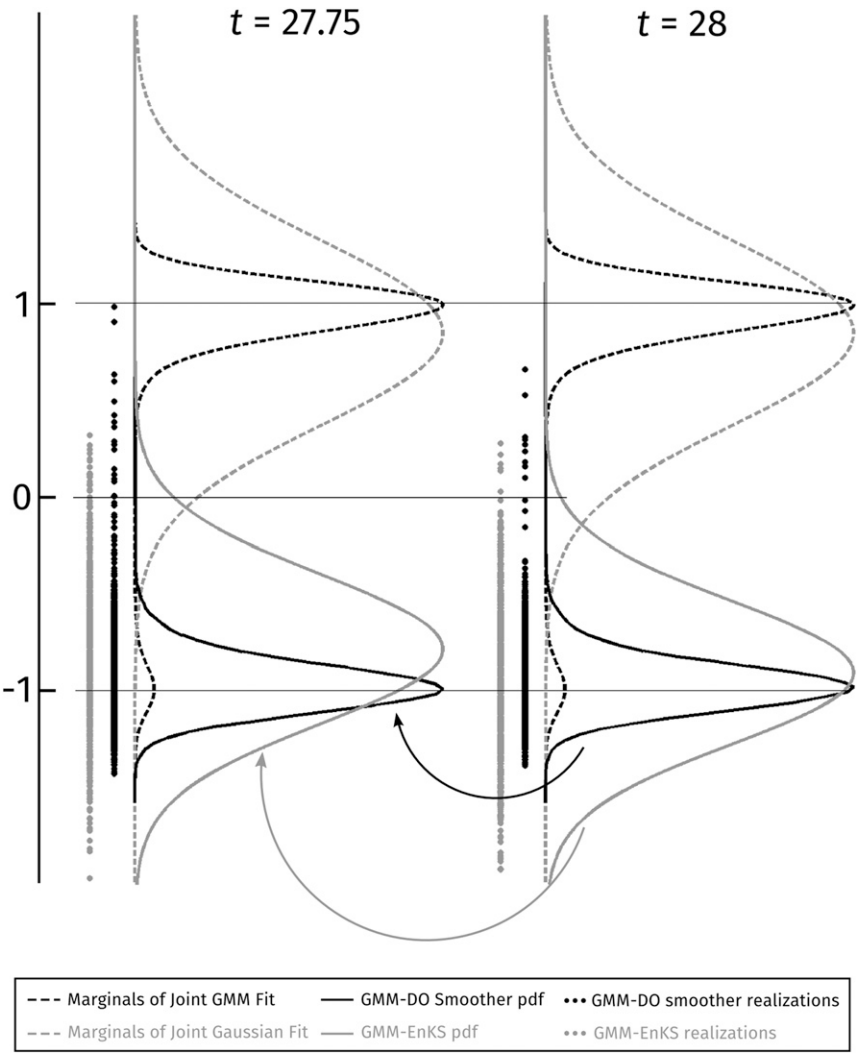

FIG. 3. Double-well stochastic diffusion experiment. An analysis of the GMM–DO smoother and the GMM–DO–Filter–EnKS distributions during the backward-smoothing pass. Owing to its Gaussian fit of the joint filtered distribution of $X(27.75; \omega)$ and $X(28; \omega)$, the GMM–DO–Filter–EnKS estimate erroneously drifts toward the positive well.

negative well at $t = 27.75$. In contrast, samples of the GMM–DO–Filter–EnKS become more aligned toward the positive well at $t = 27.75$ because the joint Gaussian fit of the filtered distributions is erroneously concentrated at the positive wells (see the marginals). This phenomenon becomes more pronounced over time and eventually leads to a striking disparity in the GMM–DO smoother and GMM–DO–Filter–EnKS solutions (see Fig. 1b).

The results discussed so far consider the effect of all seven observations on estimating the particle history. Figure 4 shows the results of the GMM–DO smoother, executed directly as soon as observations arrive through time. For the first three observations (not shown), the smoother and filter solutions are identical. The fourth observation at $t = 24$ contradicts the smoother's strong belief that the particle is located in the positive well. Consequently, the smoother only slightly aligns its estimate toward the negative well. The fifth observation is in line with the fourth, and provides more evidence of a transition. Accounting for this information, the smoother retrospectively improves the statistics of the transition. The variance of the smoother's estimate of the transition is reduced after the sixth observation is assimilated.

If the observations are sparser or noisier, or if the model noise (κ) is large, the differential between the estimates of the GMM–DO smoother and Gaussian smoothers increases. This confirms the filter results of Sondergaard and Lermusiaux (2013b), in that schemes

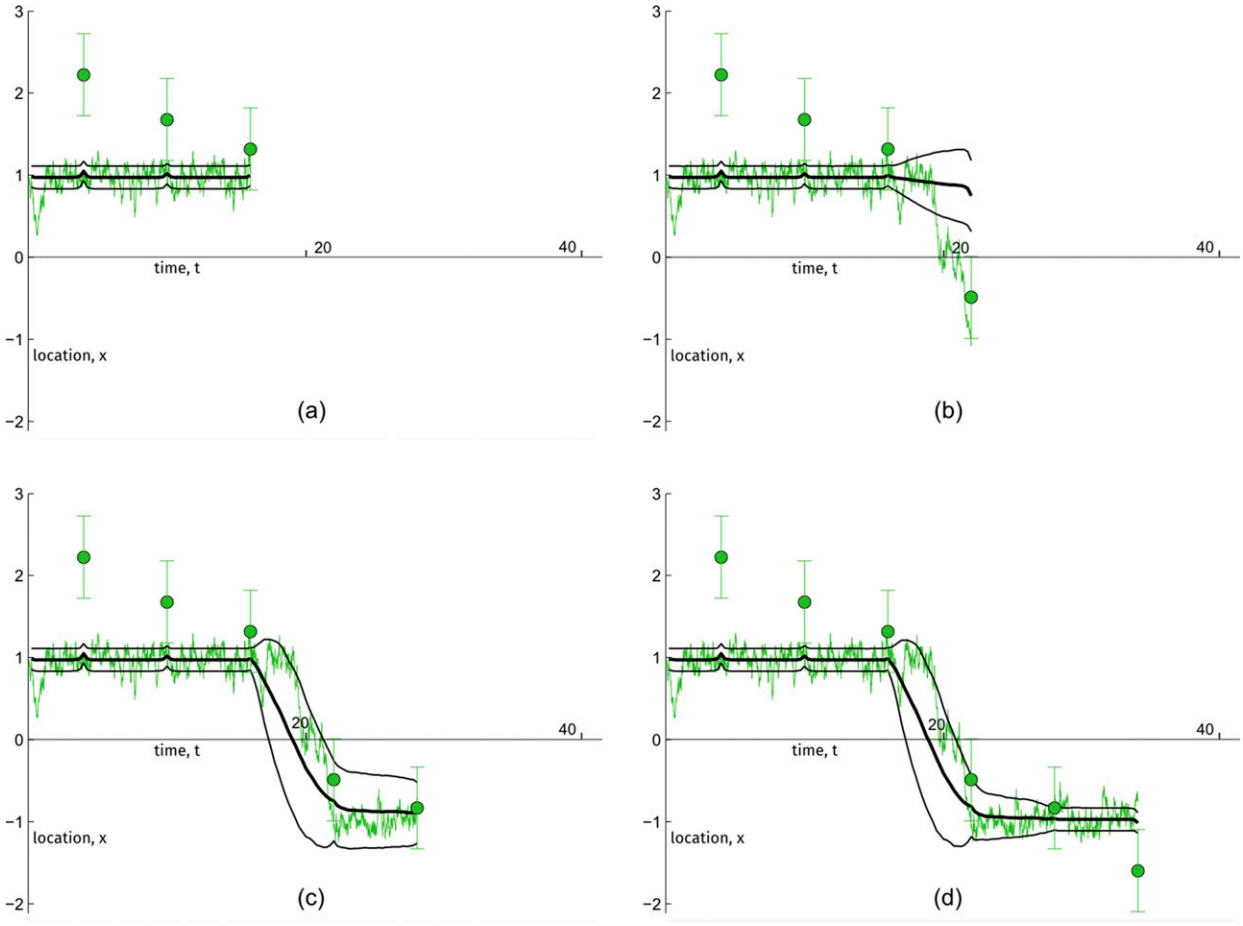

FIG. 4. Double-well stochastic diffusion experiment. The various panels depict the results of the GMM–DO smoother executed as and when observations arrive over time.

that accurately capture the pdfs are especially needed when data are sparse or noisy, which is common in oceanic and atmospheric studies. As such, if the nonlinear deterministic model is accurate (i.e., low model noise), rigorous non-Gaussian state distributions are crucial: erroneous or ad hoc approximations then severely undermine the smoothers. On the other hand, when the model uncertainty is Gaussian and large, then a Gaussian smoother suffices (Lolla 2016). Nonetheless, a key advantage of the GMM–DO smoother is that it adapts to all of these situations as they occur.

3. Passive tracer advection in a swirl flow

We now consider the passive advection of a tracer by a divergence-free, deformational flow in a unit square basin. Such test cases are often used to evaluate numerical schemes [e.g., for flux limiters, see Durran (1999)]. Here, we employ this test case because it admits an *exact non-Gaussian smoothed* solution (i.e., the Bayesian posterior) and thus allows a validation of the GMM–DO smoother. First, we show how to calculate such exact smoothed distributions for a certain class of reversible dynamical systems. Second, we evaluate the GMM–DO smoother by comparison with this exact smoothed solution. Finally, we compare the GMM–DO smoother results with these of the GMM–DO ESSE subspace smoother, a composite Gaussian smoother that uses the final GMM–DO filter estimate to initialize the backward ESSE smoother pass. In this scheme, the forward-filtering pass is thus performed using the GMM–DO filter. However, during the joint-GMM fitting pass, a mixture complexity of $M = 1$ is enforced, so as to fit a Gaussian to the joint subspace coefficients. The ESSE backward-smoothing pass is thus as that of the GMM–DO smoother but with $M = 1$. We note that if the EnKS is used instead, results are similar, but the EnKS is much more expensive than this GMM–DO ESSE since it uses a backward update for each of the realizations (here, $N_r = 10\,000$ of them, as we will see)

while the GMM–DO ESSE and GMM–DO smoothers only employ a single update directly on their distributions. Next, we start by describing the test case and its numerical schemes. As in section 2, we then simulate a true solution over a suitable time window and reconstruct the tracer's evolution based on noisy, intermittent observations of the true solution.

a. Experimental setup and test procedure

We aim to track the concentration of a tracer, denoted by $X(\mathbf{r}, t; \omega)$, defined inside the unit square basin $[0, 1]^2$. It is governed by the following advection equation:

$$\frac{\partial X(\mathbf{r},t;\omega)}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}} X(\mathbf{r},t;\omega) = 0,$$
$$\mathbf{r} = (r_x, r_y) \in (0,1)^2, \quad 0 < t < 1, \tag{4}$$

where $\mathbf{v} = (v_x, v_y)$ denotes the flow field. We use the following analytical form for $\mathbf{v}$:

$$v_x = \sin^2(\pi r_x)\sin(2\pi r_y),$$
$$v_y = -\sin^2(\pi r_y)\sin(2\pi r_x). \tag{5}$$

This $\mathbf{v}$, modified from Durran (1999), represents a counterclockwise-rotating flow, centered at (0.5, 0.5): it is divergence free and vanishes at the domain boundaries. As a consequence, the tracer is transported within, and confined to the interior of the basin. The dynamics in (4) of the tracer are noise free; that is, there is no model noise in this test case. The uncertainty in X arises only through the initial conditions, and is deterministically evolved in accordance with (4).

1) Initial conditions

The initial tracer concentration is stochastic and is given by

$$X(\mathbf{r}, 0; \omega) = \frac{1}{2}\{1 + \cos[\pi D(\mathbf{r};\omega)]\}, \tag{6}$$

where

$$D(\mathbf{r};\omega) = \max[0, 1 - 4\|\mathbf{r} - \mathbf{R}_c(\omega)\|],$$
$$\mathbf{R}_c(\omega) \sim \mathcal{N}\left(\cdot; \begin{bmatrix}0.5\\0.5\end{bmatrix}, \begin{bmatrix}0.0625 & 0\\0 & 0.0625\end{bmatrix}\right). \tag{7}$$

The random variable $\mathbf{R}_c(\omega)$ generates a distribution of the initial tracer concentration $X(\mathbf{r}, 0; \omega)$. For each realization $\mathbf{r}_c^{(r)}$ of $\mathbf{R}_c(\omega)$, the corresponding realization of the initial tracer is unity everywhere except in the interior of a circular disk of radius 0.25 centered at $\mathbf{r}_c^{(r)}$. Inside the disk, the concentration is radially symmetric, with a minima at $\mathbf{r}_c^{(r)}$. The flow field $\mathbf{v}(\mathbf{r}, t)$ and the initial tracer concentration are depicted in Fig. 5. In the double-well example (section 2), the goal was to estimate the particle's location, a scalar quantity. Consequently, we could afford to use the direct Monte Carlo method to evolve the state pdf. However, in the present example, the state vector is much larger; it has $N_{\mathbf{X}} = 10^4$ components, where $N_{\mathbf{X}}$ is the number of grid points used in the spatial discretization. Hence, we employ the DO methodology to evolve the pdf of (4).

