Physica D 427 (2021) 133003

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Bayesian learning of stochastic dynamical models

Peter Lu, Pierre F.J. Lermusiaux *

Massachusetts Institute of Technology, Department of Mechanical Engineering, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 6 July 2021
Accepted 11 August 2021
Available online 5 September 2021

Keywords:
Bayesian data assimilation
Learning
GMM-DO
Dynamical system
Stochastic PDEs
Ocean and weather prediction

ABSTRACT

A new methodology for rigorous Bayesian learning of high-dimensional stochastic dynamical models is developed. The methodology performs parallelized computation of marginal likelihoods for multiple candidate models, integrating over all state variable and parameter values, and enabling a principled Bayesian update of model distributions. This is accomplished by leveraging the dynamically orthogonal (DO) evolution equations for uncertainty prediction in a dynamic stochastic subspace and the Gaussian Mixture Model-DO filter for inference of nonlinear state variables and parameters, using reduced-dimension state augmentation to accommodate models featuring uncertain parameters. Overall, the joint Bayesian inference of the state, model equations, geometry, boundary conditions, and initial conditions is performed. Results are exemplified using two high-dimensional, nonlinear simulated fluid and ocean systems. For the first, limited measurements of fluid flow downstream of an obstacle are used to perform joint inference of the obstacle's shape, the Reynolds number, and the $\mathcal{O}(10^5)$ fluid velocity state variables. For the second, limited measurements of the concentration of a microorganism advected by an uncertain flow are used to perform joint inference of the microorganism's reaction equation and the $\mathcal{O}(10^5)$ microorganism concentration and ocean velocity state variables. When the observations are sufficiently informative about the learning objectives, we find that our posterior model probabilities correctly identify either the true model or the most plausible models, even in cases where a human would be challenged to do the same.

1. Introduction

Stochastic dynamical systems [1] are everywhere, from oceanic and ecological systems, power grids and communications networks, to financial markets and social networks. The mathematical tools that have been developed for investigating stochastic dynamical systems are thus highly versatile and have been applied in a wide range of fields [2–13].

Quantitative investigations of a stochastic dynamical system typically assume that the mathematical model formulated for the system is an accurate description of its governing processes. Uncertainty in the system's state variables are often assumed to originate solely from uncertainty in the system's initial and boundary conditions, and stochastic forcings with known statistical properties. This assumption of absolute model formulation validity however is not always defensible. For example, when dealing with complex systems for which governing equations have not yet been derived from known first principles, the assumption is surely inappropriate. In general, uncertainty in model formulation can originate from the choice of state variables themselves, from the functional forms of the model equations, boundary conditions or initial conditions, and from the definition of the (spatial) domain of integration. Both the deterministic and stochastic components of the model formulation can be uncertain. In what follows, when possible, we will refer to model formulation uncertainty simply as model uncertainty.

Model uncertainty can be difficult to quantify and is thus often ignored. This is not damaging when model uncertainty is insignificant. For example, when we model ballistic dynamics on Earth, one can have confidence in Newton's laws of motion. In other cases however, it can lead to significant underestimation of uncertainty. [14,15], and [16] review poignant examples from statistics in which ignorance of model uncertainty resulted in overconfidence in state estimates, which subsequently led to tragically flawed conclusions.

Ignoring model uncertainty is also antithetic to the scientific method, which entails the comparison of competing hypotheses by means of observations. If multiple models are considered, the observations used to perform inference with one model can also be used to learn the relative validity of each of the models. This process of model learning can reveal valuable insights regarding the fundamental mechanisms of the system. If model uncertainty is ignored however and only one model – one hypothesis – is assumed, this opportunity for scientific discovery in the classic sense is forfeited.

* Corresponding author.
E-mail address: pierrel@mit.edu (P.F.J. Lermusiaux).

https://doi.org/10.1016/j.physd.2021.133003
0167-2789/

1.1. Progress to date

Several methods have been obtained to handle the coupled issues of model uncertainty and model learning in stochastic dynamical systems. Directed search methods for model learning typically proceed by first performing state variable inference for a large set of candidate models, then scoring the inferred state variables relative to observations using metrics derived from frequentist statistics. Computational schemes are employed to search through expansive sets of plausible candidate models, with the search process directed by results from successive rounds of candidate evaluations. A premier example of this strategy is [17]. In this work, the authors employed a heuristic optimization scheme known as symbolic regression [18] to search through a space of algebraic expressions with the goal of finding the fundamental physical laws that govern several simple dynamical systems, such as single and double pendula. The authors were able to identify conservation laws for energy and momentum without any prior information regarding the laws' functional forms. Their approach is highly versatile but exceptionally demanding in terms of computational cost, even for the low-dimensional systems they considered. Model learning for the double pendulum system, a non-linear system with two state variables, required over 30 h of computational time in a 32-core parallelized implementation. Extensions of their approach, and other directed search model learning methods (e.g. [19–23]), to high-dimensional systems will likely prove to be computationally challenging.

Hierarchical Bayesian modeling is a general approach to handling model uncertainty whereby full stochastic dynamical models are represented as hierarchies of simpler, analytically tractable sub-models [24,25]. If these sub-models are properly formulated, inference can be performed separately for each by exploiting their conditional independences, with the sub-models aggregated afterwards to achieve global model learning. An oceanographic application of this approach is demonstrated in [26], where the authors used a hierarchical Bayesian model to formulate a stochastic dynamical model of the surface wind streamfunction over a region of the Labrador Sea using satellite surface wind velocity data. The aggregate wind model was decomposed into sub-models for observational data, boundary conditions, and the numerical streamfunction, which enabled the quantification of boundary condition uncertainties on the posterior distribution of streamfunction values. An ecological application of this hierarchical Bayesian modeling is [27], where the authors predict the spatial distribution of ground flora based on sparse data. Sub-models were formulated that enabled the incorporation of geographic covariates, a source of model uncertainty that, when accounted for, significantly enhanced flora distribution predictions. Further applications of hierarchical Bayesian modeling to problems of spatiotemporal statistics include [28–31], and [32]. Multiresolution Bayesian modeling, a variation of the approach for application to signal and image processing, is reviewed in [33,34], and [35]. In [36], an extension of the hierarchical formulation to graphical models allows for more complex interdependencies between model components, at the cost of more computationally intensive inference algorithms.

Though many model learning techniques, such as the directed search methods, can be effective when system dimensions are small and candidate model spaces are readily explored, computational difficulties arise when they are applied to high-dimensional systems, such as those encountered in geosciences [37,38]. Reduced-order modeling (ROM) techniques [39,40] are designed to find low-dimensional representations of high-dimensional models, for which model uncertainty quantification and inference are more readily performed. They include proper orthogonal decomposition (POD) [41–43], centroidal Voronoi tessellation [44], neural networks [45,46], Volterra series [47], kriging [48], dynamic data-driven ROMs [49], certified reduced basis methods [50], empirical emulators [51,52], error subspace statistical estimation [53,54] and its use in adaptive modeling [55–57], and the dynamically orthogonal (DO) evolution equations [58–60].

Machine learning methods have been applied recently to the discovery of model equations. The sparse regression-based methods (SINDy) [61,62] are promising as they do not require prior knowledge, but they often need large data sets. Variations of SINDy have been obtained such as weak SINDy to learn PDEs [63], adaptive generation of features to expand the library of models [64], and extensions to Bayesian identification [65]. Methods such as DeepMOD/DL-PDE/DLGA-PDE [66–70] leverage neural networks to eliminate the need for data with high temporal resolution. Some methods use genetic algorithms [71] and reinforcement learning [72–74] to search in the space of potential models. However, most of these methods do not provide uncertainty estimates for the discovered models. Schemes have also combined prior knowledge about underlying governing equations for model recovery and refinement. For example, [75] successfully used Gaussian processes to learn the values of the parametric response of partially-known differential equations. PDENet [76], which connects differential operators and convolutional filters to learn spatial derivatives, can now relax the need for prior information about the form of the PDE by using a symbolic neural network [77]. Finally, physics-based machine learning have been combined with POD reductions [78] and Long Short-Term Memory networks have resolved improperly modeled dynamics [79].

The challenges for many methods however are to predict and evolve the probability distribution of high-dimensional fields accurately, capturing nonlinear dynamics and non-Gaussian statistics, to assimilate the observed information with Bayes rules in accord with all prior uncertainties, and to accurately marginalize for the posterior distribution of the models and thus achieve rigorous Bayesian learning of model formulations.

1.2. Problem statement

System: We consider a system with state vector $\boldsymbol{X} \in \mathbb{R}^{N_{\boldsymbol{X}}}$ governed by an uncertain stochastic dynamical model $\mathcal{M}$ with uncertain parameter vector $\boldsymbol{\Theta} \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$, where $N_{\boldsymbol{X}}$ and $N_{\boldsymbol{\Theta}} \in \mathbb{N}$ are the dimensions of the state and parameter vectors. For realizations $\boldsymbol{x}$, $\boldsymbol{\theta}$, and $\mathcal{M}_n$ of $\boldsymbol{X}$, $\boldsymbol{\Theta}$, and $\mathcal{M}$ respectively, we have,

$$\frac{\mathrm{d}\boldsymbol{x}(t;\omega)}{\mathrm{d}t} = \mathcal{M}_n\left[\,\boldsymbol{x}(t;\omega), \boldsymbol{\theta}(\omega), t; \omega\,\right] \tag{1}$$

$$\mathcal{M}_n \equiv \left(\mathcal{D}_n, \mathrm{SG}_n, \mathrm{BC}_n, \mathrm{IC}_n\right), \tag{2}$$

where t denotes time, ω an index over stochastic realizations (a random event), $\mathcal{D}_n$ the set of stochastic dynamical equations, SG_n the geometry (e.g. spatial domain), BC_n the boundary conditions, and IC_n the initial conditions. All of the model components represented in (2) are allowed to be uncertain. They can all contain uncertain formulations (e.g. candidate functions for the model, domain shape, BC or IC). The joint probability distribution over $\boldsymbol{X}$, $\boldsymbol{\Theta}$, and $\mathcal{M}$ is denoted $p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n)$ and is time-dependent.

Observations: Stochastic observations $\boldsymbol{Y} \in \mathbb{R}^{N_{\boldsymbol{Y}}}$ are available at data times. Their probability distribution is conditionally independent of both the model and parameters, given state variables

$$p_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{y}\,|\,\boldsymbol{x},\boldsymbol{\theta},\mathcal{M}_n) = p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}\,|\,\boldsymbol{x}) \equiv \mathcal{L}\,(\boldsymbol{y}\mid\boldsymbol{x}) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{3}$$

where $N_{\boldsymbol{Y}} \in \mathbb{N}$ is the dimension of the observation vector, $\boldsymbol{y}$ a realization of the observation vector, and $\mathcal{L}\,(\boldsymbol{y}\mid\bullet)$ the observation likelihood function or observation model for the system.

Denoting by $\mathcal{N}(\bullet;\, \boldsymbol{\mu},\, \boldsymbol{\Sigma})$ a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, a particular observation likelihood function we will use is,

$$\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{y};\, \mathbf{H}\boldsymbol{x},\, \mathbf{R}) \quad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{4}$$

where $\mathbf{H} \in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}}$ is the linear observation matrix and $\mathbf{R} \in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{Y}}}$ the observation covariance matrix. This is equivalent to,

$$\boldsymbol{Y} = \mathbf{H}\boldsymbol{X} + \boldsymbol{V}\,,$$

where $\boldsymbol{V} \in \mathbb{R}^{N_{\boldsymbol{Y}}}$ represents zero-mean Gaussian noise with covariance $\mathbf{R}$.

Goal: Our goal is to address the three challenges, specifically:

1. Efficiently evolve $p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{x}, \boldsymbol{\theta}, \mathcal{M}_n)$ in time, accounting for all forms of uncertainty encapsulated in (1) and (2).
2. When observations are available, use (3) and perform the Bayesian update of the joint probability distribution,

   $$p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}}(\boldsymbol{x}, \boldsymbol{\theta}, \mathcal{M}_n) \rightarrow p_{\boldsymbol{X},\boldsymbol{\Theta},\mathcal{M}|\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{\theta}, \mathcal{M}_n \,|\, \boldsymbol{y})$$

3. Finally, obtain marginal distributions, especially the model distribution $p_{\mathcal{M}}(\mathcal{M}_n)$ and its Bayesian update when observations are available,

   $$p_{\mathcal{M}}(\mathcal{M}_n) \rightarrow p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n \,|\, \boldsymbol{y})$$

2. Methodology

We now define the computations needed for Bayesian model learning including marginal likelihoods and then develop our rigorous Bayesian learning for high-dimensional nonlinear stochastic dynamical models.

2.1. Bayesian model learning

2.1.1. Candidate models

A common approach to represent model uncertainty is to formulate a finite set of candidates for the true model. These candidates or beliefs may be derived from first principles, inspired by previous observations, or based on a combination of theoretical and empirical prior knowledge. A discrete probability distribution $p_{\mathcal{M}}(\bullet)$ can be defined over the set of candidate models to represent the probabilities that each of the candidates are the true model. Each candidate model can be used to predict the evolution of the system's state, independent of the other models, producing state variable probability distributions that are conditional on the candidates being the true model. If the candidates are assumed to be independent, a state variable probability distribution that accounts for the uncertainty in the formulation of the system's model can be estimated at any time as simply the weighted average of these conditional distributions,

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x} \,|\, \mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{5}$$

where $p_{\boldsymbol{X}}(\bullet)$ is the state variable distribution, $N_{\mathcal{M}} \in \mathbb{N}$ the total number of candidate models, $\mathcal{M}_n$ the nth candidate model, and $p_{\boldsymbol{X}|\mathcal{M}}(\bullet \,|\, \mathcal{M}_n)$ the state variable distribution conditioned on $\mathcal{M}_n$ (the nth model-conditional state variable distribution). Versions of (5) have been used before and are known by many names, including Bayesian model averaging [15,80,81], multimodel estimation [82,83], multimodel fusion [84], and (multimodel or super-) ensemble modeling [85–87]. If the candidate models $\mathcal{M}_n$ are correlated or if the space of model formulation/structures is continuous (instead of discrete as in (5)), the distribution (5) becomes a correlated weighting or an integral over the continuous model formulation/structures. Our formalism directly extends to these cases.

If (5) holds, its linear nature w.r.t. candidate models leads to several useful properties (see [88]). First, marginal distributions for subsets of state variables can be found as weighted averages of the corresponding model-conditional marginal distributions, i.e. if $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}^{\mathrm{T}}$ with $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ mutually exclusive, we have $p_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}_1|\mathcal{M}}(\boldsymbol{x}_1 \,|\, \mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n)$. Second, the state variable mean can be found as a weighted average of the model-conditional state variable means, i.e. $\mathrm{E}[\boldsymbol{X}] = \sum_{n=1}^{N_{\mathcal{M}}} \mathrm{E}[\boldsymbol{X} \,|\, \mathcal{M}_n]\, p_{\mathcal{M}}(\mathcal{M}_n)$.

2.1.2. Joint state and model learning

When observations of state variables are made, both the model-conditional state variable distributions and model distribution within (5) can theoretically be updated using Bayes' theorem [89],

$$p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x} \,|\, \boldsymbol{y}, \mathcal{M}_n) = \frac{p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y} \,|\, \boldsymbol{x}, \mathcal{M}_n)}{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n)}\, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x} \,|\, \mathcal{M}_n) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}, \forall n \in \{1, \ldots, N_{\mathcal{M}}\}\,, \tag{6}$$

$$p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n \,|\, \boldsymbol{y}) = \frac{p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n)}{p_{\boldsymbol{Y}}(\boldsymbol{y})}\, p_{\mathcal{M}}(\mathcal{M}_n) \qquad \forall n \in \{1, \ldots, N_{\mathcal{M}}\}\,. \tag{7}$$

In (6), the model $\mathcal{M}_n$ plays the role of a 'given parameter'. The distributions $p_{\boldsymbol{X}|\mathcal{M}}(\bullet \,|\, \mathcal{M}_n)$ and $p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\bullet \,|\, \boldsymbol{y}, \mathcal{M}_n)$ are the prior and posterior conditional state variable distributions for the nth candidate model, while $p_{\mathcal{M}}(\bullet)$ and $p_{\mathcal{M}|\boldsymbol{Y}}(\bullet \,|\, \boldsymbol{y})$ are the prior and posterior model distributions, respectively.

If the candidate models are assumed independent, as in (5), the posterior model-conditional state variable distributions and model distribution can be combined to form the posterior state variable distribution,

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{X}|\boldsymbol{Y},\mathcal{M}}(\boldsymbol{x} \,|\, \boldsymbol{y}, \mathcal{M}_n) p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n \,|\, \boldsymbol{y}) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,. \tag{8}$$

Equation (6) represents Bayesian state variable inference and can be performed for each model-conditional state variable distribution independently, ignoring model uncertainty. Techniques for state variable inference abound, ranging from the Kalman filter [90,91] to particle filters [92], Markov chain Monte Carlo (MCMC) algorithms [93], and forward–backward algorithms [94]. State variable inference and data assimilation have roots in optimal estimation theory e.g. [95–97] and control theory [98,99], with now many applications in environmental sciences and engineering (e.g. [12,24,100–105]).

Equation (7) represents Bayesian model learning. The comparison of the posteriors for each $\mathcal{M}_n$, $p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n \,|\, \boldsymbol{y})$, is Bayesian hypothesis testing for competing models and each $p_{\mathcal{M}|\boldsymbol{Y}}(\mathcal{M}_n \,|\, \boldsymbol{y})$ is often referred to as model evidence. Though cosmetically simpler than (6), (7) is in fact the more challenging of the two Bayesian updates to perform. The chief difficulty lies in the calculation of the marginal likelihood $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n)$, which represents the strength of the observational evidence for $\mathcal{M}_n$, i.e. the likelihood of model $\mathcal{M}_n$, for all states $\boldsymbol{X}$. While $p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y} \,|\, \boldsymbol{x}, \mathcal{M}_n)$ is equivalent to the observation likelihood function $\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x})$ – the function that defines the probability distribution for observations when state variables are known – an explicit expression for the likelihood $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n)$ is not available. Instead, $p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n)$ (the probability distribution for the observation vector realization conditioned on a given candidate model) must be found through difficult, very large-dimension integrations [14,24,106],

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y} \,|\, \mathcal{M}_n) = \int p_{\boldsymbol{Y}|\boldsymbol{X},\mathcal{M}}(\boldsymbol{y} \,|\, \boldsymbol{x}, \mathcal{M}_n)\, p_{\boldsymbol{X}|\mathcal{M}}(\boldsymbol{x} \,|\, \mathcal{M}_n)\, \mathrm{d}\boldsymbol{x}$$

$$= \int \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x})\, p_{\boldsymbol{X}\mid\mathcal{M}}(\boldsymbol{x}\mid\mathcal{M}_n)\, \mathrm{d}\boldsymbol{x} \qquad \forall n \in \{1,\ldots,N_{\mathcal{M}}\}\,. \tag{9}$$

Note that although $p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y}\mid\mathcal{M}_n)$ also appears in (6), there it serves only as the normalization constant for the posterior conditional state variable distribution and its explicit calculation is usually side-stepped by Bayesian state variable inference schemes. Likewise, $p_{\boldsymbol{Y}}(\boldsymbol{y})$ appearing in (7) is not of concern for learning, as it serves only as the normalization constant for the posterior model distribution. Once all marginal likelihoods have been found, $p_{\boldsymbol{Y}}(\boldsymbol{y})$ can be computed as simply their weighted summation,

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{n=1}^{N_{\mathcal{M}}} p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y}\mid\mathcal{M}_n)\, p_{\mathcal{M}}(\mathcal{M}_n)\,. \tag{10}$$

2.2. Marginal likelihood computation

Several approximations have been used to calculate the integral in (9). A simple one used in control theory is to approximate both $\mathcal{L}(\boldsymbol{y}\mid\bullet)$ and $p_{\boldsymbol{X}\mid\mathcal{M}}(\bullet\mid\mathcal{M}_n)$ as Gaussian distributions, in which case an analytical expression for (9) can be found [82,83]. This handles poorly non-Gaussian state variable distributions as well as high-dimensional systems with large covariance matrices. When the functional forms of $\mathcal{L}(\boldsymbol{y}\mid\bullet)$ and $p_{\boldsymbol{X}\mid\mathcal{M}}(\bullet\mid\mathcal{M}_n)$ do not allow for the analytical calculation of the integral in (9), an alternative is to use closed-form approximations of the integral that are exact in the limit of infinite observations [16]. The accuracy of these approximations generally increases as the density of $\mathcal{L}(\boldsymbol{y}\mid\bullet)\, p_{\boldsymbol{X}\mid\mathcal{M}}(\bullet\mid\mathcal{M}_n)$ increases near its maximum (i.e. the more the density peaks, the better). These asymptotic approximations are thus best-suited for systems featuring unimodal state variable distributions and large numbers of observations. Unfortunately, in high-dimensional, nonlinear systems such as those encountered in oceanography and meteorology, multimodality and sparse observations are the norm (e.g. [10,103,107–110]).

For cases where analytical solutions to (9) are not available and asymptotic approximations are inappropriate, computational techniques are used. They include importance sampling [111,112], bridge sampling [113], path sampling [114], annealed importance sampling [115], and nested sampling [116,117]. Though these methods are generally robust to variations in the functional forms of $\mathcal{L}(\boldsymbol{y}\mid\bullet)$ and $p_{\boldsymbol{X}\mid\mathcal{M}}(\bullet\mid\mathcal{M}_n)$, they remain limited in the dimension of the systems they can handle, e.g. $\mathcal{O}(10^2)$ [114]. This falls short of the $\mathcal{O}(10^5$–$10^9)$ state variables frequently encountered in geophysical sciences [37]. In order to perform our Bayesian model learning, the major challenge is thus to obtain inference schemes that efficiently reduce dimensions, capturing and exploiting dominant nonlinear dynamics and non-Gaussian statistics, and that accurately compute the marginal likelihood integral in (9). This is addressed next.

2.3. Bayesian learning schemes

To enable principled Bayesian model learning for high-dimensional nonlinear stochastic dynamical systems using (9), we utilize the DO evolution equations and the Gaussian Mixture Model (GMM)-DO filter. The DO equations achieve an optimal dynamic dimension-reduction [60] and evolve the dominant model-conditional state variable distributions introduced in (5). The GMM-DO filter [118,119] is first employed to perform the Bayesian updates (6) of the model-conditional distributions in the evolving DO subspaces. In doing so, reduced-dimension state augmentation is used to also update and learn uncertain model parameters. The marginal likelihoods (9) are then calculated analytically using the GMMs in the DO subspace of each candidate model, providing the structural flexibility to represent nonlinear multimodal state variable distributions. Finally, these marginal likelihoods are used to perform the Bayesian update (7) of the model distribution, thus accomplishing Bayesian model learning. Each of these components is outlined next, with an integrated account provided at the end. A compendium of the notation is provided in Table 3.

2.3.1. DO evolution equations

The DO differential equations provide an instantaneously optimal dynamic reduction of nonlinear stochastic differential equations [59,60,120,121]. The stochastic dynamical state vector is truncated using a DO expansion (a generalized, time-dependent Karhunen–Loeve decomposition [122]),

$$\boldsymbol{x}(t;\omega) \approx \bar{\boldsymbol{x}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\, \tilde{\boldsymbol{x}}_i(t)\,, \tag{11}$$

where $\bar{\boldsymbol{x}}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}}$ is the state vector mean, $\tilde{\boldsymbol{x}}_i(t) \in \mathbb{R}^{N_{\boldsymbol{X}}}$ the ith of N_{DO} orthonormal basis vectors, and $\phi_i(t;\omega) \in \mathbb{R}$ the ith of N_{DO} zero-mean stochastic processes. The basis vectors and stochastic processes are referred to as the DO modes and coefficients, respectively. The DO modes define a stochastic subspace $\mathbf{V}_{\mathrm{DO}} = \mathrm{span}\{\tilde{\boldsymbol{x}}_i(t)\}_{i=1}^{N_{\mathrm{DO}}}$ embedded in $\mathbb{R}^{N_{\boldsymbol{X}}}$ within which the majority of the state uncertainty resides. At any given time, a reduced-dimension probability distribution for the N_{DO} coefficients then efficiently represents the full probability distribution for the N_{X} state variables, as (11) relates the two sets of variables through an affine transformation.