2) DO equations for stochastic tracer advection

The stochastic tracer concentration, at any time t, is represented by the DO decomposition,

$$X(\mathbf{r}, t; \boldsymbol{\omega}) = \bar{x}(\mathbf{r}, t) + \sum_{i=1}^{s} \Phi_i(t;\omega)\tilde{x}_i(\mathbf{r}, t).$$

The DO equations in (16) for this system take the following form:

$$\frac{\partial \bar{x}(\mathbf{r},t)}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}}\bar{x} = 0, \tag{8a}$$

$$\frac{\partial \tilde{x}_i(\mathbf{r},t)}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}}\tilde{x}_i - \sum_{j=1}^{s}\langle \mathbf{v}\cdot\nabla_{\mathbf{r}}\tilde{x}_i,\ \tilde{x}_j\rangle\tilde{x}_j = 0, \quad \text{and} \tag{8b}$$

$$\frac{d\Phi_i(t;\omega)}{dt} + \sum_{j=1}^{s}\langle \mathbf{v}\cdot\nabla_{\mathbf{r}}\tilde{x}_j, \tilde{x}_i\rangle\Phi_j(t;\omega) = 0. \tag{8c}$$

3) Initialization of DO decomposition

Let the vectors $\mathbf{X}(t; \omega)$, $\bar{\mathbf{x}}(t)$, and $\tilde{\mathbf{x}}_i(t) \in \mathbb{R}^{N_{\mathbf{X}}}$ denote the spatially discretized fields of $X(\mathbf{r}, t; \omega)$, $\bar{x}(\mathbf{r}, t)$, and $\tilde{x}_i(\mathbf{r}, t)$, respectively. We first construct the realizations $\mathbf{x}^{(r)} \in \mathbb{R}^{N_{\mathbf{X}}}$ of the discretized initial tracer concentration $\mathbf{X}(0; \omega)$ by drawing independent samples $\mathbf{r}_c^{(r)}$, and transforming them in accordance with (7) and (6).

- The mean vector, $\bar{\mathbf{x}}(t)$, is set to the average of all the realizations $\mathbf{x}^{(r)}(0)$ [i.e., $\bar{\mathbf{x}}(0) = (1/N_r)\sum_{r=1}^{N_r}\mathbf{x}^{(r)}(0)$].
- For $i = 1\ 2, \ldots, s$, the orthonormal mode vectors $\tilde{\mathbf{x}}_i(0)$ are set to be the s leading eigenvectors of the sample covariance matrix $\mathbf{C}$, given by $\mathbf{C} = 1/(N_r - 1)\sum_{r=1}^{N_r}[\mathbf{x}^{(r)}(0) - \bar{\mathbf{x}}(0)][\mathbf{x}^{(r)}(0) - \bar{\mathbf{x}}(0)]^{\mathrm{T}}$.
- The ensemble members $\phi_i^{(r)}(t)$ of the stochastic coefficients $\Phi_i(t; \omega)$ are initialized as

$$\phi_i^{(r)}(0) = \tilde{\mathbf{x}}_i^{\mathrm{T}}(0) \times [\mathbf{x}^{(r)}(0) - \bar{\mathbf{x}}(0)],$$
$$i = 1, 2, \ldots, s, \quad r = 1, 2, \ldots, N_r. \tag{9}$$

The sample mean of the stochastic coefficients thus generated is zero. Typically, the number of realizations

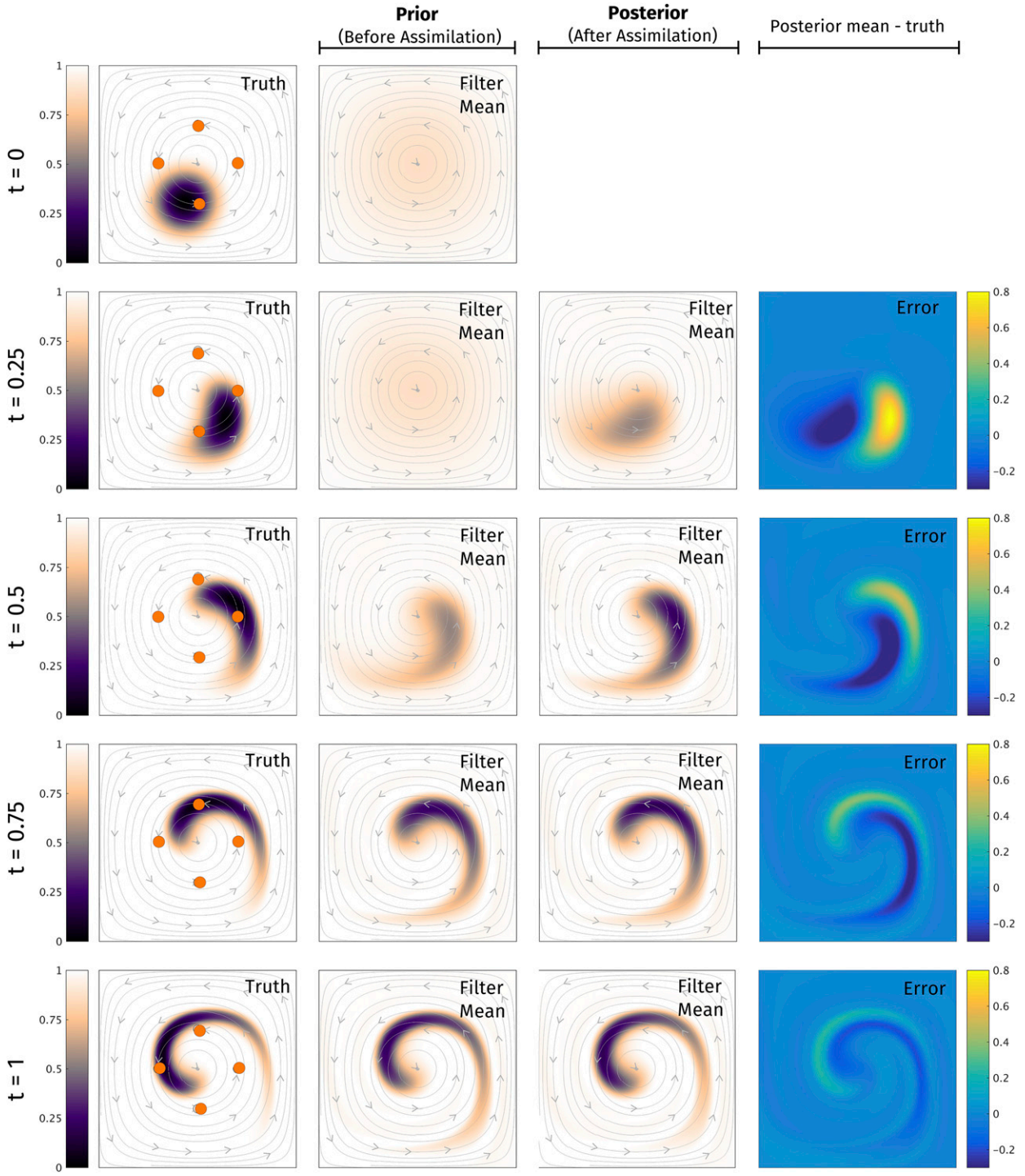

FIG. 5. Passive tracer advection in swirl flow. Time evolution of the mean tracer field $\bar{x}(\mathbf{r}, t)$ estimated by the GMM–DO filter, displayed alongside the true tracer field. The difference between the posterior mean field and the truth is also shown. The swirl flow is depicted by its streamlines, and the stationary sensors are indicated by white markers.

N_r, is much larger than s such that the dynamic structure of the pdf within the evolving subspace may be accurately represented. In all the results presented here, we use $s = 20$ modes and $N_r = 10^4$ ensemble members in the DO decomposition.

4) True solution and observation model

We initialize the true tracer concentration in an unbiased manner using a random sample $\mathbf{r}_c^{\text{true}}$ drawn from the initial distribution of $\mathbf{R}_c(\omega)$, given by (7). In this case, the sample equals $\mathbf{r}_c^{\text{true}} = (0.43, 0.31)$. Hence, the true concentration is initially unity everywhere except within a disk centered at $\mathbf{r}_c^{\text{true}}$. The significant initial error, shown in the first row in Fig. 5, is representative of data assimilation in common sea exercises, where the initial mean estimate can be quite far from the true solution (Lermusiaux et al. 2006; Haley et al. 2009; Ramp et al. 2009). The true field is evolved deterministically forward using (4).

We make four sets of observations ($K = 4$) of the tracer concentration, at times 0.25, 0.50, 0.75, and 1, through four stationary sensors, located at {(0.7, 0.05), (0.5, 0.7), (0.3, 0.5), (0.5, 0.3)}. The observation noises are set to be independent, both across sensors and through time, and of zero-mean Gaussian with variance $\sigma_0^2 = 0.08$. This variance is about twice as large as the variance in the tracer field expected at the measurement locations at time 0.25, the time of the first assimilation.

5) Numerical method

We solve the DO equations in (8) using a modular finite-volume framework. The geometry is discretized using a uniform, staggered C grid of 100 elements in each direction. The advection operator is discretized using a total variation diminishing (TVD) scheme with a monotonized central limiter (Van Leer 1977). A first-order forward Euler method is used to evolve the mean $\bar{x}$ and the modes $\tilde{x}_i$, and a four-stage Runge–Kutta scheme is employed for the stochastic coefficients (Φ_i). A time step of 5×10^{-4} is used, satisfying the Courant–Friedrichs–Lewy (CFL) condition.

b. *Validation of GMM–DO smoother: Computing the exact smoothed solution*

In this Bayesian setting, the skill of the GMM–DO smoother should be judged based on how close the pdf of the smoother estimate is to that of the exact smoothed pdf. We now obtain this exact smoothed pdf. First, recall from Lolla and Lermusiaux (2017) that the smoothed state-space distribution $p_{\mathbf{X}_k|\mathbf{Y}_{1:K}}$ is governed by the recursive smoothing equation:

$$p_{\mathbf{X}_k|\mathbf{Y}_{1:K}}(\mathbf{x}_k \,|\, \mathbf{y}_{1:K}) = \int p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:K}}(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}, \mathbf{y}_{1:K}) \times p_{\mathbf{X}_{k+1}|\mathbf{Y}_{1:K}}(\mathbf{x}_{k+1} \,|\, \mathbf{y}_{1:K})\, d\mathbf{x}_{k+1}. \tag{10}$$

A sample from this pdf may be drawn as follows:

$$\mathbf{x}_{k|1:K}^{(r)} \sim p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}(\cdot \,|\, \mathbf{x}_{k+1|1:K}^{(r)}, \mathbf{y}_{1:k}), \quad \text{where} \quad \mathbf{x}_{k+1|1:K}^{(r)} \sim p_{\mathbf{X}_{k+1}|\mathbf{Y}_{1:K}}(\cdot \,|\, \mathbf{y}_{1:K}). \tag{11}$$

Thus, given a set of samples of the smoothed state $\mathbf{X}_{k+1|1:K}$, one can use (11) to march backward in time and obtain the corresponding samples of $\mathbf{X}_{k|1:K}$. As discussed in Lolla and Lermusiaux (2017), the conditional distribution $p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}$ is not available in closed form for irreversible dynamical systems and nonlinear systems where the dynamics operator cannot be inverted analytically. However, $p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}$ can be evaluated here since the dynamics in (4) is reversible. To see this, let $\mathcal{L}_k^{k+1}$ denote the discrete operator that maps $\mathbf{X}_k$ to $\mathbf{X}_{k+1}$; that is, $\mathbf{X}_{k+1} = \mathcal{L}_k^{k+1}(\mathbf{X}_k)$. Although not known in closed form, $\mathcal{L}_k^{k+1}$ is implicitly given by the numerical solution of (4) from t_k to t_{k+1}. Furthermore, since the tracer advection is deterministic, the conditional pdf $p_{\mathbf{X}_{k+1}|\mathbf{X}_k}(\cdot \,|\, \mathbf{x}_k)$ is a Dirac delta distribution centered at $\mathcal{L}_k^{k+1}(\mathbf{x}_k)$; namely, $p_{\mathbf{X}_{k+1}|\mathbf{X}_k}(\mathbf{x}_{k+1} \,|\, \mathbf{x}_k) = \delta[\mathbf{x}_{k+1} - \mathcal{L}_k^{k+1}(\mathbf{x}_k)]$.