To evolve the probability density of the state vector, equations for the terms in the expansion (11) are obtained from the original stochastic dynamical model equations governing the evolution of the state vector

$$\frac{\mathrm{d}\boldsymbol{x}(t;\omega)}{\mathrm{d}t} = \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\,. \tag{12}$$

We assume for now that the true model for the system is known and hence use $\mathcal{M}$ in (12) as opposed to the $\mathcal{M}_n$ used in (1). Specifically, evolution equations for $\bar{\boldsymbol{x}}(t)$, $\tilde{\boldsymbol{x}}_i(t)$, and $\phi_i(t;\omega)$ [59] are obtained by insertion of (11) into (12), constraining the evolution of the DO vectors to directions orthogonal to $\mathbf{V}_{\mathrm{DO}}$

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} \perp \mathbf{V}_{\mathrm{DO}} \Leftrightarrow \left\langle \frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t}\,,\, \tilde{\boldsymbol{x}}_j(t) \right\rangle = 0 \qquad \forall i,j \in \{1,\ldots,N_{\mathrm{DO}}\}\,, \tag{13}$$

where the operator $\langle \mathbf{a}, \mathbf{b}\rangle$ represents the vector inner product of $\mathbf{a}$ and $\mathbf{b}$. Using (13) in conjunction with the expansion (11) and stochastic dynamical model (12), the DO evolution equations can be derived for the state vector mean, DO vectors, and coefficients,

$$\frac{\mathrm{d}\bar{\boldsymbol{x}}(t)}{\mathrm{d}t} = \mathrm{E}\left[\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\,, \tag{14}$$

$$\frac{\mathrm{d}\tilde{\boldsymbol{x}}_i(t)}{\mathrm{d}t} = \sum_{j=1}^{N_{\mathrm{DO}}} \mathbf{C}^{-1}_{(i,j)}\, \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}} \left[\,\mathrm{E}\left[\,\phi_j(t;\omega)\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right]\right] \qquad \forall i \in \{1,\ldots,N_{\mathrm{DO}}\}\,, \tag{15}$$

$$\frac{\mathrm{d}\phi_i(t;\omega)}{\mathrm{d}t} = \left\langle \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] - \mathrm{E}\left[\,\mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right]\right],\, \tilde{\boldsymbol{x}}_i(t)\right\rangle \qquad \forall i \in \{1,\ldots,N_{\mathrm{DO}}\}\,, \tag{16}$$

where $\mathrm{E}[\bullet]$ represents the expectation operator, $\mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}[\mathbf{a}]$ represents the projection of the vector $\mathbf{a}$ onto the space orthogonal to $\mathbf{V}_{\mathrm{DO}}$

$$\mathcal{P}_{\mathbf{V}_{\mathrm{DO}}^{\perp}}[\mathbf{a}] = \mathbf{a} - \mathcal{P}_{\mathbf{V}_{\mathrm{DO}}}[\mathbf{a}] = \mathbf{a} - \sum_{k=1}^{N_{\mathrm{DO}}} \left\langle \mathbf{a}, \tilde{\boldsymbol{x}}_k(t) \right\rangle \tilde{\boldsymbol{x}}_k(t)$$

and $\mathbf{C}^{-1}_{(i,j)}$ is the (i,j)th entry of the inverse of the DO coefficient covariance matrix $\mathbf{C}_{(i,j)} = \mathrm{E}\left[\phi_i(t;\omega)\phi_j(t;\omega)\right]$.

Numerical implementations of the deterministic (14) and (15) can be based on classic solvers [121,123]. The stochastic evolution of the DO coefficients (16) can be carried out using Monte Carlo (MC) sampling methods, whereby $N_{\mathrm{MC}} \gg N_{\mathrm{DO}}$ samples are drawn from the initial coefficient distribution and evolved by solving (16) as an ODE of dimension N_{DO} for each sample [123]. These evolved samples can constitute a rich description of this distribution since $N_{\mathrm{MC}} \gg N_{\mathrm{DO}}$.

2.3.2. GMM-DO filter

The affine transformation (11) between DO coefficients and state variables can be written as

$$\boldsymbol{x}(t;\omega) \approx \bar{\boldsymbol{x}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\tilde{\boldsymbol{x}}_i(t) = \bar{\boldsymbol{x}}(t) + \boldsymbol{\mathcal{X}}(t)\boldsymbol{\phi}(t;\omega)\,, \tag{17}$$

where $\boldsymbol{\mathcal{X}}(t) \in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\mathrm{DO}}}$ is a matrix whose columns are DO vectors and $\boldsymbol{\phi}(t;\omega) \in \mathbb{R}^{N_{\mathrm{DO}}}$ a realization of the vector of DO coefficients $\boldsymbol{\Phi}$. The GMM-DO filter takes advantage of (17) and proceeds as follows [118,119].

When observations are made (3), the prior probability distribution for DO coefficients in the DO subspace is approximated using a GMM,

$$p_{\boldsymbol{\Phi}}(\boldsymbol{\phi}) \approx \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi},j} \times \mathcal{N}\left(\boldsymbol{\phi};\, \boldsymbol{\mu}_{\boldsymbol{\Phi},j},\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\right) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\mathrm{DO}}}\,, \tag{18}$$

where N_{GMM} is the to-be-determined number of GMM components, $\pi_{\boldsymbol{\Phi},j}$ the jth component weight, $\boldsymbol{\mu}_{\boldsymbol{\Phi},j}$ the jth component mean vector, and $\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$ the jth component covariance matrix. This approximation is found by performing a semiparametric fit to the Monte Carlo samples used to numerically evolve the stochastic coefficients. Specifically, the expectation–maximization (EM) algorithm [124] is used to find maximum likelihood estimates for the parameters $\pi_{\boldsymbol{\Phi},j}$, $\boldsymbol{\mu}_{\boldsymbol{\Phi},j}$, and $\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$, while the number of GMM components N_{GMM} is selected using the Bayesian information criterion (BIC) [125].

Due to (17), the prior GMM for $\boldsymbol{\Phi}$ (18) equivalently represents a prior GMM for $\boldsymbol{X}$,

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \approx \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X},j} \times \mathcal{N}\left(\boldsymbol{x};\, \boldsymbol{\mu}_{\boldsymbol{X},j},\, \boldsymbol{\Sigma}_{\boldsymbol{X},j}\right) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{19}$$

where

$$\pi_{\boldsymbol{X},j} = \pi_{\boldsymbol{\Phi},j} \tag{20}$$

$$\boldsymbol{\mu}_{\boldsymbol{X},j} = \bar{\boldsymbol{x}} + \boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j} \tag{21}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X},j} = \boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\boldsymbol{\mathcal{X}}^{\mathrm{T}} \tag{22}$$

are the jth component weight of the prior state variable GMM, jth component mean vector, and jth component covariance matrix, respectively. Further, if the Gaussian observation likelihood function,

$$\mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x}) = \mathcal{N}\left(\boldsymbol{y};\, \mathbf{H}\boldsymbol{x},\, \mathbf{R}\right), \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}} \tag{23}$$

as first defined in (4) is used, the Bayesian update of the GMM prior (19) is another GMM by conjugacy [118]; the posterior state variable distribution is thus,

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) = \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}|\boldsymbol{Y},j} \times \mathcal{N}\left(\boldsymbol{X};\, \boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j},\, \boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j}\right) \qquad \forall \boldsymbol{x} \in \mathbb{R}^{N_{\boldsymbol{X}}}\,, \tag{24}$$

with parameters

$$\pi_{\boldsymbol{X}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{X},j} \times \mathcal{N}\left(\boldsymbol{y};\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j},\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)}{\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X},k} \times \mathcal{N}\left(\boldsymbol{y};\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},k},\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},k}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)}\,,$$

$$\boldsymbol{\mu}_{\boldsymbol{X}|\boldsymbol{Y},j} = \boldsymbol{\mu}_{\boldsymbol{X},j} + \mathbf{K}_j\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X},j}\right)\,,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X}|\boldsymbol{Y},j} = \left(\mathbf{I} - \mathbf{K}_j\mathbf{H}\right)\boldsymbol{\Sigma}_{\boldsymbol{X},j} \qquad \forall j \in \{1,\ldots,N_{\mathrm{GMM}}\}$$

and gain matrices defined by

$$\mathbf{K}_j = \boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X},j}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \qquad \forall j \in \{1,\ldots,N_{\mathrm{GMM}}\}\,.$$

Though analytically accessible, the posterior GMM state variable distribution (24) cannot be directly computed nor stored for systems (12) of high dimension. A key advantage of the GMM-DO filter is that the update of (19) is equivalently obtained from the following update of (18):

$$p_{\boldsymbol{\Phi}|\boldsymbol{Y}}(\boldsymbol{\phi}|\boldsymbol{y}) = \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} \times \mathcal{N}\left(\phi;\, \boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j},\, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j}\right) \quad \forall \boldsymbol{\phi} \in \mathbb{R}^{N_{\mathrm{DO}}}\,, \tag{25}$$

where

$$\pi_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \frac{\pi_{\boldsymbol{\Phi},j} \times \mathcal{N}\left(\tilde{\boldsymbol{y}};\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j},\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\right)}{\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi},k} \times \mathcal{N}\left(\tilde{\boldsymbol{y}};\, \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},k},\, \tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi},k}\tilde{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}\right)}\,,$$

$$\boldsymbol{\mu}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} - \sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k}\,,$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\boldsymbol{Y},j} = \left(\mathbf{I} - \tilde{\mathbf{K}}_j\tilde{\mathbf{H}}\right)\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}\,, \qquad \forall j \in \{1,\ldots,N_{\mathrm{GMM}}\}$$

and the transformed observation vector realization, transformed observation matrix, transformed gain matrices, and intermediate component mean vectors are defined by,

$$\begin{aligned}
\tilde{\boldsymbol{y}} &= \boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}\,,\\
\tilde{\mathbf{H}} &= \mathbf{H}\boldsymbol{\mathcal{X}}\,,\\
\tilde{\mathbf{K}}_j &= \boldsymbol{\mathcal{X}}^{\mathrm{T}}\mathbf{K}_j\,,\\
\boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},j} &= \boldsymbol{\mu}_{\boldsymbol{\Phi},j} + \tilde{\mathbf{K}}_j\left(\tilde{\boldsymbol{y}} - \tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi},j}\right) \qquad \forall j \in \{1,\ldots,N_{\mathrm{GMM}}\}\,.
\end{aligned} \tag{26}$$

This posterior GMM coefficient distribution (25) is equivalent to (24) through (17) if the state vector mean is also updated according to,

$$\bar{\boldsymbol{x}}(t^+) = \bar{\boldsymbol{x}}(t^-) + \boldsymbol{\mathcal{X}}\sum_{k=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{\Phi}|\boldsymbol{Y},k} \times \boldsymbol{\mu}'_{\boldsymbol{\Phi}|\boldsymbol{Y},k}\,.$$

We first note that whereas the explicit calculation of (24) is infeasible, the calculation of (25) is untroublesome. Critically, no matrices of size larger than $N_{\boldsymbol{X}} \times N_{\mathrm{DO}} \ll N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}$ are manipulated in the update (25), rendering the analytical Bayesian update of the DO coefficient distribution computationally tractable for high-dimensional systems. Finally, new Monte Carlo samples are drawn from the posterior GMM coefficient distribution (25), which are dynamically evolved with the DO evolution equations until new observations are made and the efficient GMM-DO filter is applied again.

2.3.3. Reduced-dimension state augmentation

Parameter uncertainty in stochastic dynamical models can be directly coupled to state uncertainty using state augmentation [95]. The conceptual premise of state augmentation is simple: treat all uncertain parameters as time-invariant uncertain state variables. Concatenating vectors of state variables and parameters, using the notation in Table 3, then leads to the augmented state vector $\boldsymbol{X}_{\boldsymbol{\Theta}}(t) = \begin{bmatrix} \boldsymbol{X}(t) \\ \boldsymbol{\Theta} \end{bmatrix} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}$, whose realizations are given by $\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega) = \begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix}$. The augmented stochastic dynamical model is then,

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega)}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix} = \begin{bmatrix} \mathcal{M}\left[\boldsymbol{x}(t;\omega);\omega\right] \\ 0 \end{bmatrix} \\ &= \mathcal{M}_{\boldsymbol{\Theta}}\left[\boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega);\omega\right] . \end{aligned} \tag{27}$$

Reduced-dimension state augmentation utilizes an expansion as (11) for the augmented state vector,

$$\begin{aligned} \boldsymbol{x}_{\boldsymbol{\Theta}}(t;\omega) = \begin{bmatrix} \boldsymbol{x}(t;\omega) \\ \boldsymbol{\theta}(\omega) \end{bmatrix} &\approx \begin{bmatrix} \bar{\boldsymbol{x}}(t) \\ \bar{\boldsymbol{\theta}} \end{bmatrix} + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega) \begin{bmatrix} \tilde{\boldsymbol{x}}_i(t) \\ \tilde{\boldsymbol{\theta}}_i(t) \end{bmatrix} \\ &= \bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(t) + \sum_{i=1}^{N_{\mathrm{DO}}} \phi_i(t;\omega)\tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}(t) , \end{aligned} \tag{28}$$

where $\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}$ is the mean of the augmented state vector, $\tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}(t) \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}}$ the ith augmented DO vector, $\bar{\boldsymbol{\theta}} \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$ the mean of the parameter vector, and $\tilde{\boldsymbol{\theta}}_i(t) \in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$ the parameter elements of the ith augmented DO vector. Given (27), DO equations as in Section 3.2 (i.e. (14)–(16)) can be used to evolve the augmented state vector mean, augmented DO vectors, and DO coefficients defined in (28).

State augmentation also enables joint Bayesian inference of state variables and parameters using the GMM-DO filter. The observation likelihood $\mathcal{L}(\boldsymbol{y} \mid \bullet)$ is augmented to an equivalent likelihood function

$$\mathcal{L}_{\boldsymbol{\Theta}}(\boldsymbol{y} \mid \boldsymbol{x}_{\boldsymbol{\Theta}}) = \mathcal{L}\left(\boldsymbol{y} \mid \begin{bmatrix} \mathbf{I}^{N_{\boldsymbol{X}}} & \mathbf{0} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\Theta}}\right) \quad \forall \boldsymbol{x}_{\boldsymbol{\Theta}} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}} \tag{29}$$

that simply ignores the parameter elements of the augmented state vector. If the original state likelihood function is a Gaussian function, as assumed by the GMM-DO filter in (23), the augmented likelihood becomes

$$\begin{aligned} \mathcal{L}_{\boldsymbol{\Theta}}(\boldsymbol{y} \mid \boldsymbol{x}_{\boldsymbol{\Theta}}) &= \mathcal{N}\left(\boldsymbol{y}\,;\, \mathbf{H}\begin{bmatrix} \mathbf{I}^{N_{\boldsymbol{X}}} & \mathbf{0} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\Theta}}\,,\, \mathbf{R}\right) = \mathcal{N}\left(\boldsymbol{y}\,;\, \begin{bmatrix} \mathbf{H} & \mathbf{0} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\Theta}}\,,\, \mathbf{R}\right) \\ &= \mathcal{N}\left(\boldsymbol{y}\,;\, \mathbf{H}_{\boldsymbol{\Theta}}\boldsymbol{x}_{\boldsymbol{\Theta}}\,,\, \mathbf{R}\right) \quad \forall \boldsymbol{x}_{\boldsymbol{\Theta}} \in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}}} \end{aligned} \tag{30}$$

and the GMM-DO filter can be applied as usual using the augmented observation matrix $\mathbf{H}_{\boldsymbol{\Theta}} = \begin{bmatrix} \mathbf{H} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N_{\boldsymbol{Y}} \times (N_{\boldsymbol{X}}+N_{\boldsymbol{\Theta}})}$ in place of the original observation matrix $\mathbf{H}$. As long as some of the augmented DO vectors in (28) feature both nonzero state variable and parameter elements – i.e. as long as there is some finite correlation between state variables and parameters – the probability distribution for parameters will be jointly updated with the probability distribution for state variables when the GMM-DO filter updates the distribution for the shared DO coefficients. After this GMM-DO update, the integration of (27) and (28) will utilize the updated parameter mean and evolve the joint state-parameter probability distribution in accord with the dynamics until the next observation.

In summary, reduced-dimension state augmentation extends both the DO evolution equations and the GMM-DO filter to enable joint uncertainty evolution and inference of state variables and parameters. In what follows, the augmented state vector $\boldsymbol{X}_{\boldsymbol{\Theta}}$ (and realization $\boldsymbol{x}_{\boldsymbol{\Theta}}$) will be denoted as simply $\boldsymbol{X}$ (and $\boldsymbol{x}$), but $\boldsymbol{X}$ in general encapsulates both state variables and parameters coupled through state augmentation.

2.3.4. GMM-DO marginal likelihood computation

We now show that for the GMM state variable distributions, Gaussian observation models, and likelihood function of the GMM-DO filter, an analytical solution for the marginal likelihood integral (9) is available. Specifically, using (23) and (19) for $\mathcal{L}(\boldsymbol{y} \mid \bullet)$ and $p_{\boldsymbol{X}\mid\mathcal{M}}(\bullet \mid \mathcal{M}_n)$ respectively, we obtain,

$$\begin{aligned} p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y} \mid \mathcal{M}_n) &= \int \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{x})\; p_{\boldsymbol{X}\mid\mathcal{M}}(\boldsymbol{x} \mid \mathcal{M}_n)\; \mathrm{d}\boldsymbol{x} \\ &= \int \mathcal{N}\left(\boldsymbol{y}\,;\, \mathbf{H}\boldsymbol{x}\,,\, \mathbf{R}\right) \\ &\quad \times \left(\sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \times \mathcal{N}\left(\boldsymbol{x}\,;\, \boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\right)\right) \mathrm{d}\boldsymbol{x} . \end{aligned} \tag{31}$$

Here, since we are again considering model uncertainty, we use $\mathcal{M}_n$ in place of $\mathcal{M}$ to represent a realization of the uncertain model. Since the integral in (31) is taken over all values of $\boldsymbol{x}$ (state variables and parameters), a linear transformation of the integration variable can be performed without changing the value of the integral. Then, using the linear transformation properties of Gaussian distributions, (31) is rewritten as

$$\begin{aligned} p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y} \mid \mathcal{M}_n) &= \int \mathcal{N}\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\,;\, 0\,,\, \mathbf{R}\right) \left(\sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \right. \\ &\quad \left. \times \mathcal{N}\left(\mathbf{H}\boldsymbol{x}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\right)\right) \mathrm{d}\mathbf{H}\boldsymbol{x} \\ &= \int \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \times \mathcal{N}\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\,;\, 0\,,\, \mathbf{R}\right) \\ &\quad \times \mathcal{N}\left(\mathbf{H}\boldsymbol{x}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\right) \mathrm{d}\mathbf{H}\boldsymbol{x} . \end{aligned} \tag{32}$$

Interchanging integration and summation, then factoring,

$$\begin{aligned} p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y} \mid \mathcal{M}_n) &= \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \times \int \mathcal{N}\left(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\,;\, 0\,,\, \mathbf{R}\right) \\ &\quad \times \mathcal{N}\left(\mathbf{H}\boldsymbol{x}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\right) \mathrm{d}\mathbf{H}\boldsymbol{x} \\ &= \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \times \left[\, \mathcal{N}\left(\bullet\,;\, 0\,,\, \mathbf{R}\right) \right. \\ &\quad \left. * \mathcal{N}\left(\bullet\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}}\right) \right](\boldsymbol{y}) , \end{aligned} \tag{33}$$

where $*$ represents the convolution operator, defined as: $\left[f(\bullet) * g(\bullet)\right](t) = \int f(t-\tau)g(\tau)\,\mathrm{d}\tau$. The convolutions in (33) yield new Gaussian distributions whose means and variances are equal to the sums of the means and variances of the two component distributions [126]:

$$\begin{aligned} p_{\boldsymbol{Y}\mid\mathcal{M}}(\boldsymbol{y} \mid \mathcal{M}_n) &= \sum_{j=1}^{N_{\mathrm{GMM}}} \pi_{\boldsymbol{X}\mid\mathcal{M}_n,j} \\ &\quad \times \mathcal{N}\left(\boldsymbol{y}\,;\, \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\,,\, \mathbf{H}\boldsymbol{\Sigma}_{\boldsymbol{X}\mid\mathcal{M}_n,j}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right) . \end{aligned} \tag{34}$$

Analogous to (24), the analytically accessible (34) cannot be directly computed for systems of high dimension due to the prohibitive size of the state variable GMM component covariance matrices involved. Fortunately, using the GMM-DO approach, computing (34) is possible, again because of (17). Substituting the

prior GMM-DO fit in the DO subspace, (20)–(22), into (34) yields,

$$p_{\boldsymbol{Y}|\mathcal{M}}(\boldsymbol{y}\,|\,\mathcal{M}_n) = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\mathcal{M}_n,j} \times \mathcal{N}\left(\boldsymbol{y}\,;\,\mathbf{H}\left(\bar{\boldsymbol{x}}+\boldsymbol{\mathcal{X}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\right)\,,\,\mathbf{H}\left(\boldsymbol{\mathcal{X}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\boldsymbol{\mathcal{X}}^{\mathrm{T}}\right)\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right) = \sum_{j=1}^{N_{\text{GMM}}} \pi_{\boldsymbol{\Phi}|\mathcal{M}_n,j} \times \mathcal{N}\left(\tilde{\boldsymbol{y}}\,;\,\tilde{\mathbf{H}}\boldsymbol{\mu}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\,,\,\tilde{\mathbf{H}}\boldsymbol{\Sigma}_{\boldsymbol{\Phi}|\mathcal{M}_n,j}\tilde{\mathbf{H}}^{\mathrm{T}}+\mathbf{R}\right)\,, \tag{35}$$

where the definitions in (26) have been used. The explicit evaluation of the Gaussian functions in (35) is computationally expedient. None of the matrices involved exceed $N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}$ in size and $N_{\boldsymbol{Y}}$ is typically orders of magnitude smaller than $N_{\boldsymbol{X}}$, especially for systems featuring infinite-dimensional state fields (or their large-dimensional discretized versions) [127]. Furthermore, all the quantities in (35) are computed during the Bayesian state variable inference of the GMM-DO filter, thus necessitating no new computations. Finally, (35) can be used to calculate marginal likelihoods for all candidate models, which can subsequently be used to perform Bayesian update of the model distribution as represented by (7).

2.4. Summary

Our principled Bayesian learning of stochastic dynamical models consists the following four-steps:

1. *Model formulation.* Formulate a set of candidate models: for each, specify stochastic dynamical equations, spatial geometry, boundary conditions, and initial conditions, as represented in (2).
2. *Uncertainty initialization.*
   (a) *State variable and parameter uncertainty initialization.* For each candidate model, specify an initial joint probability distribution for state variables and parameters. Initialize the augmented state vector DO expansion (28) for each candidate (mean, modes and coefficients). Computational procedures to do so when initial distributions are Gaussians or GMMs are discussed in [59,128] and in [129–131] for multivariate realistic ocean applications.
   (b) *Model uncertainty initialization.* Define an initial discrete probability distribution over the set of candidate models, i.e. assign the probability that each candidate model is the true model.
   (c) *Initial Bayesian uncertainty quantification.* As needed, compute the initial probability distribution for state variables and parameters that accounts for model uncertainty using (5).
3. *Uncertainty evolution.*
   (a) *Model-conditional uncertainty evolution.* Use the DO Eqs. (14)–(16) to evolve the augmented state vector means, modes, and coefficients for each candidate model.
   (b) *Prior Bayesian uncertainty quantification.* As needed, compute the prior probability distribution for state variables and parameters that accounts for model uncertainty using (5).
4. *Learning.*
   (a) *State variable and parameter inference.* When an observation is made, use the GMM-DO filter to perform a Bayesian update of the DO coefficient distribution for each candidate model.
   (b) *Marginal likelihood calculation.* Compute marginal likelihoods for all candidate models using (35).
   (c) *Model learning.* Use (7) and the marginal likelihoods to perform a Bayesian update of the model distribution.
   (d) *Posterior Bayesian uncertainty quantification.* As needed, compute the posterior probability distribution for state variables and parameters that accounts for model uncertainty using (5).

A graphical representation of this four-step procedure is provided in Fig. 1. N_{DO} and N_{GMM} are both time- and model-dependent but for simpler notation, this is not explicitly represented in the Figure. Similarly, the number of candidate models $N_{\mathcal{M}}$ is in general a function of time, but this too is omitted.

2.5. Remarks

Independence between candidate models. All operations in Steps 2a, 3a, 4a, and 4b (i.e. the left half of Fig. 1) are performed independently for each candidate model and can thus be completed in parallel. Another important consequence of this independence is that idiosyncratic variations of the DO evolution equations and the GMM-DO filter can be applied to any of the candidate models without affecting the others. For example, each candidate model can employ a different stochastic subspace dimension for the expansion (28), and some can dynamically evolve this subspace dimension [120,121].

Independence from model distribution. The operations in Steps 2a, 3a, 4a, and 4b are also independent of the model distribution (i.e. no arrows point from the right half of Fig. 1 to the left). In fact, as long as the marginal likelihoods of Step 4b are stored, the Bayesian update of the model distribution in Step 4c need not be performed unless the state variable distribution with model uncertainty accounted for is needed (i.e. Steps 2c, 3b, and 4d are needed). New candidate models can thus be added to the set of viable candidates, even after Steps 2a, 3a, 4a, and 4b have been carried out for existing candidates. One simply needs to respecify the initial model distribution to include the new candidates, then perform the independent steps afresh for the new candidates only. The marginal likelihoods generated for the new candidates can be combined with those already stored to perform Bayesian updates of the new model distribution.

Note that model learning can occur either immediately after each observation or only after all observations are made. Computational resources and the utility of real-time state variable and parameter distributions that account for model uncertainty can guide the choice. If real-time distributions are crucial, model learning should be performed at all observation times in order to carry out Steps 3b and 4d. If it is not, model learning may be left for post-processing. The present methodology allows for both courses of action.