Since the dynamics in (4) is a passive advection, the operator $\mathcal{L}_k^{k+1}$ is a bijective map between $\mathbf{X}_k$ and $\mathbf{X}_{k+1}$. In other words, given a realization $\mathbf{x}_{k+1}$ of the discrete stochastic tracer field $\mathbf{X}_{k+1}$, there exists a unique vector $\mathbf{x}_k$ such that $\mathcal{L}_k^{k+1}(\mathbf{x}_k) = \mathbf{x}_{k+1}$. The position of each fluid parcel at time t_{k+1} may be traced back to its corresponding location at t_k by simply reversing the flow. As a consequence, the inverse of $\mathcal{L}_k^{k+1}$ (hereafter denoted by $\mathcal{L}_{k+1}^k$) exists, and is obtained by reversing the flow in (4), that is, replacing $\mathbf{v}$ with $-\mathbf{v}$. Mathematically, this is also equivalent to reversing time in (4). Moreover, since the time-reversed dynamics are also noise free, we obtain $p_{\mathbf{X}_k|\mathbf{X}_{k+1}}(\mathbf{x}_k \,|\, \mathbf{x}_{k+1}) = \delta[\mathbf{x}_k - \mathcal{L}_{k+1}^k(\mathbf{x}_{k+1})]$. Similar to $\mathcal{L}_k^{k+1}(\mathbf{x}_k)$, the inverse map $\mathcal{L}_{k+1}^k(\mathbf{x}_{k+1})$ is evaluated numerically [i.e., running (4) backward in time]. Now, using Bayes's theorem and the Markov property to rewrite $p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}(\cdot \,|\, \mathbf{x}_{k+1|1:K}^{(r)}, \mathbf{y}_{1:k})$, we obtain

$$\begin{aligned} &p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}(\mathbf{x}_k \,|\, \mathbf{x}_{k+1|1:K}^{(r)}, \mathbf{y}_{1:k}) \\ &\quad \propto p_{\mathbf{Y}_{1:k}|\mathbf{X}_k}(\mathbf{y}_{1:k} \,|\, \mathbf{x}_k) p_{\mathbf{X}_k|\mathbf{X}_{k+1}}(\mathbf{x}_k \,|\, \mathbf{x}_{k+1|1:K}^{(r)}) \\ &\quad = p_{\mathbf{Y}_{1:k}|\mathbf{X}_k}(\mathbf{y}_{1:k} \,|\, \mathbf{x}_k) \delta[\mathbf{x}_k - \mathcal{L}_{k+1}^k(\mathbf{x}_{k+1|1:K}^{(r)})]. \end{aligned}$$

The Dirac delta pdf $\delta[\mathbf{x}_k - \mathcal{L}^k_{k+1}(\mathbf{x}^{(r)}_{k+1|1:K})]$ vanishes for all $\mathbf{x}_k \neq \mathcal{L}^k_{k+1}[\mathbf{x}^{(r)}_{k+1|1:K}]$. Thus, as long as the likelihood $p_{\mathbf{Y}_{1:k}|\mathbf{X}_k}(\mathbf{y}_{1:k}\,|\,\mathbf{x}_k)$ is finite for such x_k, $p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}(\mathbf{x}_k\,|\,\mathbf{x}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k})$ must also vanish. Since $\mathcal{L}^k_{k+1}[\mathbf{x}^{(r)}_{k+1|1:K}]$ is uniquely defined, using the fact that $p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}}(\mathbf{x}_k\,|\,\mathbf{x}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k})$ must integrate to unity, we obtain

$$p_{\mathbf{X}_k|\mathbf{X}_{k+1},\mathbf{Y}_{1:k}} = \delta[\mathbf{x}_k - \mathcal{L}^k_{k+1}[\mathbf{x}^{(r)}_{k+1|1:K}]].$$

Therefore, from (11), $\mathbf{x}^{(r)}_{k|1:K} = \mathcal{L}^k_{k+1}[\mathbf{x}^{(r)}_{k+1|1:K}]$. In other words, samples of $\mathbf{X}_{k|1:K}$ may be obtained by evolving the samples of $\mathbf{X}_{k+1|1:K}$ backward in time in accordance with (4). Presently, since we use the DO methodology to predict pdfs, the exact smoothed pdf is obtained by solving the DO equations in (8) backward in time, starting from the filtered solution $\mathbf{X}_{K|1:K}$ at the final assimilation time $t_k = 1$. For a fair evaluation of the smoothing skill, we use the GMM–DO filter estimate of $\mathbf{X}_{K|1:K}$ to initialize the backward pass of the GMM–DO smoother and of the GMM–DO ESSE smoother, as well as the computation of the exact smoothed pdf.

It is important to note that this approach of reversing time is only applicable to dynamical systems that are reversible, and numerically stable upon reversing time. In the present case, the advection operator remains numerically stable upon reversing $\mathbf{v}$, since the time-reversed dynamics are well posed. However, this property does not hold for general nonlinear dynamical systems. For example, a fluid flow governed by the viscous Navier–Stokes equations is not reversible. In fact, the time-reversed dynamics are ill-posed and, as a consequence, its numerical solution becomes unstable. The present example was specifically chosen so as to highlight this time-reversal property and we use it to determine the exact smoothed solution.

c. Results and discussion

Figure 5 depicts the evolution of the true tracer concentration field, the mean field of the GMM–DO filter estimate (both prior and posterior), and the difference between the posterior mean and the true field. At $t = 0$, the true solution is uniform everywhere except in the interior of a disk of radius 0.25 centered at $\mathbf{r}^{\text{true}}_c = (0.43, 0.31)$. The initial mean field of the GMM–DO filter is largely uniform as a result of the distribution of $\mathbf{R}_c$. Thus, the largest initial difference is observed within the interior of this disk. Because of the shear flow, the true tracer distribution becomes increasingly deformed, appearing as a long thin arc at $t = 1$, the end of the simulation. As observations are assimilated over time, the GMM–DO filter's mean estimate improves, as indicated by the magnitudes in the difference plots. The largest mean innovations (posterior mean minus prior mean) and error reductions occur during the second assimilation $t = 0.5$) because the sensor location then samples the patch of low values (note that since only advection is present, the sign and magnitude of the tracer is arbitrary). Subsequently, the fourth and third assimilations are still informative because two sensor locations surround the now-elongated patch of low values. By the end of the run ($t = 1$), the filter mean bears a strong visual resemblance to the truth.

The normalized RMS difference between the filter mean and true tracer field is plotted and is shown later (Fig. 9a). It quantifies the error reduction in the means shown in Fig. 5. We note that the RMS difference remains constant between successive assimilation times. This happens because the flow $\mathbf{v}$ is nondivergent and vanishes at the boundaries. Additionally, the tracer dynamics are purely advection based and does not contain any source of randomness. As a consequence, the uncertainty does not grow or decay with time. Over the course of the entire simulation, the RMS error of the GMM–DO filter mean estimate reduces from 18% at $t = 0$ to 6% at $t = 1$. Note that here these errors contain a component due to the filter approximation but also a large component due to the initial mean error and the limited observations (see Fig. 5, first row): the exact Bayesian filter solution does not have a mean equal to the truth right away either.

In Fig. 6, we depict the evolution of the first three modes ($\tilde{x}_i$, for $i = 1, 2, 3$) and the marginal pdfs of their respective stochastic coefficients (Φ_i). The prior pdfs are shown by dotted lines whereas the posterior pdfs are depicted by solid lines. We observe that the stochastic coefficients are composed of several far-from-Gaussian features. The updates of the GMM–DO filter capture these features by optimally adjusting the mixture complexity (M). Furthermore, the GMM fitting and analysis is performed in the low-dimensional subspace spanned by the modes $\tilde{x}_i$. We can also observe that upon assimilation, the spread of the stochastic coefficients is reduced, as their pdfs become taller and thinner. This indicates that the GMM–DO filter becomes increasingly confident in its estimate of the true tracer concentration over time. As expected, the pattern of time evolution of the modes is synchronous to the flow magnitude and direction. The flow field in (5) vanishes at the domain center (0.5, 0.5) and at the edges. Since the dynamics in (4) are advection driven, the modes at these locations do not evolve with time. At other locations, the modes evolve in an axisymmetric fashion, with a speed depending on the magnitude of the swirl flow at that

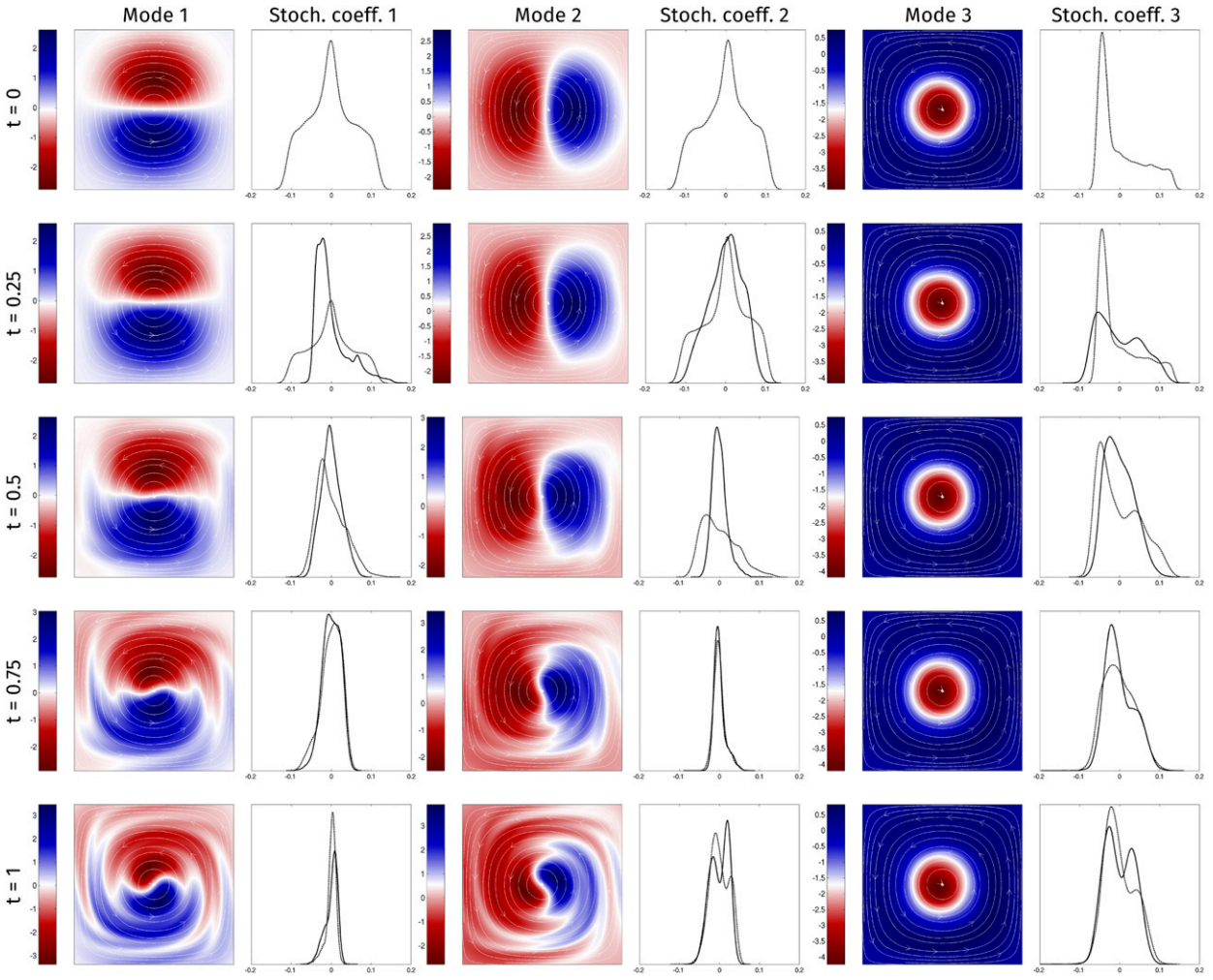

FIG. 6. Passive tracer advection in swirl flow. Time evolution of the first three DO modes $\tilde{x}_i(\mathbf{r}, t)$ of the GMM–DO filter, displayed alongside the kernel density estimates of the corresponding marginal pdfs of $\Phi_i(t; \omega)$. The prior pdfs are shown as dotted lines, while posterior pdfs are depicted as solid lines.

location. In particular, the third mode $\tilde{x}_3(\mathbf{r}, t)$ is initially axisymmetric (up to roundoff), and hence the term $\mathbf{v} \cdot \nabla_{\mathbf{r}} \tilde{x}_3$ is close to zero. Thus, (8b) implies that the rate of change of $\tilde{x}_3$ is also small. Hence, $\tilde{x}_3$ remains axisymmetric with time.

Next, the joint GMM-fitting pass of filtered states and the backward-smoothing pass are completed, starting from the filtered solution at $t = 1$. The exact smoothed pdf is obtained by solving the DO equations in (8) backward in time, from $t = 1$, until $t = 0$. Figure 7 compares the mean tracer field of the GMM–DO smoother with the exact smoothed mean at the four assimilation times. The difference between the GMM–DO smoother mean and the true smoothed mean are also shown. We observe that the mean field of the GMM–DO smoother closely matches that of the exact smoothed pdf at all times. The largest errors again occur close to the deformed disk. Ignoring numerical errors in the calculation of the true smoothed mean (e.g., due to numerical diffusion), the differences in the fields arise only because of the joint-GMM fitting and the finite number of samples used to represent the smoothed stochastic coefficients. Later, we show the normalized RMS error between these mean field quantities (Fig. 9b). At the final time, $t = 1$, the normalized RMS error is zero because the GMM–DO backward-smoothing pass and exact smoothed solution are each initialized with the GMM–DO filter solution. As we approach $t = 0$, the GMM–DO smoother mean begins to slightly depart from the exact smoothed mean, owing to the imperfect joint subspace GMM fits and also to the imperfect DO filter

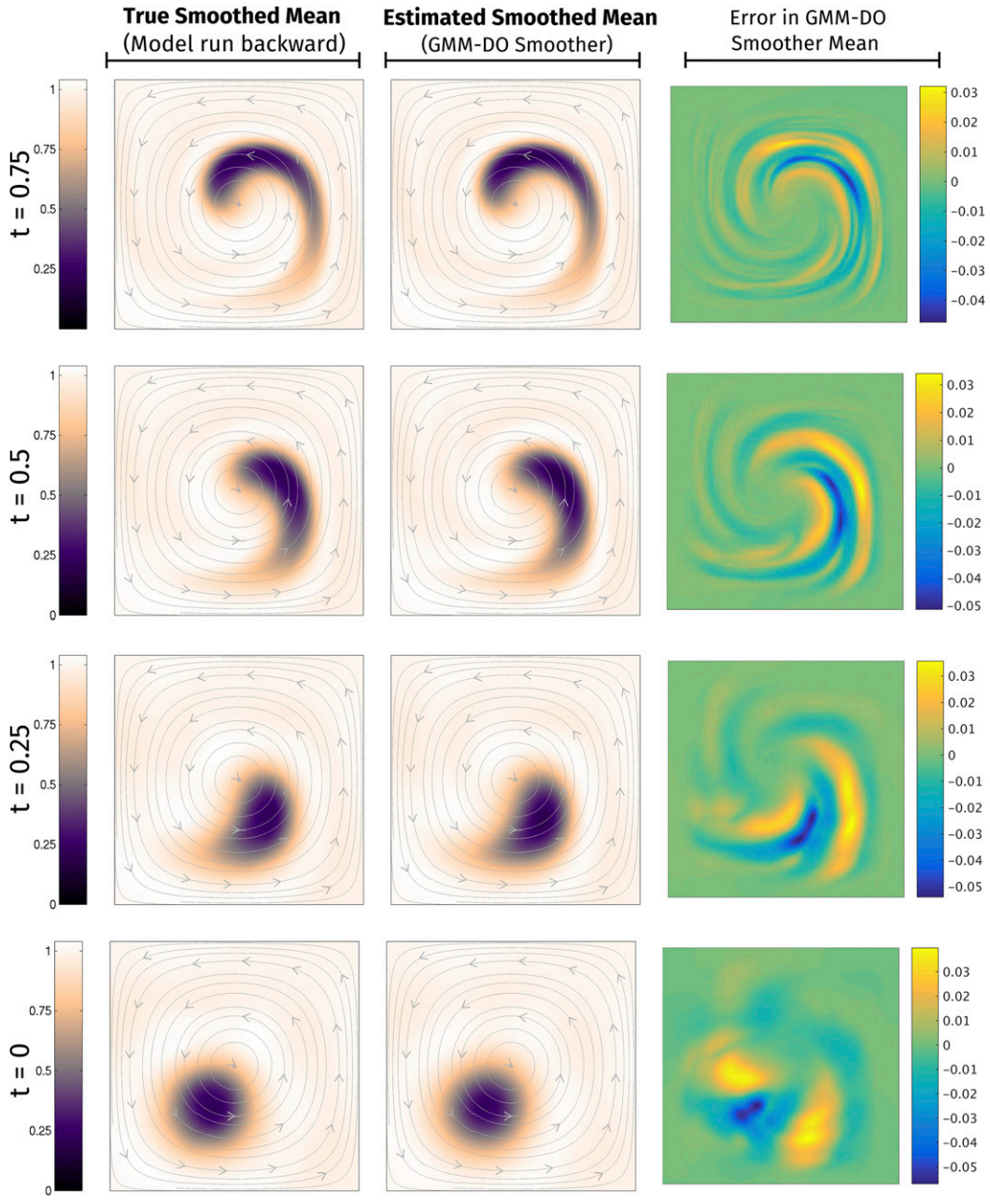

FIG. 7. Passive tracer advection in swirl flow. Time evolution of the mean tracer field estimated by the GMM–DO smoother plotted alongside the corresponding exact smoothed mean field. The exact smoothed mean is computed by reversing the flow from the final filtered solution. The error in the smoother mean is also depicted.

modes and coefficients at final time $t = 1$ (the initial conditions of the GMM–DO smoothing). Nevertheless, the GMM–DO smoother mean remains within 1% of the exact smoothed mean throughout the simulation time window.

Figure 8 compares the first three modes and marginal pdfs of the stochastic coefficients of the GMM–DO smoother (shown in blue) to those of the exact smoothed solution (shown in black), at $t = 0.75$, 0.5, 0.25, and 0. Also shown (in gray) are the marginal pdfs of the coefficients when the Gaussian ESSE smoother updates are used in the backward pass. It is clear that both sets of modes are in very good agreement with each other at the times shown. We observe that at $t = 0.75$, the GMM–DO

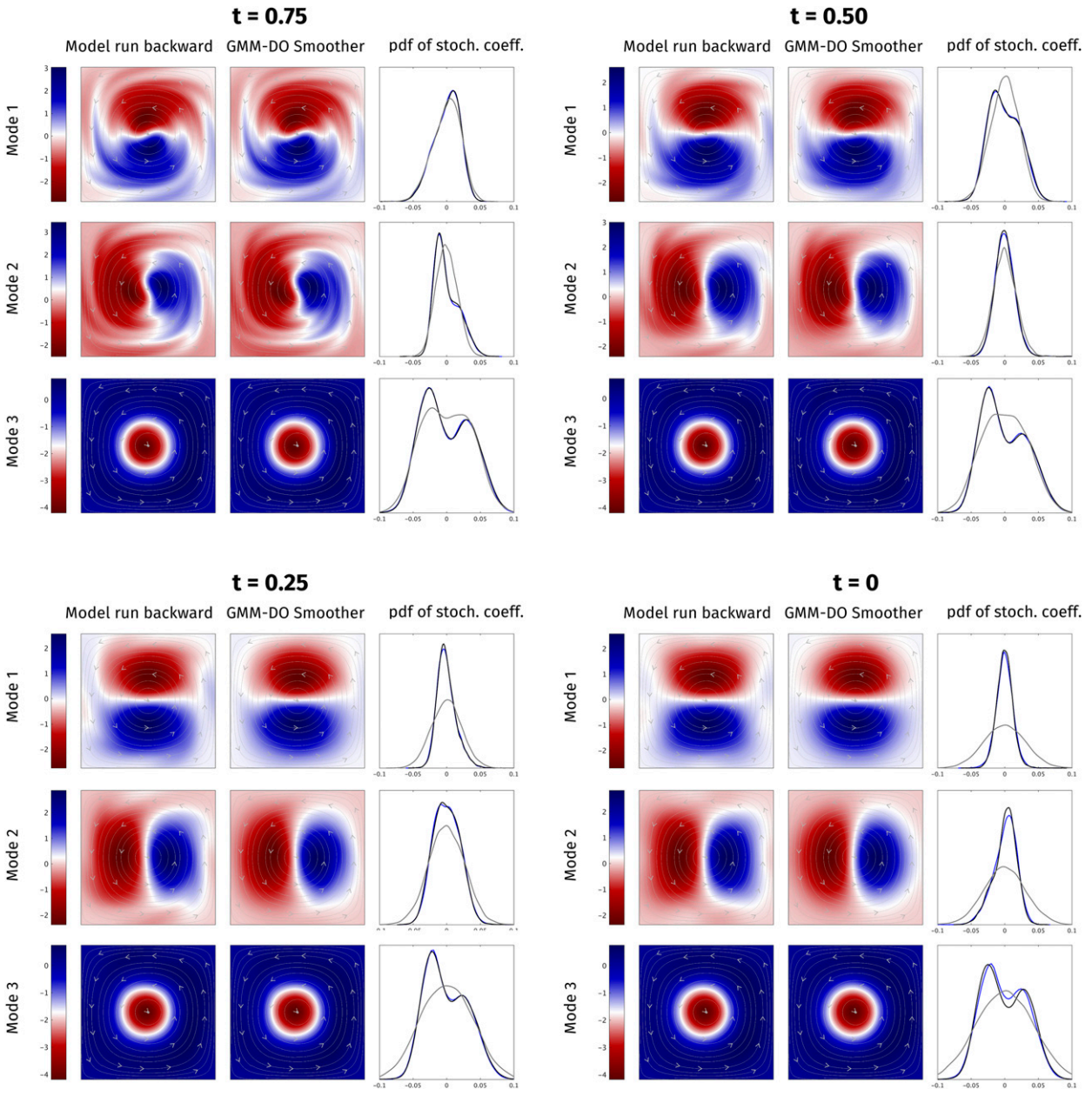

FIG. 8. Passive tracer advection in swirl flow. Time evolution of the first three modes and the marginal pdfs of their respective stochastic coefficients (blue) during the backward-smoothing pass of the GMM–DO smoother. The marginal pdfs shown in black correspond to the stochastic coefficients of the exact smoothed tracer field (they almost overlap with these of the GMM–DO smoother). The marginal pdf estimates from the GMM–DO ESSE smoother are depicted in gray (they become more and more Gaussian as time decays).

smoother pdfs match those of the exact solution. The pdf of the first coefficient is well captured by GMM–DO ESSE, but differences are noticeable in the second and third coefficients, where the non-Gaussian features are apparent. As the schemes march farther back in time, the discrepancies between the GMM–DO ESSE and GMM–DO smoother pdfs grow larger. The Gaussian smoother diffuses the pdfs, thereby losing its non-Gaussian features. In contrast, the pdfs of the GMM–DO smoother coefficients and the exact solution match closely at all times. Finally, although we have only shown the first three modes here, the marginal pdfs of the remaining modes are also very similar. This implies that the GMM–DO smoother accurately captures not

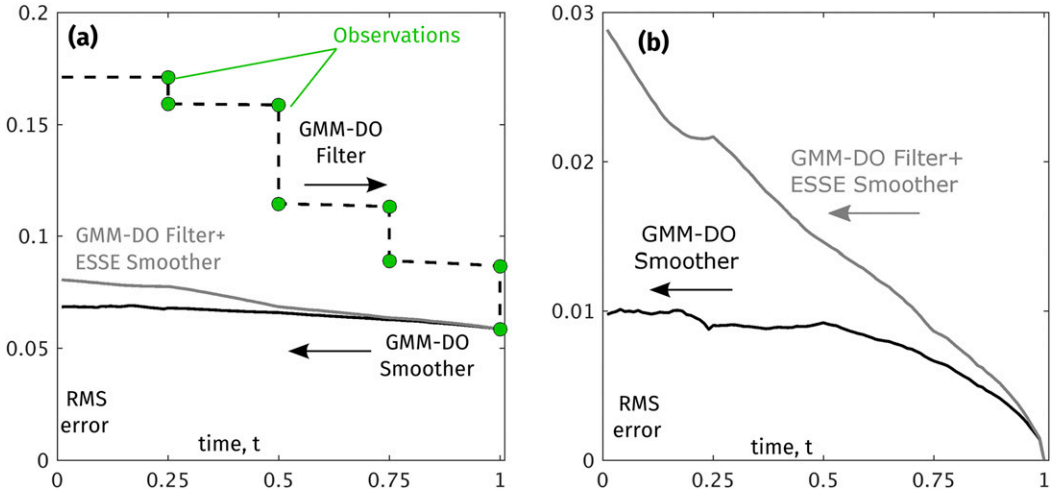

FIG. 9. Passive tracer advection in swirl flow. (a) Normalized RMS error between the mean estimates of the GMM–DO filter (dashed black), the GMM–DO smoother (solid black), the GMM–DO ESSE smoother (solid gray), and the true tracer field. We see that the smoother means are much closer to the truth than the filter mean. (b) Normalized RMS error between the smoother means (GMM–DO smoother, solid black; GMM–DO ESSE smoother, solid gray) and the exact smoothed mean, computed by reversing the flow from the final-time filter solution (this is not the true tracer field). The GMM–DO smoother error remains less than 1% during the entire simulation window.

just the mean field (Fig. 9), but also the exact smoothed pdf of the state vector. This completes its stochastic validation.

4. Sudden expansion flows

We now examine the performance of the GMM–DO smoother in more complex, high-dimensional dynamics, consisting of variable jets and eddies. Specifically, we consider two-dimensional *sudden expansion* flows (Fearn et al. 1990). Such flows are quite common in the coastal ocean. They occur when a uniform barotropic jet exits a narrow strait or estuary, leading to meander and vortex generation as the jet exits the constriction. These circulation features are highly variable: the number of vortices, their locations, the strength of the jet, and its meanders are all time-varying quantities, leading to nonstationary distributions.

This dynamical system was utilized in Sondergaard and Lermusiaux (2013b) to test and evaluate the GMM–DO filter. Here, we contrast the GMM–DO smoother against the GMM–DO ESSE scheme, as in the previous passive tracer advection in section 3. As in the previous two examples, the true solution is generated by drawing a random sample from the stochastic initial conditions and evolving the sample deterministically. Interestingly, this third example develops dynamic multimodal statistics associated with the breaking of flow symmetries at moderate to large Reynolds numbers. Hence, it is an ideal irreversible test to evaluate the GMM–DO smoother, which aims to capture and retain such non-Gaussian features backward in time. Overall, this example demonstrates the applicability of the GMM–DO smoother in practical high-dimensional dynamical systems. Next, we describe the experimental setup and test procedure, and then we discuss the results.

a. *Experimental setup and test procedure*

Figure 10 shows the schematic setup of the 2D sudden expansion test case. A uniform, horizontal flow of nondimensional speed $U = 1$ enters a narrow conduit of width $h = 1/3$ at the far-left end of the domain. It develops into a steady, parabolic flow, attaining a maximum speed $U_{\max} = 1.5$ at the center line. At $x = 3$, the flow reaches the abrupt expansion into a channel of larger width. The flow exits the channel at the right end of the domain, $L = 20$ units downstream of the expansion.

The flow behavior closely depends on the Reynolds number (Re). For low Re, the flow remains steady and symmetric about the centerline, with recirculation zones appearing at the corners of the expansion (Durst et al. 1974). At larger Re, the flow downstream of the expansion develops asymmetry and exhibits vortex shedding, which can either be steady or unsteady. Here, we consider an intermediate value of Re = 250, at which we expect the symmetric inlet flow to break toward one side of the centerline, depending on the initial perturbations (see Fig. 11a for an illustration of the flow field). Farther downstream, a second region of circulation forces the

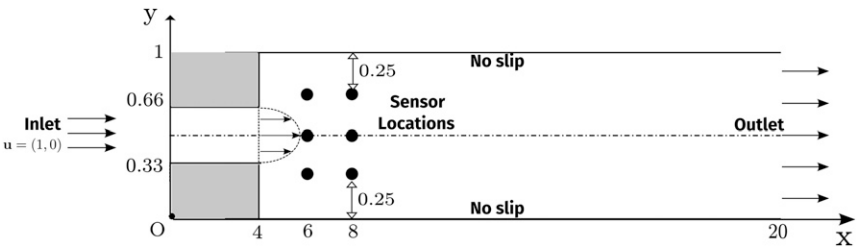

FIG. 10. Sudden expansion flow setup. Six sensors located at $(x_{\text{obs}}, y_{\text{obs}}) = \{(6, 0.25), (6, 0.5), (6, 0.75), (8, 0.25). (8, 0.5), (8, 0.75)\}$.

flow to the opposite side before symmetry is eventually restored (Sondergaard and Lermusiaux 2013b).