Recursive model learning. Our model learning procedure can be extended to a form of 'recursive model learning' that is similar to parameter inference. In this case, when the model distribution is updated at each observation time, the model itself can be treated as a parameter: we would then sample model uncertainty as we sample parameter uncertainty. Consequently, the set of candidate models could be updated with time, e.g. using the

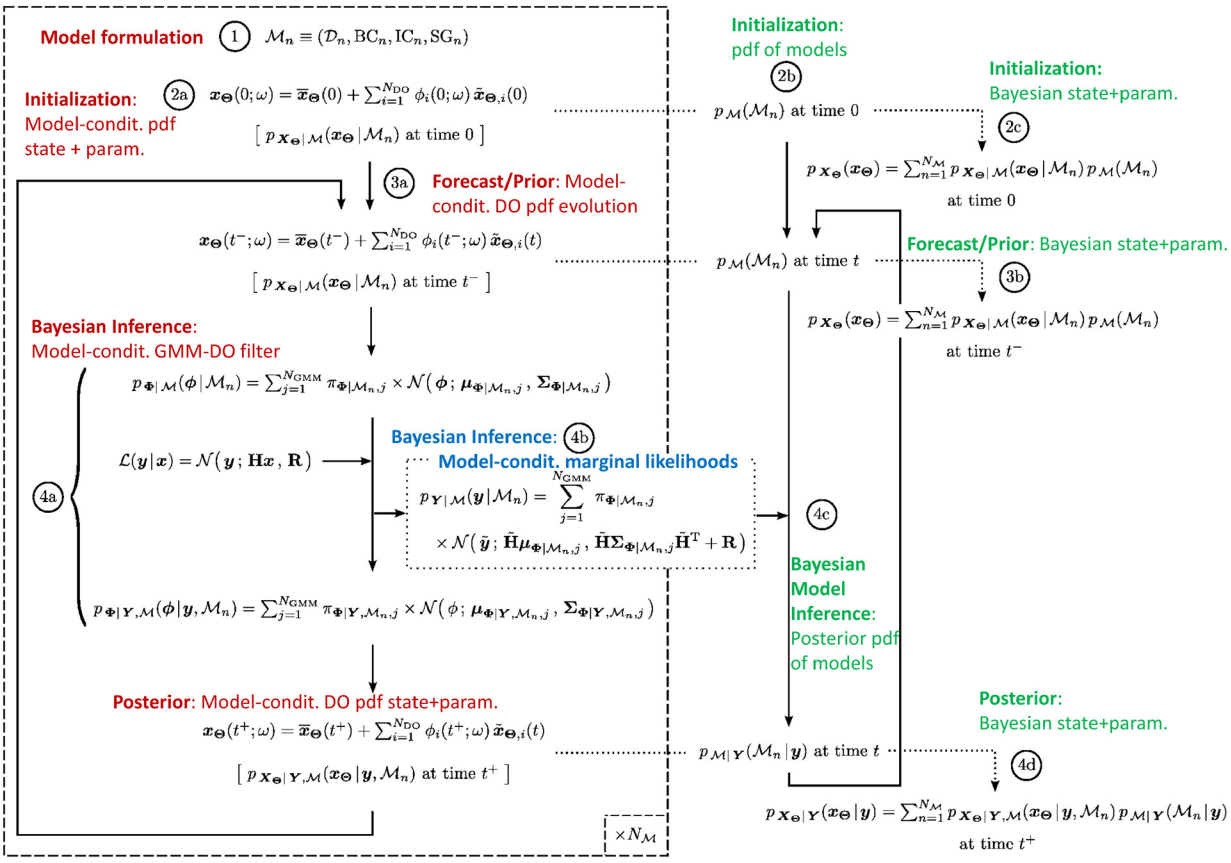

Fig. 1. Four-step procedure for Bayesian learning of stochastic dynamical models.

GMM-DO filter, and these updates would then influence model predictions at future times. With this extension, model learning and state inference would not be independent: updates to the set of candidate models would influence the dynamic evolution of the state variables, just as updates to parameter distributions influence state variables in the present learning.

In sum, the present methodology is a versatile rigorous approach to Bayesian learning of stochastic dynamical models. Applications to two nonlinear systems are presented next.

3. Flow past an obstacle

The flow past a cylinder or idealized island is a classic fluid dynamical system. Its simplest rendition is a two-dimensional flow around an impermeable fixed circular obstacle, with an incompressible and Newtonian fluid, and a uniform and steady flow at an infinite distance away from the obstacle. This system has been extensively studied due to both its simple formulation and the startling intricacy of the fluid flow patterns it can evoke [132–137]. The key parameter is the non-dimensional Reynolds number [138],

$$\mathrm{Re} = \frac{V_\infty L}{\nu}\,, \tag{36}$$

where V_∞ is the fluid velocity at infinity from the obstacle, L the projected width of the obstacle, and ν the dynamic viscosity of the fluid. For Re below approximately 40, the fluid flow downstream of the obstacle is symmetric, with a pair of stable recirculation zones appearing behind the obstacle for Re greater than 5. These recirculation zones increase in size with increasing Re. At Re greater than 40, the recirculation zones become unstable and exhibit periodic, asymmetric vortex shedding. This results in dynamic flow patterns known as von Kármán vortex streets downstream of the obstacle. Systems featuring Re greater than 200 exhibit more complex aperiodic patterns while Re greater than 10^5 result in turbulent flow [139].

Countless variations of this system have been investigated, involving non-Newtonian fluids [140], rotating obstacles [141], oscillating obstacles [142], noncircular obstacles [143], multiple obstacles [144], and three-dimensional fluids and obstacles [145]. The system and its variants have also been used for multiple purposes, one of which being practical models of real systems found in both natural and engineered environments [146].

3.1. Stochastic flow past an obstacle system

The first stochastic dynamical system considered for our Bayesian model learning (Section 2) is a stochastic variation of the flow past a cylinder. The two-dimensional (2D) spatial domain is shown in Fig. 2, with fluid flow proceeding left to right in the positive r_1 direction. Uncertainty in the fluid velocity field is introduced in three ways: (1) initial velocity field; (2) Reynolds number; and (3) shape of the obstacle.

For Re between 40 and 200, the vortex shedding frequency and morphology are predictable. The phase of vortex shedding however is highly unpredictable, as shedding is triggered chaotically by asymmetric perturbations of the otherwise symmetric fluid domain. Even mild uncertainty in the initial conditions can thus lead to substantial uncertainty in the shedding phase, leading to substantial uncertainty in the velocity field. With uncertainty in Re, both the frequency and morphology of vortex shedding are uncertain (higher Re correspond to greater shedding frequencies and vortices of higher energy, see [88,147]).

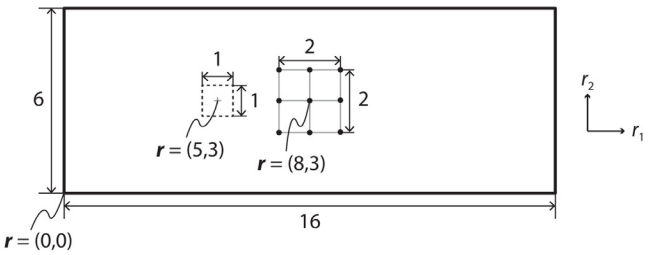

Fig. 2. 2D spatial domain of the stochastic flow past an obstacle system. All lengths and coordinates are non-dimensional. The bounding area for the uncertain obstacle is indicated by the dashed square. The observation array consists of the nine locations at the center of the domain.

Finally, uncertainty in the shape of the obstacle can also lead to significant uncertainty in the velocity field. As exhibited in Fig. 3, different obstacle shapes result in different downstream fluid flow patterns. Even when projected widths are identical, different shapes can lead to different vortex shedding frequencies and morphologies for the same Re. For our Bayesian learning, five shapes are hypothesized, with spatial dimensions as given in Fig. 4. The details of the flow patterns for these five shapes differ (Fig. 3). Specifically, the circle and square exhibit both similar near-field and far-field vortex shedding patterns, but with slight variations in vortex intensities (most notably in the near-field). The other three shapes also exhibit similar near-field vortex shedding patterns, but the far-field pattern of the downstream-pointing triangle differs markedly from those of the upstream-pointing triangle and diamond. These differences can be exploited to learn the correct obstacle shape.

3.2. Description of the learning experiments

3.2.1. Model formulation

The five shapes lead to five possible configurations for the spatial domain in Fig. 2. Each of these domains represents a distinct candidate model for the system, with each candidate assumed to be equally likely a priori (i.e. a uniform initial model distribution is used). We now describe their components (2).

$\mathcal{D}_n$ — The same set of stochastic dynamical equations is used for each of the candidate models: the 2D Navier–Stokes PDEs for an incompressible, Newtonian fluid in non-dimensional form

$$\frac{\partial v_1}{\partial t} = -\frac{\partial (v_1^2)}{\partial r_1} - \frac{\partial (v_1 v_2)}{\partial r_2} + \frac{1}{\mathrm{Re}(\omega)}\left(\frac{\partial^2 v_1}{\partial r_1^2} + \frac{\partial^2 v_1}{\partial r_2^2}\right) - \frac{\partial p}{\partial r_1}\,, \tag{37}$$

$$\frac{\partial v_2}{\partial t} = -\frac{\partial (v_1 v_2)}{\partial r_1} - \frac{\partial (v_2^2)}{\partial r_2} + \frac{1}{\mathrm{Re}(\omega)}\left(\frac{\partial^2 v_2}{\partial r_1^2} + \frac{\partial^2 v_2}{\partial r_2^2}\right) - \frac{\partial p}{\partial r_2}\,, \tag{38}$$

$$\frac{\partial v_1}{\partial r_1} + \frac{\partial v_2}{\partial r_2} = 0\,, \tag{39}$$

where $v_1 = v_1(\boldsymbol{r}, t; \omega)$ and $v_2 = v_2(\boldsymbol{r}, t; \omega)$ denote the r_1 and r_2 components of the non-dimensionalized velocity field $\boldsymbol{v}(\boldsymbol{r}, t; \omega)$, respectively [148]. The Reynolds number for the system $\mathrm{Re}(\omega)$ is an uncertain parameter to be inferred. The non-dimensionalized pressure field, $p = p(\boldsymbol{r}, t; \omega)$, is solved for implicitly from the velocity field using the continuity Eq. (39).

SG_n — The spatial geometries for the five candidate models have their own obstacle shape (Fig. 4). The projected widths of all the obstacles are identical and are all 1 non-dimensional length unit. For numerical implementation, the spatial domain for each of the candidate models is discretized using a regular rectangular grid of 320 and 120 elements along the r_1 and r_2 directions respectively ($\Delta r_1 = \Delta r_2 = \frac{1}{20}$). This spatial discretization results in approximately 320×120 r_1-velocity and 320×120 r_2-velocity state variables for each of the candidate models, with slight variations due to the different surface areas of the five possible obstacle shapes. The augmented state vector for any of the candidate models is then

$$\boldsymbol{x}_{\boldsymbol{\Theta}}(t; \omega) = \begin{bmatrix} \boldsymbol{v_1}(t; \omega) \\ \boldsymbol{v_2}(t; \omega) \\ \mathrm{Re}(\omega) \end{bmatrix}\,, \tag{40}$$

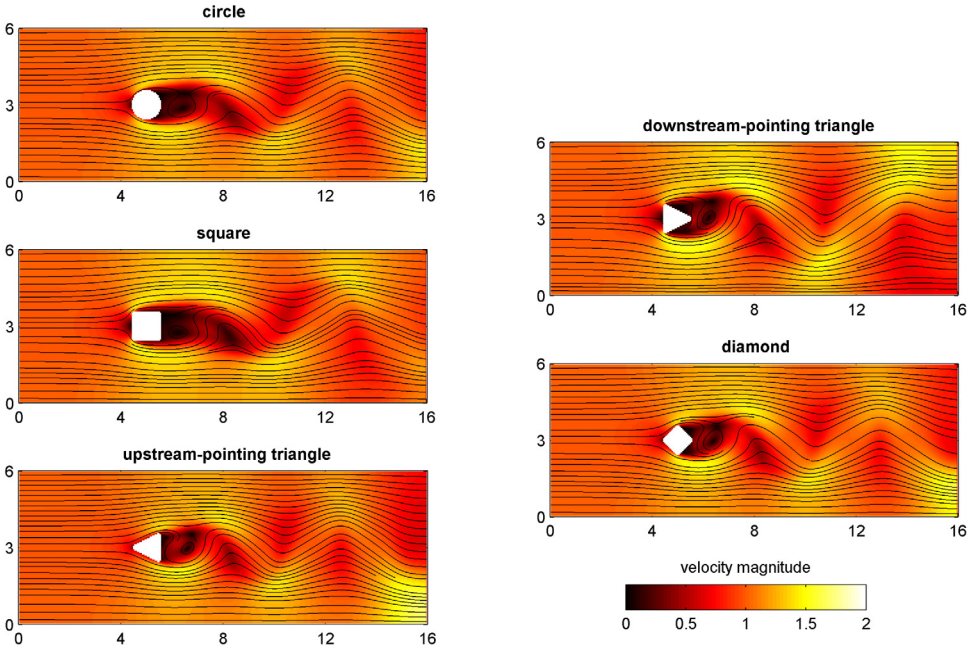

Fig. 3. Fluid flow patterns for realizations of the stochastic flow past an obstacle system featuring different obstacle shapes, all at the same non-dimensional time and with a Re of 75. The projected width of all the obstacles is identical. Phase shifts indicate different vortex shedding frequencies.

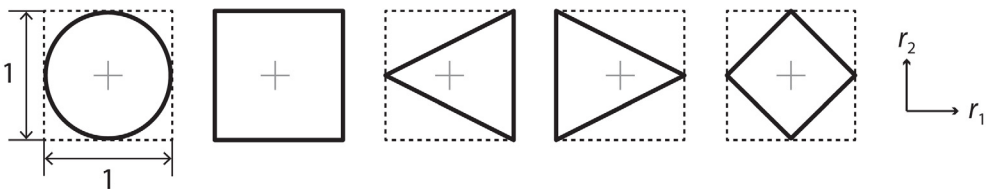

Fig. 4. Five possible obstacle shapes: circle, square, upstream-pointing triangle, downstream-pointing triangle, and diamond. All lengths are non-dimensional. The bounding area for each shape is the area indicated in Fig. 2.

where $\boldsymbol{v_1}(t;\omega)$ and $\boldsymbol{v_2}(t;\omega)$ are the vectors of r_1-velocity and r_2-velocity state variables, respectively. The dimension of each augmented state vector, $N_{\boldsymbol{X}} + N_{\boldsymbol{\Theta}}$, is thus of order 10^5.

BC_n — Stochastic Dirichlet conditions are used for the left domain limits for all five candidate models

$$\boldsymbol{v} = (V_\infty(\omega), 0) = \left(\frac{\mathrm{Re}(\omega)}{\mathrm{Re}'}, 0\right) \text{ for } r_1 = 0\,, \tag{41}$$

where $V_\infty(\omega)$ denotes the stochastic non-dimensionalized inlet fluid velocity and $\mathrm{Re}' = \frac{V'L'}{\nu}$ is a constant comprised of the non-dimensionalizing velocity and length constants V' and L' as well as the dynamic viscosity of the fluid ν. V', L', and ν are chosen to be 1, 1, and $\frac{1}{80}$, respectively. For all candidate models, deterministic Neumann conditions are used for the top, bottom, and right domain limits,

$$\frac{\partial \boldsymbol{v}}{\partial r_2} = 0 \;\text{ at } r_2 = 0 \text{ and } r_2 = 6\,,$$
$$\frac{\partial \boldsymbol{v}}{\partial r_1} = 0 \;\text{ at } r_1 = 16\,,$$

Deterministic Dirichlet conditions impose no-slip conditions on the surfaces of the obstacles.

IC_n — For each of the five candidate models, an initial uniform distribution from 50 to 100 is assumed for $\mathrm{Re}(\omega)$. The mean of this distribution—$\mathrm{Re} = 75$—corresponds to a mean non-dimensionalized inlet fluid velocity of $\frac{75}{80}$, using (41). Initial means of the fluid velocity for each of the candidate model spatial domains are found by numerically solving the continuity Eq. (39) with this mean inlet velocity and the other boundary conditions given above. Figs. 5 and 6 illustrate the stochastic system at the initial non-dimensional time ($t = 0$). For each candidate model, a covariance matrix is constructed for the velocity fields that respects symmetry about the $r_2 = 3$ centerline using the boundary-mollified spatial covariance method [59]. Discrete Karhunen–Loève transforms are performed on these covariance matrices to initialize eight pairs of augmented DO vectors and coefficients for each candidate model. The symmetry properties of this procedure lead to unbiased initialization of uncertainty in vortex shedding phase (i.e. no information regarding the phase of vortex shedding is assumed at the start), as shown in Fig. 6. In the wakes behind the obstacles, the velocity standard deviations reach 25 to 50 percent of the total velocity magnitude, confirming the large uncertainty in the initial velocity field. For each candidate model, a ninth augmented DO vector is used that contains only a single nonzero entry for the uncertain $\mathrm{Re}(\omega)$, for a total DO subspace dimension $N_{\mathrm{DO}} = 9$.

3.2.2. True solution generation

We generate a true solution for the learning experiments using a circular obstacle and a Reynolds number of 80. A deterministic initial velocity field satisfying the continuity Eq. (39) is first constructed, with arbitrary asymmetries introduced to induce vortex shedding. The evolution of the velocity and pressure fields is computed by numerically solving the Navier–Stokes Eqs. (37)–(39) using our second-order finite-volume scheme [149] with a regular rectangular grid of 320 and 120 elements along the r_1 and r_2 directions, respectively, and a non-dimensional time-step of $\Delta t = \frac{1}{120}$ up to a final non-dimensional time $T = 50$. This deterministic solution is a realization of the stochastic flow past an obstacle system, representing the true obstacle shape, true Reynolds number, and true dynamical velocity field that are to be jointly learned.

Table 1
Numerical properties for the stochastic flow past an obstacle system.

Property	$N_{\mathcal{M}}$	$N_{\boldsymbol{X}}$	$N_{\boldsymbol{\Theta}}$	$N_{\boldsymbol{Y}}$	N_{DO}	N_{MC}	Δr_1	Δr_2	Δt	T
Value	5	∼75,000	1	18	9	10^4	1/20	1/20	1/120	50

3.2.3. Observations and learning parameters

Every 1 non-dimensional time unit (i.e. every 120 time-steps), noisy observations of both r_1-velocity and r_2-velocity are made at nine locations (Fig. 2), for a total of $N_{\boldsymbol{Y}} = 18$ scalar data at every data time. A single non-dimensional time unit for this system corresponds to a phase shift of about $\frac{5\pi}{13}$ when $\mathrm{Re} = 80$. Unbiased Gaussian noise is applied to the observations with a standard deviation equal to approximately 10% of the mean fluid velocity, which are then used by the GMM-DO filter to perform joint model-conditional inference of the velocity state variables and Reynolds number, for each of the candidate models. The observation matrix $\mathbf{H}$ in Eq. (23) selects the velocity state variables at the nine data locations and the covariance matrix $\mathbf{R}$ is diagonal of elements equal to the variance of the observation noise.

As in [118], the BIC and EM algorithm are used to select the optimal number of GMM components N_{GMM} at each data time. Typical BIC-optimized values for N_{GMM} were found to lie between 20 and 60 for the present experiments. Marginal likelihoods for the candidate models are calculated using (35) after every GMM-DO filtering step and the model distribution is updated using (7).

3.2.4. Numerical method

For each of the candidate models, the augmented state vector mean, DO vectors, and DO coefficients are governed by the DO Eqs. (14)–(16) for the Navier–Stokes Eqs. (37)–(39) with stochastic boundary conditions (41). They are numerically integrated using the finite-volume methodology developed in [123], with a non-dimensional time-step of $\Delta t = \frac{1}{120}$. A total of $N_{\mathrm{MC}} = 10^4$ Monte Carlo samples is used for the stochastic evolution of the DO coefficients. The numerical properties used are summarized in Table 1.

3.3. Learning results

Figs. 5 and 6 showed the system at $t = 0$, before observations are made. The fluid velocity field of the chosen true deterministic simulation (Fig. 5, top left) is in a time-periodic state. Regular asymmetric vortex shedding is occurring downstream of the obstacle, with a period of approximately 5.4 non-dimensional time

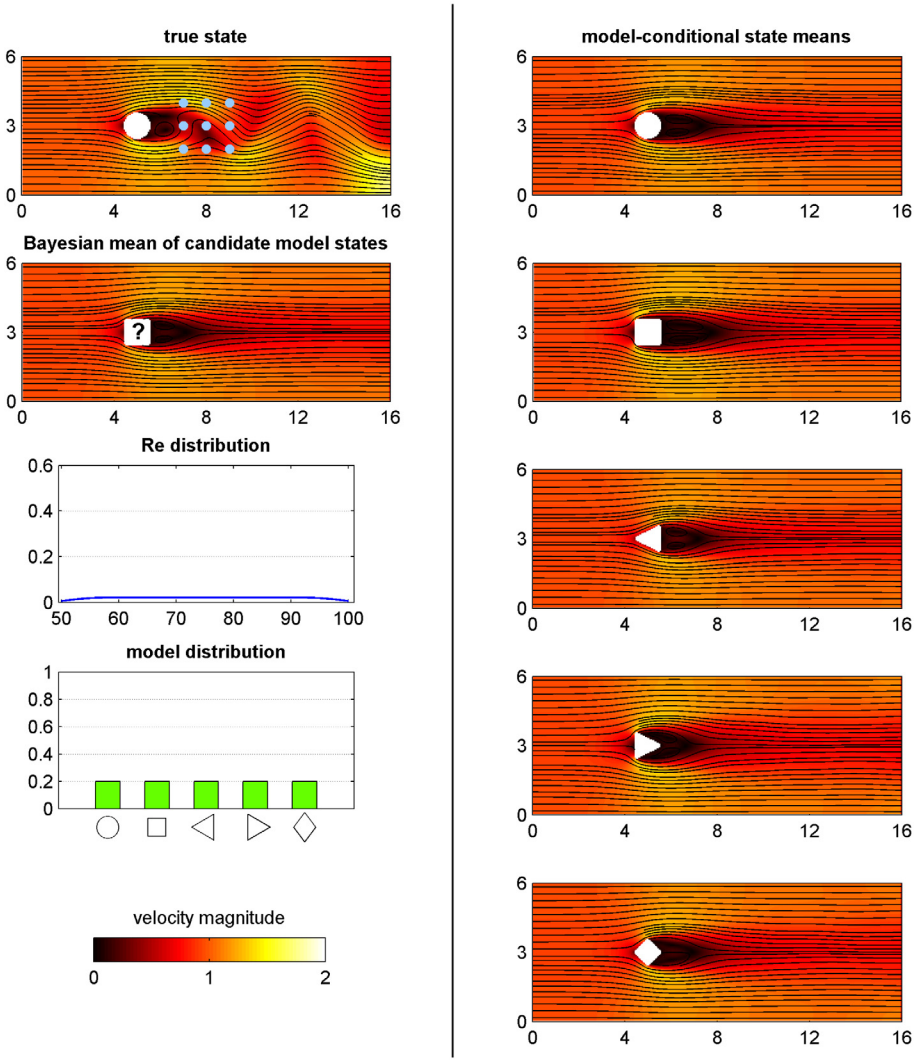

Fig. 5. The stochastic flow past an obstacle system at non-dimensional time $t = 0$. Left column: Observations are made from a deterministic simulated truth featuring the circular obstacle (top), with observation locations indicated on the true state field in light blue. The bounding area for the obstacle in the Bayesian mean of the candidate model states is labeled with a question mark since the obstacle shape is uncertain. The initial distributions for the continuous Re(ω) and for the discrete model obstacle shape are shown below, above the velocity magnitude colorbar. Right column: Model-conditional velocity field means, for each of the five candidate models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

units. Each of the five model-conditional velocity field means (i.e. $\bar{\boldsymbol{x}}$ in (11)) for the candidate models is symmetric about the $r_2 = 3$ centerline, indicating no bias towards any particular vortex shedding phase. The Bayesian mean of these five model-conditional velocity fields (i.e. the mean of $p_{\boldsymbol{X}}(\bullet)$ in (5) or equivalently the mean of the five $\bar{\boldsymbol{x}}$ of the candidate models) is consequently also symmetric. The model-conditional Re distributions are uniform, as is the model distribution, i.e. $p_{\mathcal{M}}(\bullet)$ in (5). The model-conditional velocity standard deviations are also symmetric about the $r_2 = 3$ centerline. Crucially, these initial model-conditional flow uncertainties are not small, the largest standard deviations reach the total flow variability.

Figs. 7 and 8 illustrate the system after 10 non-dimensional time units and 10 observation episodes. Each of the candidate models performs 10 Bayesian GMM-DO data assimilation (see Sections 2.3.2 and 2.3.3, details not illustrated here). As a result, all five of the model-conditional velocity field means are beginning to align in phase with the true velocity field (Fig. 7, right column). Consequently, the Bayesian mean of these conditional velocity field means is also beginning to align in phase with the true field (Fig. 7, left column, top two panels), though the energy of the vortices in this Bayesian mean is still lower than that of those in the true field. The Reynolds number distribution is beginning to shift towards the true Re of 80. Finally, the probability of one of the incorrect candidate models, the one featuring the downstream-pointing triangular obstacle, has fallen quickly. The Bayesian learning identifies that it is indeed the obstacle most dissimilar dynamically to the true circular one: it has a large frontal area and downstream-pointing edges.