The flow is governed by the incompressible Navier–Stokes equations without source terms. The corresponding DO equations in (16) are solved to forecast the uncertainties, with the specifics as described in Ueckermann et al. (2013). The state vector $\mathbf{X}(t; \omega)$ consists of the spatially discretized x and y components of the flow field. In this example, uncertainty arises only through the stochastic initial conditions.

1) INITIALIZATION OF DO DECOMPOSITION

The DO initialization employed here is simple and inspired by the multiscale multivariate initialization of ESSE simulations used for real-time regional ocean probabilistic predictions (Lermusiaux et al. 2000; Lermusiaux 2002).

- Mean vector $\bar{\mathbf{x}}(0)$: the x component of the mean field velocity is initialized to 1 everywhere in the inlet and to ⅓ in the channel so as to satisfy continuity. The y component of the mean flow is initialized to zero everywhere.
- Modes $\tilde{\mathbf{x}}_i(0)$: following (Sapsis and Lermusiaux 2009; Sondergaard and Lermusiaux 2013b), the orthonormal modes of the streamfunction are generated by retaining the dominant singular vectors of the correlation operator $C(\cdot, \cdot)$, defined by

$$C(\mathbf{r}_1, \mathbf{r}_2) = \mathcal{M}(\mathbf{r}_1, \mathbf{r}_2)\tilde{C}(\|\mathbf{r}_1 - \mathbf{r}_2\|_2), \tag{12}$$

  where $||\cdot||_2$ denotes the L2 norm, $\mathcal{M}$ is a mollifier function that is unity everywhere except near solid boundaries, at which it vanishes smoothly. In addition, $\tilde{C}(r)$ takes the form

$$\tilde{C}(r) = \left(1 + r + \frac{r^2}{3}\right) \exp(-r). \tag{13}$$

  We initialize the stochastic subspace by retaining the 20 most dominant eigenvectors of C; that is, we use $s = 20$.
- Stochastic coefficients $\Phi_i(0; \omega)$: for $1 \le i \le 20$, the coefficients are initialized as uncorrelated, zero-mean normal random variables, with variance proportional to the corresponding eigenvalue of the correlation operator defined in (12). We generate $N_r = 10{,}000$ subspace realizations $\phi_i^{(r)}$ of Φ_i and solve (A3c) in a particle-wise manner. We use $N_r \gg s$ so as to accurately capture the dynamic structure of the pdf in the evolving subspace.

2) TRUE SOLUTION AND OBSERVATIONAL MODEL

We initialize the true solution by drawing an arbitrary field according to the aforementioned initial pdf. We then propagate this sample forward in time using the deterministic Navier–Stokes equations for a total time of $T = 70$.

We make a total of four sets of noisy measurements of the x and y components of the true flow field at $t = 40$, 50, 60, and 70. The measurements are provided by six stationary sensors, placed at locations indicated in Fig. 10. Their measurement noises are independent and of zero-mean Gaussian distribution with variance of $\sigma_0^2 = 0.1$. As with the tracer advection example, this variance is about as large as the predicted variance in the flow field at the first assimilation time. Conditioned on these noisy data, we wish to compute the smoothed pdf of the flow field.

3) NUMERICAL METHOD

We numerically solve the Navier–Stokes DO equations using a finite-volume method (Ueckermann et al. 2013). The sudden expansion geometry is discretized onto a uniform structured grid of 50 × 30 nodes in the x and y directions, respectively. Thus, $N_{\mathbf{x}} = 3000$, since we solve for both the x and y velocity components (numerically, note we also solve for pressure). The diffusion operator is discretized a using second-order central differencing scheme. The advection operator makes use of a TVD scheme with a monotonized central limiter. The modes are evolved using a first-order semi-implicit projection method,

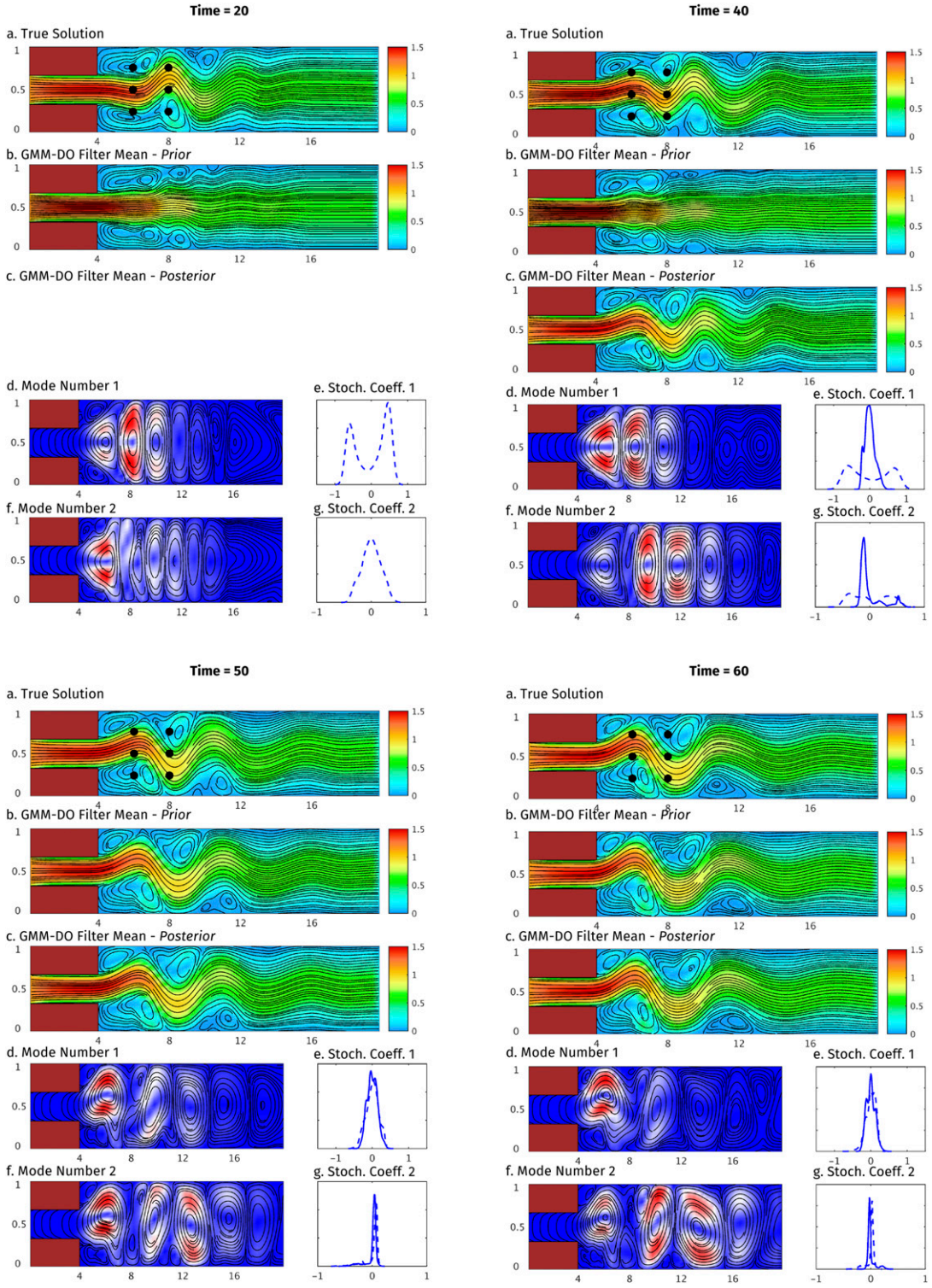

FIG. 11. Sudden expansion flow. The time evolution of (a) true flow field; (b) the prior GMM–DO filter mean field; (c) the posterior GMM–DO filter mean field (when available); (d),(f) the first two modes of the GMM–DO filter; and (e),(g) the marginal distributions of the corresponding stochastic coefficients (prior, dotted; posterior, solid).

where the diffusion and pressure terms are treated implicitly, and advection is treated explicitly. A time step of 0.025 is used, satisfying the CFL condition. The stochastic coefficients are integrated in time using a four-stage Runge–Kutta scheme. As shown in Fig. 10, no-slip boundary conditions are used at all the solid boundaries. At the open outlet, we impose a zero Neumann boundary condition for both components of the velocity. Across the inlet, we impose a uniform horizontal velocity of 1.

b. *Results and discussion*

Though the circulation features observed in this example are highly dynamic, we mainly focus on the results at $t = 20$, 40, 50, 60, and 70, which show snapshots of the DO evolution at three intermediate assimilation times, and the final time. At each of these times, Figs. 11 and 12 depict the true solution in panel a; the prior mean field in panel b; the posterior mean field (only shown at assimilation times) in panel c; the two most dominant modes, $\tilde{x}_1$, $\tilde{x}_2$, in panels (d) and (f), respectively; and the marginal filtered pdfs of their stochastic coefficients Φ_1, Φ_2, in panels (e) and (g), respectively. The dotted and solid curves represent the prior and posterior marginal pdfs, respectively. All fluid flows are shown by streamlines overlaid onto a color map of velocity magnitude.

Figure 11 illustrates the GMM–DO filter pass, showing how the flow and uncertainties evolve and ultimately settle. At $t = 20$, the initial perturbations in the true solution have already broken the symmetry of the flow, whereas the DO mean field forecast remains symmetric as no data have been assimilated. For Φ_1, we attain a bimodal distribution, where each mode corresponds to a direction in which the sudden expansion flow may break. However, Φ_2 remains unimodal at this time, not having evolved much from its Gaussian initialization. Nonetheless, the variances of both Φ_1 and Φ_2 grow from their initial values, reflecting the increase in uncertainty over time. The modes $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_2$ also remain symmetric since no observations have been assimilated so far. Before the first assimilation at $t = 40$, the DO mean field forecast is still symmetric and therefore substantially different from the truth. Similarly, the modes are also symmetric at this time. However, Φ_2 now starts to develop bimodal structures, indicating the possible secondary bifurcations of weaker downstream eddies at $x = 9$. After the assimilation at $t = 40$, the mean field shows a marked improvement, correctly identifying the direction of the flow breakage. This is confirmed by the normalized RMS difference between the GMM–DO filter mean and the true solution shown in Fig. 13a. The posterior coefficients remain far from Gaussian, but have a lower variance than their corresponding priors, suggesting a greater confidence in the estimate. Evolved forward in time, the posterior mean field breaks the symmetry of the modes at $t = 50$, and the uncertainty is largely at the edges of the main jet. Additionally, the eddies on either side of the main jet grow between $t = 40$ and $t = 50$. During $t = 50$–60, the meander around $x = 12$ evolves farther downstream. However, this evolution is also uncertain, as indicated by the widening shapes in $\tilde{\mathbf{x}}_2$ during this time interval. Additionally, the prior filter mean closely resembles the corresponding true field. Subsequent assimilations continue to further reduce the RMS difference (see Fig. 13a). The tall, thin marginal pdfs of Φ_2 indicate that the stochastic energy of the second mode has almost completely evanesced, leaving the first mode to be the main contributor of uncertainty. Moreover, at $t = 60$, the posterior pdf of Φ_1 exhibits similar levels of variance as the prior, indicating that the variance of the prior is now much smaller than that of the observation noise projected onto the first mode. This is further evidenced by the modest improvement in the RMS difference achieved by the assimilation at $t = 60$ (see Fig. 13a). At the end of the forward-filter pass ($t = 70$), the RMS difference between the mean and the truth is down to 6%, from a value of 25% prior to the first assimilation. The GMM–DO filter correctly forecasts that its final $t = 70$ estimate (Fig. 12) has accurately captured the true solution: the prior variances of Φ_1 and Φ_2 are indeed much reduced.