In Fig. 8, the predicted model-conditional velocity standard deviations for each candidate model (left column) have decayed when compared to their initial values (see Fig. 6). Importantly, these predicted standard deviations agree with the differences between the model-conditional velocity field means and the true field (see Fig. 7). The five model-conditional Re distributions (right column), that can be obtained by marginalization (see Section 2.1.1), show that each model refines its initial uniform distribution but with a different mean Re estimate. This is due to the model biases (wrong obstacle shape) and remaining field errors (e.g. phases, frequency, and strength of eddy shedding). For example, the upstream-pointing triangle has a too low Re mean estimate while the downstream-pointing triangle has a too

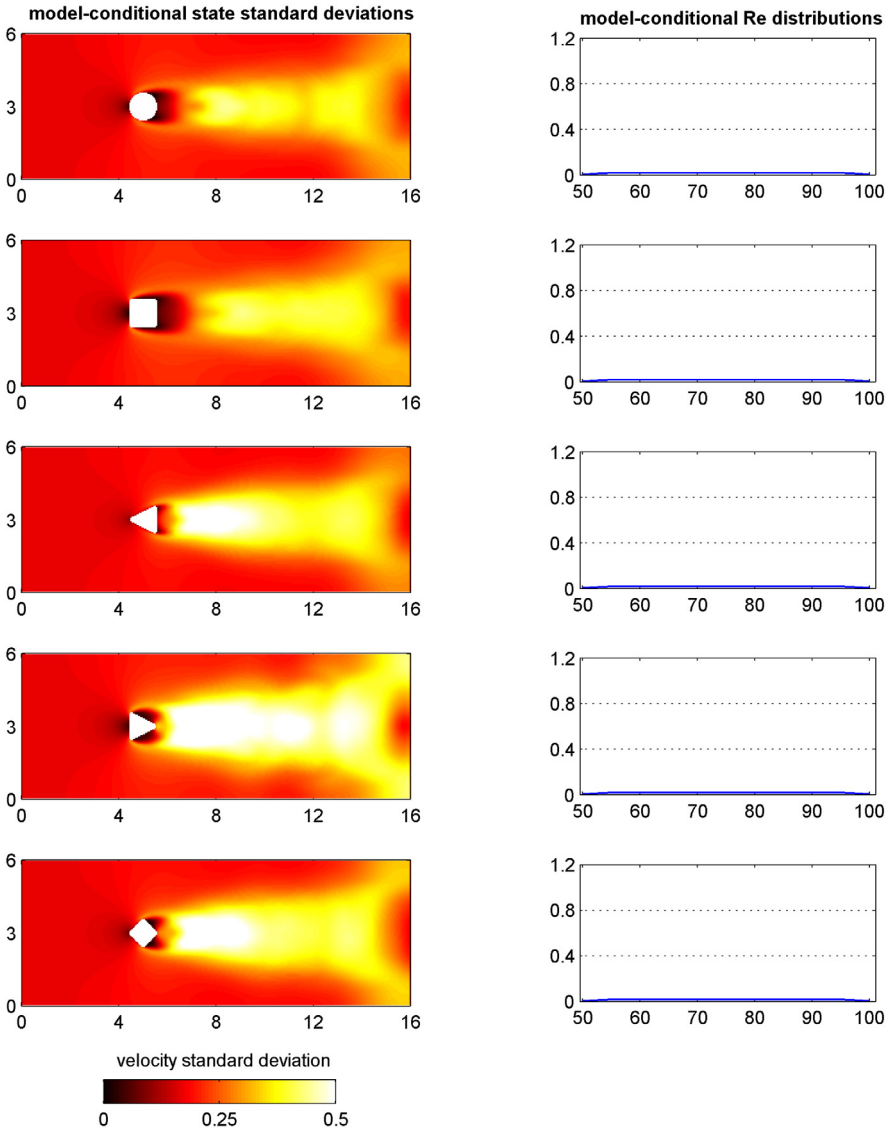

Fig. 6. As Fig. 5, but showing model-conditional uncertainties at non-dimensional time $t = 0$. Specifically, model-conditional velocity field standard deviations for each of the five candidate models are displayed on the left. Model-conditional Reynolds number distributions for each of the five candidate models are displayed on the right.

high Re mean estimate, in each case to compensate for the wrong shedding (e.g. Strouhal number).

Figs. 9 and 10 illustrate the system after 50 non-dimensional time units and 50 observation episodes and Bayesian GMM-DO assimilation. Phase alignments between the five conditional velocity field means and the true field have all improved significantly and hence the Bayesian velocity field mean is now largely consistent with the true velocity field (Fig. 9, left column, top two panels). The Reynolds number distribution has peaked around the true Re of 80, but small secondary peaks (e.g. around 74 and 86) are also visible, indicating again that some of the models are biased. Importantly, our Bayesian learning accurately captures such non-Gaussian distributions. Finally, the candidate model featuring the circular obstacle is now by far the most plausible model for the system, as indicated by the highly peaked model distribution.

In Fig. 10, the predicted model-conditional velocity field standard deviations for all five candidate models have decreased significantly, with the greatest decreases exhibited by the circular obstacle (the true model) and the square obstacle (the candidate model 'closest' to the circle, as seen in Fig. 3). Again, these predicted standard deviations agree with the actual errors of the model-conditional velocity field means (see Fig. 9). Since the five obstacle shapes produce different vortex shedding frequencies (Strouhal number) for the same Re however (see Fig. 3), the five model-conditional Re distributions have further concentrated around different values. For the candidate model with the circular obstacle, the conditional Re distribution has simply concentrated around 80, the true Re. For the other candidates though, the conditional distributions have concentrated around values that allow the candidates to better match the shedding frequency of the true field despite featuring incorrect obstacle shapes. The tendencies after 10 Bayesian assimilation have strengthen after 50 assimilation: the downstream-pointing triangle and square have a too high Re mean, with the square the closest to the truth, while the upstream-pointing triangle and diamond have too low

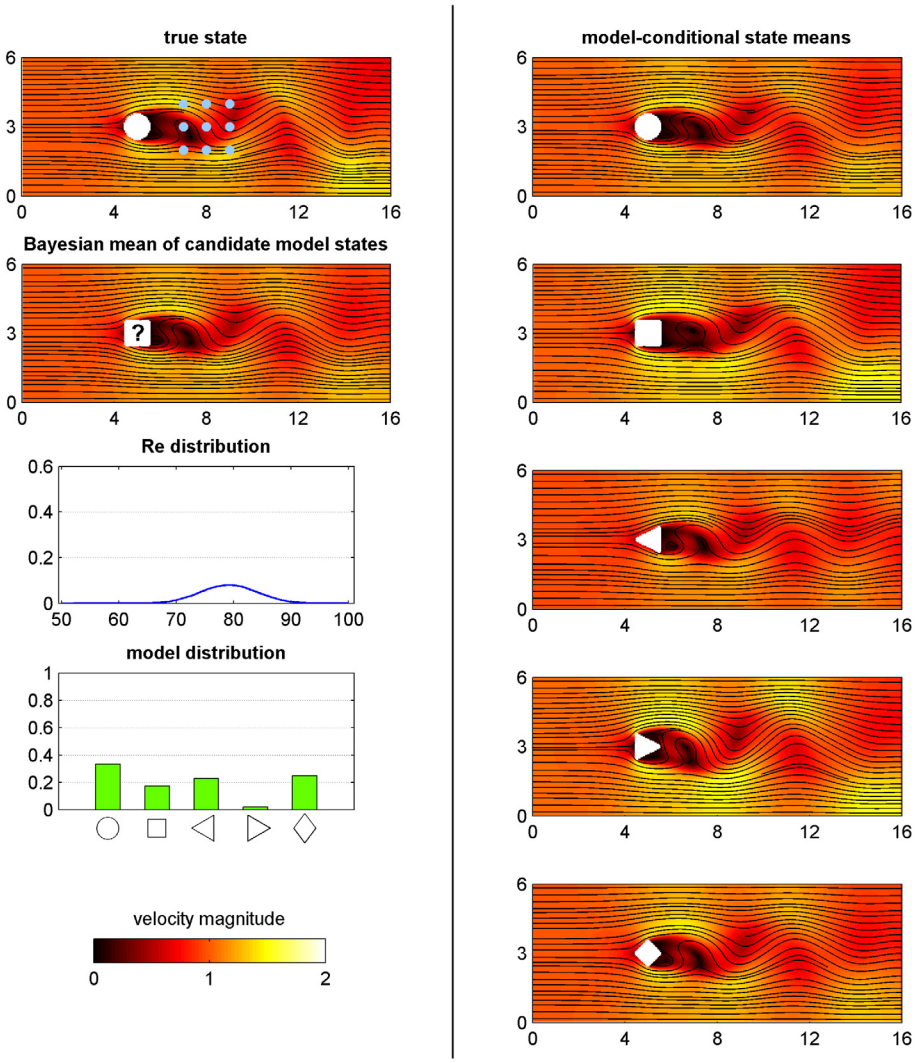

Fig. 7. As Fig. 5, but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

Re mean, with the diamond the closest to the truth. This clearly explains that the model biases (wrong obstacle shape) lead to the multi-modal Re distribution seen in Fig. 9 and to the variations in the now smaller error fields (due to the wrong eddy shedding properties, e.g. Strouhal number).

3.3.1. Learning metrics

Three metrics are used to evaluate the success of our Bayesian learning methodology for the flow past an obstacle. The first metric is the root mean square error (RMSE) between the velocity state variables of the deterministic simulation from which observations are made – i.e. the true velocity state variables – and the predicted Bayesian mean of the candidate model velocity state variables (see Section 2.1.1). A RMSE approaching 0 indicates successful state inference. The second metric is the integral of the predicted Bayesian probability distribution for the Reynolds number of the system from 77.5 to 82.5, a range that is centered on the Re of the deterministic simulations – i.e. the true Reynolds number – and encompasses only 10% of the total initial prior range assumed for the Reynolds number. This metric is referred to as the true Re probability and is calculated by marginalization of the augmented state vector (40) (see Section 2.1.1). A true Re probability approaching 1 indicates successful parameter inference. The third metric is simply the predicted probability for the model of the deterministic simulation from which observations are made – i.e. the true model – within the model distribution. A true model probability approaching 1 indicates successful model learning (since in our example, the true model is among the candidate models). This 'probability of the model' metric easily extends to cases where the truth is not part of the candidate models (e.g. is a linear combination) or is a probabilistic combination of several candidate models.

Fig. 11 illustrates the time progression of the three Bayesian learning metrics introduced above. The RMSE approaches 0, the true Reynolds number probability approaches 1, and the true model probability approaches 1, all indicating successful learning. Even though the nine velocity observations are very sparse in time and space, and the learning space is high-dimensional, our multivariate GMM-DO Bayesian learning can jointly learn the full velocity field, the Reynolds number, and the shape of the obstacle.

Of course, Bayesian learning should not always lead to prefect identification of the true model, true state variables, and/or true parameters. The goal is to evolve the prior and posterior probabilities of the fields, parameters, and model formulations, given

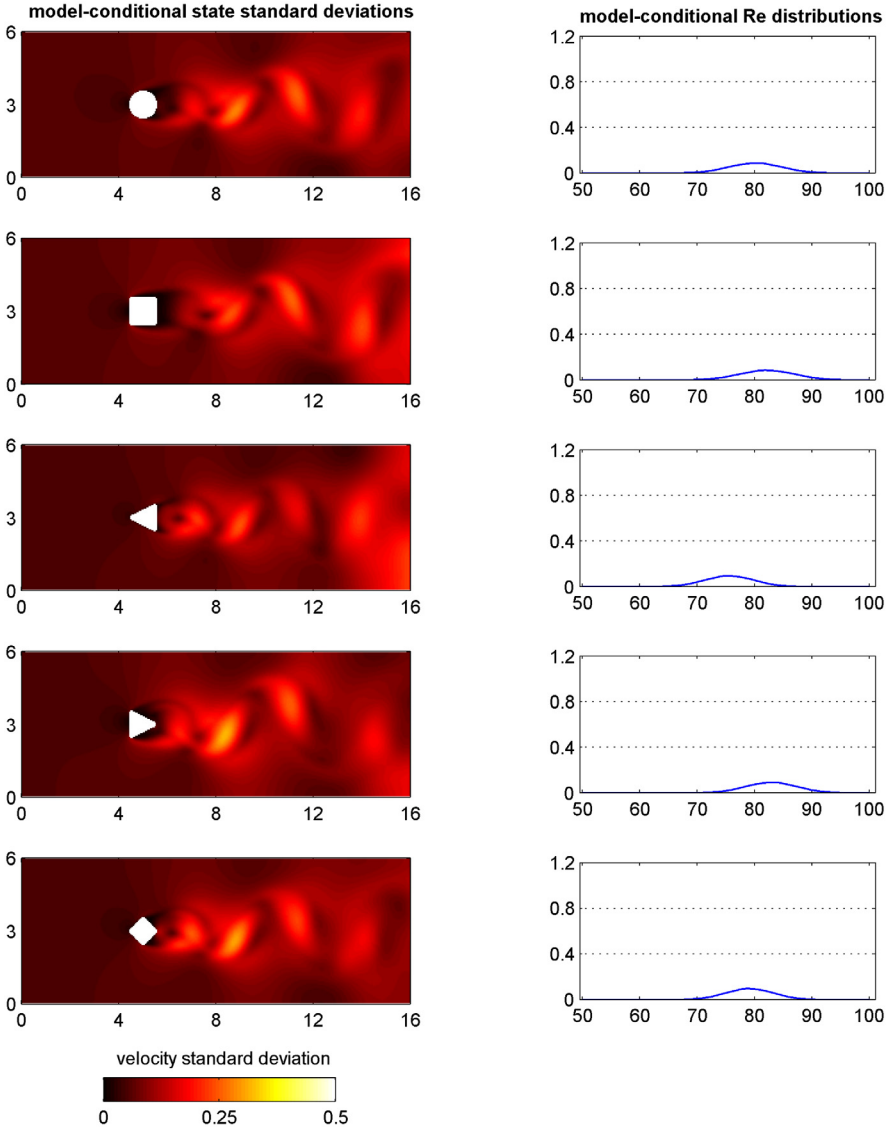

Fig. 8. As Fig. 6, but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

the observations available and their uncertainties. The quality of the learning depends on the observations collected and on the properties of the stochastic dynamical system including its initial and boundary conditions. If the observations are not sufficiently informative about the learning objectives, the Bayesian machine should thus indicate that it cannot learn much. It is only when the exact posteriors indicate successful model identification that this should also be the posterior estimate and conclusion of our GMM-DO Bayesian learning.