Figure 14 depicts the results of the joint subspace GMM-fitting pass, at $t = 39$ and $t = 40$ (i.e., immediately prior to the first assimilation) for the two most dominant modes. The individual realizations $\phi_1^{(r)}$ and $\phi_2^{(r)}$, $r = 1, 2, \ldots, N_r$, are shown in orange, in the form of a scatterplot. Based on the BIC criterion, we determine the optimal mixture complexity to be $M = 16$. We display the one standard deviation contours of the mixture components (shown in black) marginalized across pairs of coefficients (2D joint pdfs). We further project the optimal GMM onto the domain of each coefficient, thus obtaining their respective 1D marginal pdfs (also shown in black). The kernel density estimate of the marginal pdfs (shown in blue) are also plotted for comparison. Finally, we also show, in gray, the pairwise-joint and marginal pdfs obtained by a multivariate Gaussian fit of the joint subspace coefficients (i.e., the GMM with a forced mixture complexity of $M = 1$). It is clear that the joint-GMM fit successfully captures the complex, multimodal characteristics of the stochastic coefficients. In particular, the 2D projections

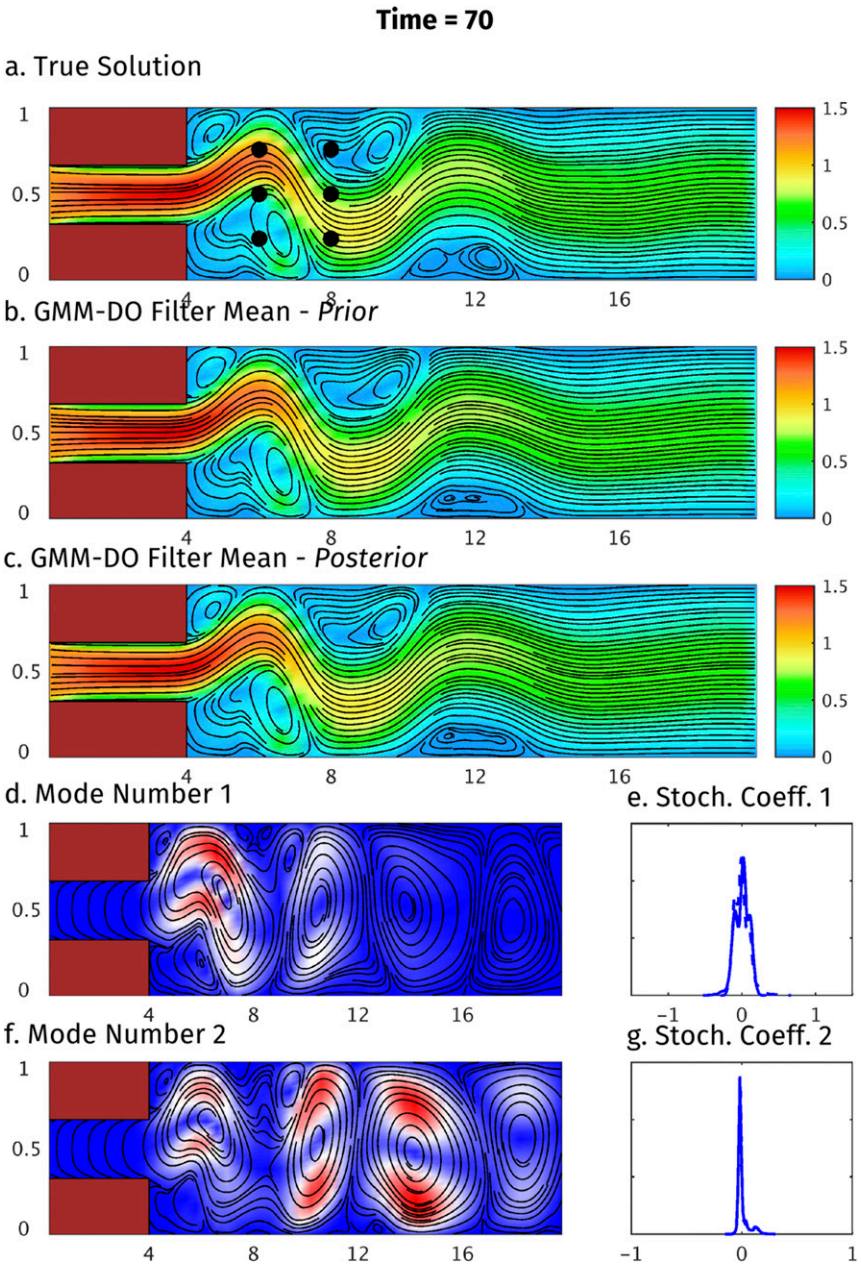

FIG. 12. Sudden expansion flow. The GMM–DO filter estimates at the final time, $t = 70$. The description of individual panels is the same as in Fig. 11. This final filter solution is used to initialize the backward-smoothing pass of the GMM–DO smoother.

of the GMM identify the localized regions in the subspace that dynamical realizations visit most, and the 1D GMM marginals differ little from their respective kernel density estimates. On the other hand, the Gaussian fit results in a severe loss of dynamical information as it cannot capture the multimodal structure of the pdfs. Of further interest is the shape of the joint pdf of the stochastic coefficients across time. We observe that the joint pdf of $\Phi_1(39)$ and $\Phi_1(40)$ takes on an elongated shape, with a majority of realizations concentrated at the two extremes. This shape can be primarily attributed to the noise-free dynamics of the fluid flow, coupled with the fact that the two times are not far enough apart for the dynamics to spread the samples farther out. We remark that it is this joint pdf of coefficients across time that enables smoothing.

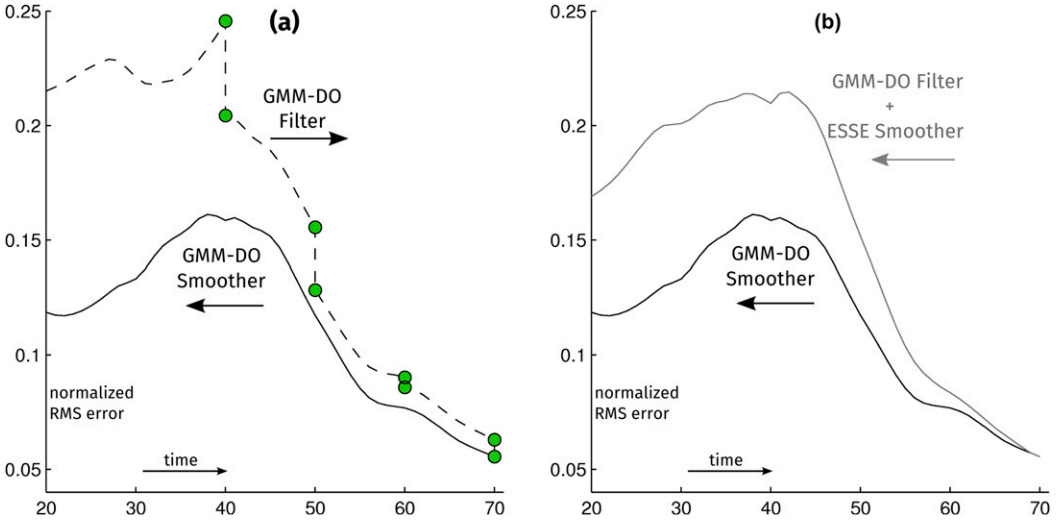

FIG. 13. Sudden expansion flow. (a) The normalized RMS error between the mean estimates of the GMM–DO filter (dashed black), GMM–DO smoother (solid black), and the true flow field. (b) The normalized RMS error between the mean estimates of the GMM–DO ESSE scheme (solid gray) and the true field. The plot of the GMM–DO smoother error (solid black) is the same as in (a). We note that these RMS errors should not be zero theoretically; they should be equal to the RMS errors of the true Bayesian mean estimate, which depends on the initial error conditions.

Consequently, it is crucial to accurately represent this pdf. We see that the GMM captures it, whereas the Gaussian fit puts too much weight toward the center and is, consequently, detrimental to the smoother's performance.

We illustrate the results of the backward-smoothing pass, at $t = 60$, 50, 40, and 20 (as in Fig. 11), showing the true field in Fig. 15a, the mean field of the GMM–DO smoother in Fig. 15b, the two most dominant modes in Fig. 15c, and the corresponding marginal pdfs of the smoothed stochastic coefficients in Figs. 15e and 15f. We clearly find that the GMM–DO smoother mean is much closer to the truth than the filter mean: at all times, the smoother accurately captures the direction in which the main meander breaks, as well as the location of the secondary eddies farther downstream. The smoother is logically most superior over the filter at past times of $t = 40$ and $t = 20$, when the filter mean is still significantly different from the truth. The statistical coupling across time due to the joint-subspace GMM fits allows the smoother to detect the break sooner than the filter. At $t = 50$ and 60, the improvement due to the smoother is smaller, since the filter mean was already close to the truth.

We also note that the smoother estimate remains precise over time, as indicated by the low variance in the two most energetic modes. Figure 16 shows, in black, the realizations of the joint smoothed coefficients Φ_1, Φ_2 at $t = 40$ and $t = 39$. As in Fig. 14, the orange-colored scatter points correspond to realizations of the pairwise filtered coefficients. Also shown in Fig. 16 are the 1D marginal pdfs of the smoothed coefficients (solid line). From these plots, we see that the smoothed coefficients have a much lower variance, but remain far from Gaussian. Therefore, we can expect the GMM–DO smoother to achieve a superior level of performance over Gaussian-based smoothers.

In Fig. 13, we examine the performance of the GMM–DO smoother against the GMM–DO ESSE smoother (again, if the Gaussian EnKS is used, results are similar, but the EnKS is much more expensive because of its backward update for each of the $N_r = 10\,000$ realizations). Figure 13 depicts the RMS differences between each of the two smoothed mean fields and the true solution. At $t = 70$, the GMM–DO ESSE estimate is similar to the GMM–DO smoother, since both are initialized with the GMM–DO filter pdf. However, as we march backward in time, the performance of the GMM–DO ESSE scheme deteriorates when compared with the GMM–DO smoother, with the relative discrepancy growing to be about 40% at $t = 20$. In fact, between times $t = 40$ and $t = 56$, the mean of the GMM–DO ESSE smoother is briefly worse than the GMM–DO filter; between $t = 40$ and $t = 20$, it recovers. As hinted earlier, the origin of the inferior performance of the GMM–DO ESSE is the loss of dynamical information

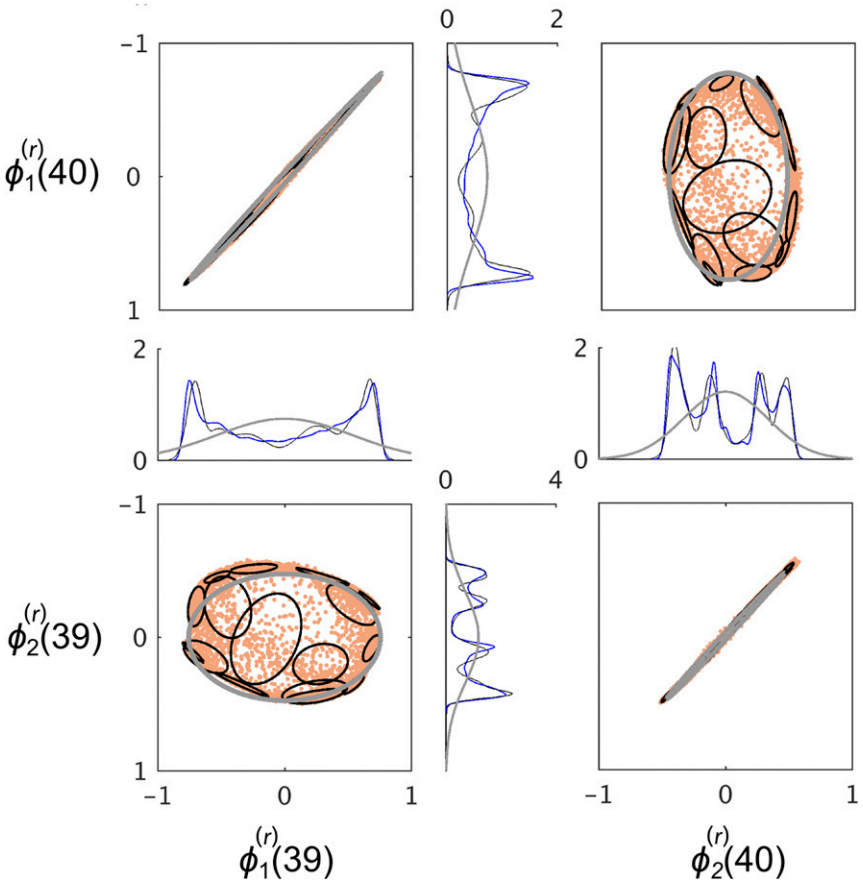

FIG. 14. Joint GMM fits in the sudden expansion flow. Pairwise joint distributions of the first two stochastic coefficients Φ_1 and Φ_2 at times $t = 39$ and $t = 40$ (i.e., immediately before the first assimilation). Ensemble members are shown in orange, and their marginalized joint GMM fits in both 1D and 2D are depicted in black. Similarly, the joint Gaussian fit (GMM fit with a forced mixture complexity of $M = 1$) is depicted in gray. The 1D marginal densities, computed by a kernel dressing scheme, are depicted in blue. Note that all the above marginal distributions are projections of joint-GMM fits performed in a $2s$-dimensional joint subspace.

due to the Gaussian fitting between the joint subspace variables. In comparison, the adaptive GMM fit performed by the GMM–DO smoother accounts for the non-Gaussian structures of the joint pdfs, thereby allowing an accurate inference. We remind the reader that all RMS errors shown in Fig. 13 contain a component due to the filter–smoother approximations but also due to the initial mean conditions and the limited observations: the exact Bayesian filter–smoother does not have a mean equal to the truth right away either.

5. Conclusions

We exemplified the properties and capabilities of the GMM–DO smoother and contrasted its performance with respect to that of other smoothers by applications to three dynamical systems, all of which admit far-from-Gaussian statistics. First, a double-well stochastic diffusion experiment was used to reconstruct the history of a particle forced under pseudo-gravity and external noise. We showed that the GMM–DO smoother accurately utilizes the multimodal structures of the state variable, and achieves a level of performance superior to that of the EnKS in detecting the transition of the particle across wells. The second example was that of an uncertain passive tracer advected in a reversible swirl shear flow, an idealization of transport phenomena occurring in geophysical fluids. This innovative smoothing test case was developed because we could derive the analytical smoothed distribution of the tracer concentration. That distribution was computed and

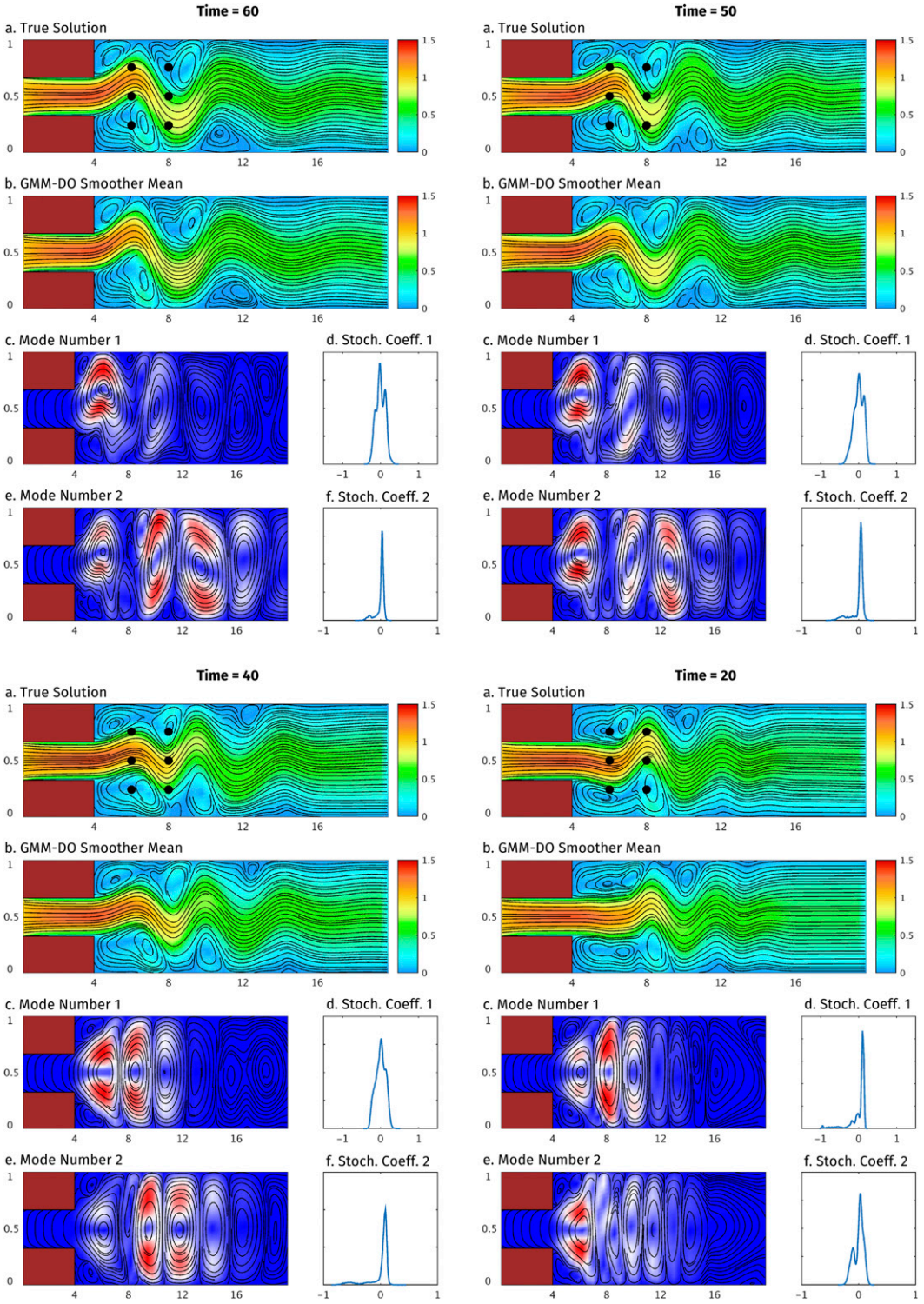

FIG. 15. Sudden expansion flow. Time-evolving estimates of the backward-smoother pass of the GMM–DO smoother as in Fig. 11, but panels are backward in time and without showing the prior (which is the GMM–DO filter estimate).

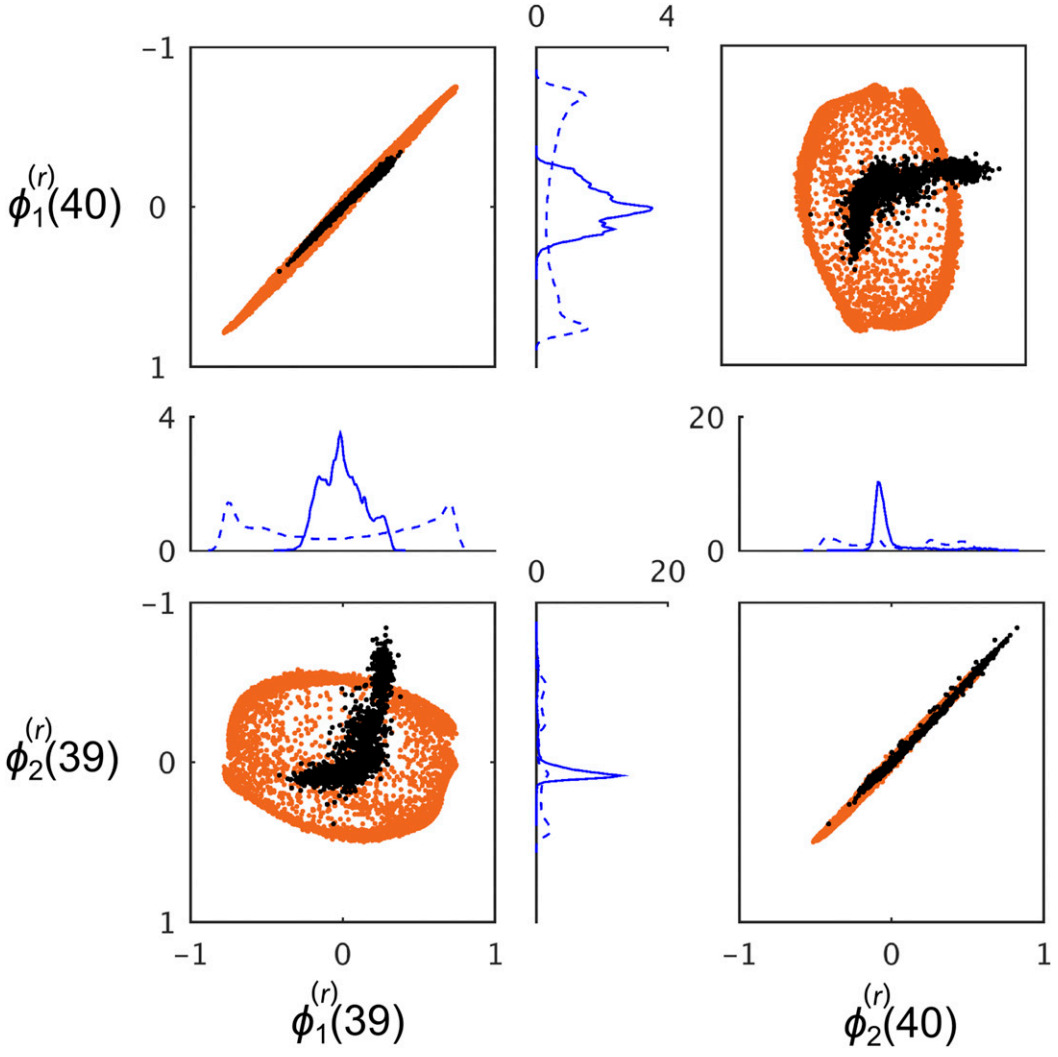

FIG. 16. A comparison of the first two stochastic coefficients at times $t = 39$ and $t = 40$ (filtered, orange; smoothed, black). The 1D marginals computed using the joint GMM-fitting pass are depicted as dashed lines, and the smoothed 1D marginal densities computed using a kernel dressing scheme are shown as solid lines.

utilized to successfully validate the GMM–DO smoother. Results showed that the GMM–DO smoother very accurately captured the full smoothed distributions, not just the mean states. Finally, we examined the performance of the GMM–DO smoother in more complex simulated ocean flows exiting a strait or estuary with jets and eddies. We found that the backward inference of the GMM–DO smoother was very efficient and accurate. To showcase the importance of exploiting the non-Gaussian properties of the conditional probabilities, comparisons were made with the Gaussian update of the ESSE smoother. For instance, in the third example, the final mean velocity fields of the GMM–DO smoother were found to have an RMS error 40% smaller than those of the GMM–DO ESSE smoother. In general, it is shown that even when the dynamics lead to only slightly multimodal joint distributions in time, backward Gaussian updates can lead to a severe loss of information. In each application, we stressed the equation-based efficiency and dynamic characteristics of the GMM–DO smoother. We also showcased the use and capabilities of the DO equations for uncertainty predictions and illustrated how the results of the new GMM–DO smoother compare to those of the GMM–DO filter.

The present three nonlinear dynamical system applications—stochastic low dimension, reversible high dimension, and irreversible high dimension—constitute complementary and effective benchmarks for the quantitative evaluation of Bayesian smoothers with nonlinear

higher-dimensional models. This by-product of our work is useful since such an evaluation of Bayesian filtering and smoothing with models governed by stochastic partial differential equations is challenging in itself. In particular, we showed that the gains of a Bayesian smoother can be quantified and that the smoothed probability distribution estimates can be compared with analytical solutions. In the future, opportunities abound for more complex oceanic and atmospheric applications of the GMM–DO smoother. This includes Bayesian reanalyses and state estimation so as to extend previous smoothing results (e.g., Lermusiaux et al. 2002; Moore et al. 2004; Wunsch and Heimbach 2007; Khare et al. 2008). Combinations within an existing assimilation framework (e.g., Anderson et al. 2009) would also be useful. Other promising directions include the use of Bayesian smoothing for adaptive sampling (Lermusiaux 2007; Choi and How 2010; Lolla 2016) and for parameter estimation (Bocquet and Sakov 2013; Särkkä 2013).

Acknowledgments. We are grateful to the Office of Naval Research for support under Grants N00014-09-1-0676 (Science of Autonomy A-MISSION), N00014-14-1-0476 (Science of Autonomy LEARNS), N00014-13-1-0518 (Multi-DA), and N00014-14-1-0725 (Bays-DA) to the Massachusetts Institute of Technology (MIT). We thank the members of the MSEAS group at MIT for helpful suggestions and discussions. We also thank the three anonymous reviewers and the editor, Jeff Anderson, for their comments.

APPENDIX

Summary of the GMM–DO Smoother

This appendix briefly summarizes the GMM–DO smoother, a particular case of the subspace-GMM smoother introduced in Lolla and Lermusiaux (2017). The steps of the GMM–DO smoother are summarized in Table 1. All the relevant symbols and notations are described in Table 2.

Problem setup: Let $X(\mathbf{r}, t; \omega)$: $\mathbb{R}^n \times [0, T] \to \mathbb{R}$ be a continuous field governed by

$$\frac{\partial X(\mathbf{r},t;\omega)}{\partial t} = \mathcal{L}[X(\mathbf{r},t;\omega);\omega], \quad t \geq 0; \tag{A1a}$$

$$X(\mathbf{r},0;\omega) = X_0(\mathbf{r};\omega) \quad \text{(initial conditions); and} \tag{A1b}$$

$$\mathcal{B}[X(\mathbf{r},t;\omega)]|_{\mathbf{r}=\boldsymbol{\xi}} = h(\boldsymbol{\xi},t;\omega) \quad \text{(boundary conditions)}. \tag{A1c}$$

Here, $\mathcal{L}[\cdot]$ is a general nonlinear differential operator, $\mathcal{B}$ is a linear boundary operator, and ξ denotes the spatial boundary coordinate. The measurements $\mathbf{Y}(t_k; \omega)$ obey the observation model:

$$\mathbf{Y}(t_k;\omega) = \mathbf{H}\mathbf{X}(t_k;\omega) + \boldsymbol{\Upsilon}(t_k;\omega),$$
$$\boldsymbol{\Upsilon}(t_k;\omega) \sim \mathcal{N}(\cdot;\mathbf{0},\mathbf{R}), \quad k = 1, 2, \ldots, K, \tag{A2}$$

where $\mathbf{X}(t_k; \omega)$ is an $N_{\mathbf{X}} \times 1$ vector denoting the spatially discretized field $X(\mathbf{r}, t_k; \omega)$. As listed in Table 2, we denote $\mathbf{X}(t_k; \omega)$ and $\mathbf{Y}(t_k; \omega)$ by $\mathbf{X}_k$ and $\mathbf{Y}_k$, respectively. The goal is to determine the smoothed state vectors $\mathbf{X}_{k|1:K}$, that is, $\mathbf{X}_k$ conditioned on $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_K$.

Overview of GMM–DO smoother

The GMM–DO smoother computes a reduced-order representation for $\mathbf{X}_{k|1:K}$; namely, $\mathbf{X}_{k|1:K}(\omega) = \bar{\mathbf{x}}_{k|1:K} + \boldsymbol{\mathcal{X}}_k \boldsymbol{\Phi}_{k|1:K}(\omega)$, where $\bar{\mathbf{x}}_{k|1:K}(N_{\mathbf{X}} \times 1)$ is the smoothed mean, $\boldsymbol{\mathcal{X}}_k(N_{\mathbf{X}} \times s)$ is a matrix whose columns form the basis vectors of the s-dimensional stochastic subspace, and $\boldsymbol{\Phi}_{k|1:K}(\omega)$ is an $s \times 1$ coefficient vector that describes the uncertainty within the subspace. The smoother consists of the following three steps.

(i) *Forward filtering pass*—In this step, the GMM–DO filter is carried out from time $t = 0$ until $t = t_k$. The observations $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$ are assimilated as they arrive. The GMM–DO filter utilizes the DO equations to propagate the state variable $X(\mathbf{r}, t; \omega)$ forward in time. For the dynamics in (A1a)–(A1c), the DO equations take the form

$$\frac{\partial \bar{x}(\mathbf{r},t)}{\partial t} = E_\omega\{\mathcal{L}[X(\mathbf{r},t;\omega);\omega]\}, \tag{A3a}$$

$$\frac{\partial \tilde{x}_i(\mathbf{r},t)}{\partial t} = \boldsymbol{\Pi}^{\perp}(E_\omega\{\mathcal{L}[X(\mathbf{r},t;\omega);\omega]\Phi_j(t;\omega)\})C^{-1}_{\Phi_i(t)\Phi_j(t)}, \tag{A3b}$$

and

$$\frac{d\Phi_i(t;\omega)}{dt} = \langle \mathcal{L}[X(\mathbf{r},t;\omega);\omega] - E_\omega\{\mathcal{L}[X(\mathbf{r},t;\omega);\omega]\}, \tilde{x}_i(\mathbf{r},t)\rangle, \tag{A3c}$$

where $\bar{x}(\mathbf{r}, t)$, $\tilde{x}_i(\mathbf{r}, t)$, and $\Phi_i(t; \omega)$ are the components of the DO decomposition of $X(\mathbf{r}, t; \omega)$ (i.e., $X(\mathbf{r}, t; \omega) = \bar{x}(\mathbf{r}, t) + \sum_{i=1}^{s} \tilde{x}_i(\mathbf{r}, t)\Phi_i(t; \omega)$), E_ω is the expectation operator, $\langle \cdot, \cdot \rangle$ denotes the inner product $\boldsymbol{\Pi}^{\perp}[F(\mathbf{r}, t)] = F(\mathbf{r}, t) - \sum_{k=1}^{s} \langle F(\cdot, t), \tilde{x}_k(\cdot, t)\rangle \tilde{x}_k(\mathbf{r}, t)$,

and C is the covariance operator. At the assimilation times t_k, the mean vector $\bar{\mathbf{x}}_{k|1:k}$, the modes matrix $\boldsymbol{\mathcal{X}}_k$, and the realizations of the prior and posterior stochastic coefficients, namely, $\boldsymbol{\phi}^{(r)}_{k|1:k-1}$, $\boldsymbol{\phi}^{(r)}_{k|1:k}$, are stored.

(ii) *Joint-subspaces GMM fitting pass*—The realizations $\boldsymbol{\phi}^{(r)}_{k|1:k-1}$ and $\boldsymbol{\phi}^{(r)}_{k|1:k}$ are augmented to form $\boldsymbol{\phi}^{(r)}_{k,k+1|1:k}$:

$$\boldsymbol{\phi}^{(r)}_{k,k+1|1:k} = \begin{bmatrix} \boldsymbol{\phi}^{(r)}_{k|1:k} \\ \boldsymbol{\phi}^{(r)}_{k+1|1:k} \end{bmatrix}, \quad r = 1, 2, \ldots, N_r. \tag{A4}$$

For each k, a GMM is fit to the set of joint realizations $[\boldsymbol{\phi}^{(r)}_{k,k+1|1:k}]_{r=1}^{N_r}$ in accordance with the EM–BIC procedure (Sondergaard and Lermusiaux 2013a). The resulting GMMs are given by

$$p_{\boldsymbol{\Phi}_k,\boldsymbol{\Phi}_{k+1}|\mathbf{Y}_{1:k}}(\boldsymbol{\phi}_k, \boldsymbol{\phi}_{k+1} \,|\, \mathbf{y}_{1:k}) = \sum_{j=1}^{M} \pi^j \times \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\phi}_k \\ \boldsymbol{\phi}_{k+1} \end{bmatrix}; \begin{bmatrix} \bar{\boldsymbol{\mu}}^j_{k|1:k} \\ \bar{\boldsymbol{\mu}}^j_{k+1|1:k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^j_{k,k|1:k} & \boldsymbol{\Sigma}^j_{k,k+1|1:k} \\ \boldsymbol{\Sigma}^j_{k+1,k|1:k} & \boldsymbol{\Sigma}^j_{k+1,k+1|1:k} \end{bmatrix} \right). \tag{A5}$$

(iii) *Backward smoothing pass*—In this step, starting from the final time-filtered solution, $\mathbf{X}_{K|1:K}(\omega) = \bar{\mathbf{x}}_{K|1:K} + \boldsymbol{\mathcal{X}}_K \boldsymbol{\Phi}_K(\omega)$, the smoother marches backward in time to compute the smoothed mean vector $\bar{\mathbf{x}}_{k|1:K}$ and the realizations $\boldsymbol{\phi}^{(r)}_{k|1:K}$ of the smoothed stochastic coefficients $\boldsymbol{\Phi}_{k|1:K}$. At each k, the smoother draws the samples $\tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K}$ from the conditional pdfs, $p_{\boldsymbol{\Phi}_k|\boldsymbol{\Phi}_{k+1},\mathbf{Y}_{1:k}}(\cdot \,|\, \boldsymbol{\phi}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k})$, available in closed form owing to (A5). The pdfs are given by

$$p_{\boldsymbol{\Phi}_k|\boldsymbol{\Phi}_{k+1},\mathbf{Y}_{1:k}}(\boldsymbol{\phi}_k \,|\, \boldsymbol{\phi}^{(r)}_{k+1|1:K}, \mathbf{y}_{1:k}) = \sum_{j=1}^{M} \hat{\pi}^{j,(r)} \times \mathcal{N}(\boldsymbol{\phi}_k; \tilde{\boldsymbol{\mu}}^{j,(r)}_k, \tilde{\boldsymbol{\Sigma}}^j_k), \tag{A6}$$

where

$$\hat{\pi}^{j,(r)} \propto \pi^j \times \mathcal{N}(\boldsymbol{\phi}^{(r)}_{k+1|1:K}; \bar{\boldsymbol{\mu}}^j_{k+1|1:k} + \boldsymbol{\mathcal{X}}^{\mathrm{T}}_{k+1}(\bar{\mathbf{x}}_{k+1|1:k} - \bar{\mathbf{x}}_{k+1|1:K}), \boldsymbol{\Sigma}^j_{k+1,k+1|1:k}), \tag{A7a}$$

$$\tilde{\boldsymbol{\mu}}^{j,(r)}_k = \bar{\boldsymbol{\mu}}^j_{k|1:k} + \tilde{\mathbf{K}}^j[\boldsymbol{\phi}^{(r)}_{k+1|1:K} - \bar{\boldsymbol{\mu}}^j_{k+1|1:k} - \boldsymbol{\mathcal{X}}^{\mathrm{T}}_{k+1}(\bar{\mathbf{x}}_{k+1|1:k} - \bar{\mathbf{x}}_{k+1|1:K})], \quad \text{and} \tag{A7b}$$

$$\tilde{\boldsymbol{\Sigma}}^j_k = \boldsymbol{\Sigma}^j_{k,k|1:k} - \tilde{\mathbf{K}}^j \boldsymbol{\Sigma}^j_{k+1,k|1:k}. \tag{A7c}$$

The smoothed mean state $\bar{\mathbf{x}}_{k|1:K}$ is computed as

$$\bar{\mathbf{x}}_{k|1:K} = \bar{\mathbf{x}}_{k|1:k} + \boldsymbol{\mathcal{X}}_k \times \left[\frac{1}{N_r} \sum_{r=1}^{N_r} \tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K} \right], \tag{A8}$$

and the smoothed coefficients $\boldsymbol{\phi}^{(r)}_{k|1:K}$ are given by

$$\boldsymbol{\phi}^{(r)}_{k|1:K} = \tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K} - \frac{1}{N_r} \sum_{r=1}^{N_r} \tilde{\boldsymbol{\phi}}^{(r)}_{k|1:K}, \quad r = 1, 2, \ldots, N_r. \tag{A9}$$

REFERENCES

Anderson, J., T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, doi:10.1175/2009BAMS2618.1.

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press, 346 pp.

Bocquet, M., 2005: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **131**, 2191–2208, doi:10.1256/qj.04.67.

——, and P. Sakov, 2013: Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes Geophys.*, **20**, 803–818, doi:10.5194/npg-20-803-2013.

Choi, H.-L., and J. P. How, 2010: Continuous trajectory planning of mobile sensors for informative forecasting. *Automatica*, **46**, 1266–1275, doi:10.1016/j.automatica.2010.05.004.

Cohn, S. E., N. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867, doi:10.1175/1520-0493(1994)122<2838:AFLKSF>2.0.CO;2.

Cosme, E., J. Verron, P. Brasseur, J. Blum, and D. Auroux, 2012: Smoothing problems in a Bayesian framework and their linear Gaussian solutions. *Mon. Wea. Rev.*, **140**, 683–695, doi:10.1175/MWR-D-10-05025.1.

Durran, D., 1999: *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*. Texts in Applied Mathematics, Vol. 32, Springer, 465 pp.

Durst, F., A. Melling, and J. Whitelaw, 1974: Low Reynolds number flow over a plane symmetric sudden expansion. *J. Fluid Mech.*, **64**, 111–128, doi:10.1017/S0022112074002035.

Evensen, G., and P. J. Van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

Eyink, G., and S. Kim, 2006: A maximum entropy method for particle filtering. *J. Stat. Phys.*, **123**, 1071–1128, doi:10.1007/s10955-006-9124-9.

Fearn, R., T. Mullin, and K. Cliffe, 1990: Nonlinear flow phenomena in a symmetric sudden expansion. *J. Fluid Mech.*, **211**, 595–608, doi:10.1017/S0022112090001707.

Feppon, F., and P. F. J. Lermusiaux, 2017: A geometric approach to dynamical model-order reduction. *SIAM J. Sci. Comput.*, in press.

Gelb, A., 1974: *Applied Optimal Estimation*. The MIT Press, 374 pp.

Haley, P. J., Jr., and Coauthors, 2009: Forecasting and reanalysis in the Monterey Bay/California Current region for the Autonomous Ocean Sampling Network-II experiment. *Deep-Sea Res. II*, **56**, 127–148, doi:10.1016/j.dsr2.2008.08.010.

Higham, D. J., 2001: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, **43**, 525–546, doi:10.1137/S0036144500378302.

Jazwinski, A. H., 2007: *Stochastic Processes and Filtering Theory*. Dover Publications, 400 pp.

Khare, S. P., J. L. Anderson, T. J. Hoar, and D. Nychka, 2008: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus*, **60A**, 97–112, doi:10.1111/j.1600-0870.2007.00281.x.

Lermusiaux, P. F. J., 1999a: Data assimilation via error subspace statistical estimation. Part II: Mid-Atlantic Bight shelfbreak front simulations and ESSE validation. *Mon. Wea. Rev.*, **127**, 1408–1432, doi:10.1175/1520-0493(1999)127<1408:DAVESS>2.0.CO;2.

——, 1999b: Estimation and study of mesoscale variability in the Strait of Sicily. *Dyn. Atmos. Oceans*, **29**, 255–303, doi:10.1016/S0377-0265(99)00008-1.

——, 2002: On the mapping of multivariate geophysical fields: Sensitivities to size, scales, and dynamics. *J. Atmos. Oceanic Technol.*, **19**, 1602–1637, doi:10.1175/1520-0426(2002)019<1602:OTMOMG>2.0.CO;2.

——, 2007: Adaptive modeling, adaptive data assimilation and adaptive sampling. *Physica D*, **230**, 172–196, doi:10.1016/j.physd.2007.02.014.

——, and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.*, **127**, 1385–1407, doi:10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.

——, D. G. M. Anderson, and C. J. Lozano, 2000: On the mapping of multivariate geophysical fields: Error and variability subspace estimates. *Quart. J. Roy. Meteor. Soc.*, **126**, 1387–1429, doi:10.1256/smsqj.56509.

——, A. R. Robinson, P. J. Haley, and W. G. Leslie, 2002: Advanced interdisciplinary data assimilation: Filtering and smoothing via error subspace statistical estimation. *Proc. OCEANS 2002 MTS/IEEE Conf.*, Biloxi, MS, IEEE, 795–802.

——, and Coauthors, 2006: Quantifying uncertainties in ocean predictions. *Oceanography*, **19**, 92–105, doi:10.5670/oceanog.2006.93.

Lolla, T., 2016: Path planning and adaptive sampling in the coastal ocean. Ph.D. thesis, Massachusetts Institute of Technology, 315 pp. [Available online at https://dspace.mit.edu/handle/1721.1/103438.]

——, and P. F. J. Lermusiaux, 2017: A Gaussian mixture model smoother for continuous nonlinear stochastic dynamical systems: Theory and scheme. *Mon. Wea. Rev.*, **145**, 2743–2761, doi:10.1175/MWR-D-16-0064.1.

Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167–194, doi:10.1034/j.1600-0870.1999.t01-2-00002.x.

Moore, A. M., H. G. Arango, E. Di Lorenzo, B. D. Cornuelle, A. J. Miller, and D. J. Neilson, 2004: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modell.*, **7**, 227–258, doi:10.1016/j.ocemod.2003.11.001.

Ramp, S. R., and Coauthors, 2009: Preparing to predict: The second Autonomous Ocean Sampling Network (AOSN-II) experiment in the Monterey Bay. *Deep-Sea Res. II*, **56**, 68–86, doi:10.1016/j.dsr2.2008.08.013.

Sapsis, T. P., and P. F. J. Lermusiaux, 2009: Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D*, **238**, 2347–2360, doi:10.1016/j.physd.2009.09.017.

Särkkä, S., 2013: *Bayesian Filtering and Smoothing*. Institute of Mathematical Statistics Textbooks, Vol. 3, Cambridge University Press, 232 pp.

Sondergaard, T., and P. F. J. Lermusiaux, 2013a: Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. *Mon. Wea. Rev.*, **141**, 1737–1760, doi:10.1175/MWR-D-11-00295.1.

——, and ——, 2013b: Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part II: Applications. *Mon. Wea. Rev.*, **141**, 1761–1785, doi:10.1175/MWR-D-11-00296.1.

Ueckermann, M. P., P. F. J. Lermusiaux, and T. P. Sapsis, 2013: Numerical schemes for dynamically orthogonal equations of stochastic fluid and ocean flows. *J. Comput. Phys.*, **233**, 272–294, doi:10.1016/j.jcp.2012.08.041.

Van Leer, B., 1977: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection. *J. Comput. Phys.*, **23**, 276–299, doi:10.1016/0021-9991(77)90095-X.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem*. Cambridge University Press, 442 pp.

——, and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, **230**, 197–208, doi:10.1016/j.physd.2006.09.040.