3.3.2. Other learning experiments

Many similar learning experiments were performed, including with simulated truths featuring the other four obstacle shapes. First, an interesting result was that model learning proceeded much more quickly for some of the shapes (e.g. the downstream-pointing triangular obstacle) than for others (e.g. the circular obstacle), suggesting that some of the shapes are more distinct among the five possible obstacles than the others. Outcomes were all consistent with known dynamical properties of flows behind such shapes [e.g.150,151] Of course, this rate of learning varied with the dynamics, uncertainty, and observations selected. Second, this exemplifies how our GMM-DO Bayesian model learning (Section 2) can be used as a statistical determinant of the similarity or dissimilarity among stochastic dynamical models. For example, we can say that for the considered Reynolds number range, initial state uncertainty, and observations available, the upstream pointing triangle, square, and diamond obstacles are the 'closest' to the circular obstacle. Finally, we confirmed the convergence of our GMM-DO Bayesian posteriors by repeating learning experiments with an increasing number of DO modes and coefficients (not shown), until the results converged to those shown.

4. Microorganism tracer

While they share a means of physical ocean transport, the many species of phytoplankton and zooplankton in the world's oceans each grow and decay according to unique biological (reaction) dynamics. These microorganisms are of great importance to geochemistry and climate science [152–154] and critical to the health of many species and ecosystems [155–158]. Assuming the continuum hypothesis holds, one can define a microorganism

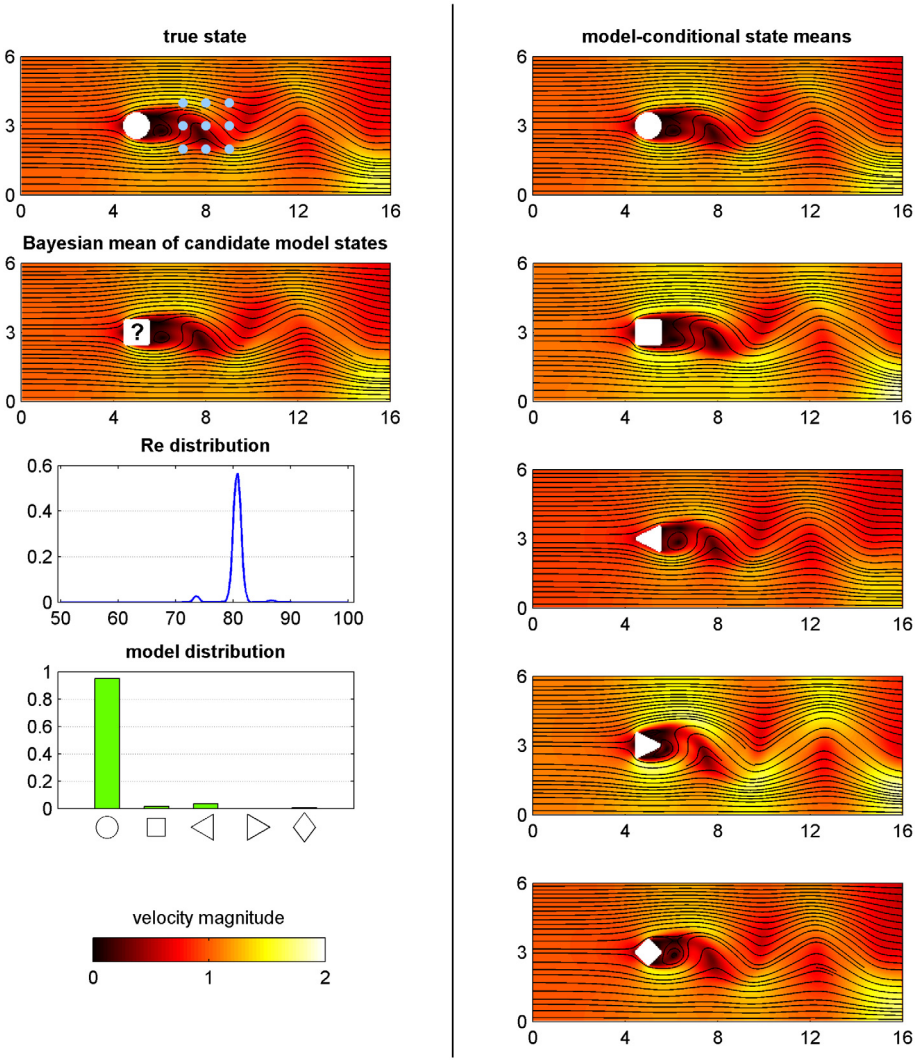

Fig. 9. As Fig. 5, but at non-dimensional time $t = 50$ (i.e. after 50 observation episodes).

concentration. The stochastic dynamical evolution of a microorganism passively advected within a fluid domain can then be described by the general equation,

$$\frac{\partial \rho}{\partial t} = \underbrace{-\boldsymbol{v} \cdot \nabla \rho}_{\text{advection}} + \underbrace{\kappa \nabla^2 \rho}_{\text{diffusion}} + \underbrace{F_\rho\left(\rho, \boldsymbol{r}, t; \omega\right)}_{\text{biology(reaction)}} , \tag{42}$$

where $\rho = \rho(\boldsymbol{r}, t; \omega)$ is the microorganism concentration field per unit mass of fluid, κ the diffusivity of the microorganism within the fluid, and F_ρ the reaction (biological behavior) term for the microorganism [159]. The reaction term commonly involves fields of chemical compounds necessary for growth and of other organisms that act as predators, preys, or competitors for resources. Physical factors also often matter, such as temperature, light abundance, and pressure. If the microorganism is not entirely immotile, motility terms are also included. Learning accurate reaction terms is thus central to marine ecosystems [11, 57,160].

4.1. Stochastic microorganism tracer system

The second stochastic dynamical system considered for our Bayesian model learning (Section 2) is an advection–diffusion–reaction (ADR) Eq. (42) for a single marine microorganism passively advected by a stochastic flow. The 2D spatial domain is shown in Fig. 12, with fluid flow proceeding left to right in the positive r_1 direction. The domain features a circular obstacle. This is an idealization of flows around an island that are known to affect plankton [e.g. 161]. A Reynolds number of 50 is chosen for the system. The microorganism is assumed to be immotile. Uncertainty in the fluid velocity and microorganism concentration fields is introduced in two ways: (1) initial velocity field and (2) reaction term of the microorganism, i.e., the governing biological formulation is itself uncertain.

As in Section 3, the phase of vortex shedding is highly sensitive to the initial fluid velocity field. Since the concentration field is closely coupled to the velocity field through (42), uncertainty in velocity leads to uncertainty in concentration. The different phase shifts of the periodic fluid flow pattern downstream of the obstacle can correspond to substantially different fluid velocity and microorganism concentration fields.

Additional uncertainty in the reaction term F_ρ can significantly increase the uncertainty in its concentration field. Basic phytoplankton reaction equations are assumed for the microorganism behavior, i.e.,

$$F_\rho\left(\rho, \boldsymbol{r}, t; \omega\right) = \underbrace{g(\boldsymbol{r}, t)\frac{\rho(\rho_{\max} - \rho)}{\rho_{\max} - \rho + k_g}}_{\text{growth}} - \underbrace{d(\boldsymbol{r}, t)\rho}_{\text{decay}} , \tag{43}$$

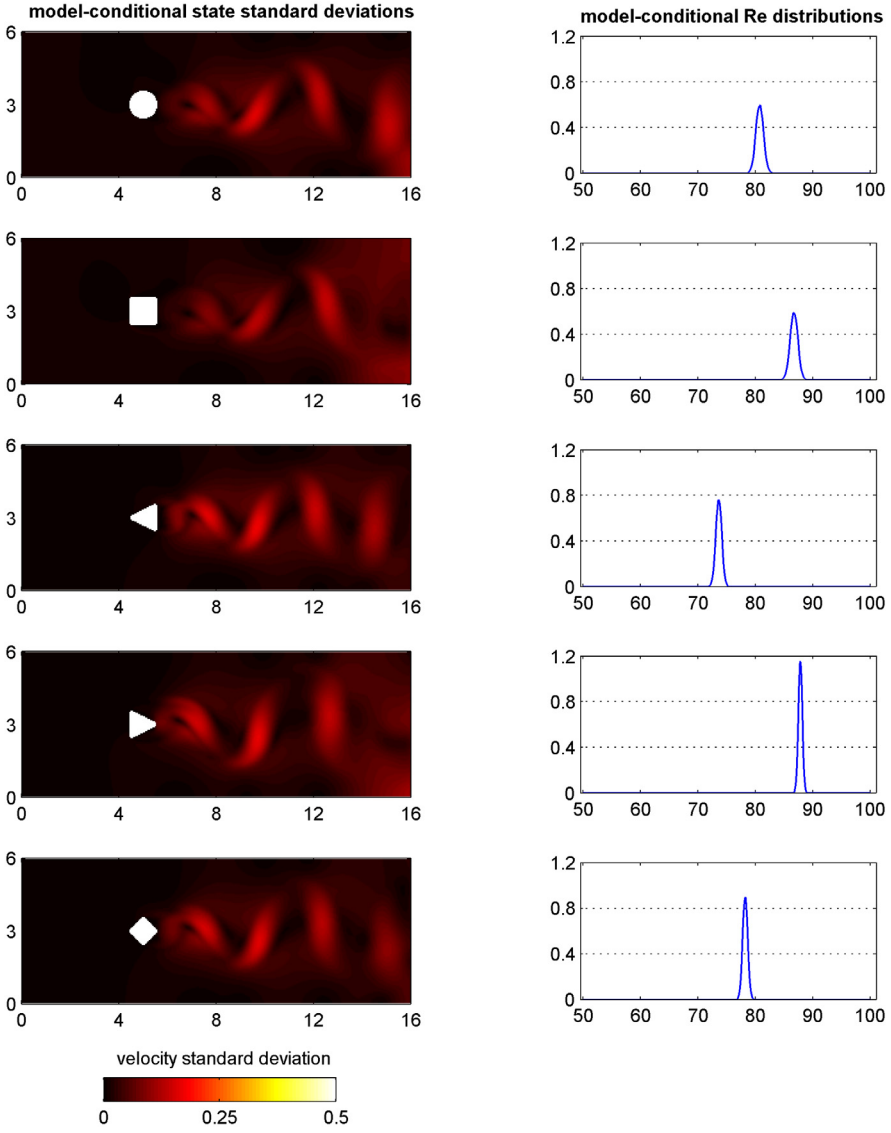

Fig. 10. As Fig. 6, but at non-dimensional time $t = 50$ (i.e. after 50 observation episodes).

where $g(\boldsymbol{r}, t)$ is the microorganism growth factor, $\rho_{\max}$ the maximum microorganism concentration, k_g the growth regularization parameter, and $d(\boldsymbol{r}, t)$ the microorganism decay factor [160]. The values of $\rho_{\max}$ and k_g are both chosen to be 1.0. The formulation of the growth $g(\boldsymbol{r}, t)$ and decay $d(\boldsymbol{r}, t)$ factors however is assumed to be uncertain fields, with three possible formulations for $g(\boldsymbol{r}, t)$,

$$g_+(\boldsymbol{r}, t) = g_+(t) = \left(1 + A_T \sin(t/T_\rho)\right) G \,, \tag{44}$$

$$g_0(\boldsymbol{r}, t) = g_0 = G \,, \tag{45}$$

$$g_-(\boldsymbol{r}, t) = g_-(t) = \left(1 - A_T \sin(t/T_\rho)\right) G \,, \tag{46}$$

and three possible formulations for $d(\boldsymbol{r}, t)$,

$$d_+(\boldsymbol{r}, t) = d_+(\boldsymbol{r}) = \left(1 + A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D \,, \tag{47}$$

$$d_0(\boldsymbol{r}, t) = d_0 = D \,, \tag{48}$$

$$d_-(\boldsymbol{r}, t) = d_-(\boldsymbol{r}) = \left(1 - A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D \,, \tag{49}$$

where $\|\bullet\|$ represents the L_2 norm operator for any given vector and $\boldsymbol{r}_0 = (3, 3)$ represents the center of the circular obstacle illustrated in Fig. 12. The temporal dependence in (44) and (46) simulates possible periodic influences in real ocean environments, such as solar light cycles or tides, while the spatial dependence in (47) and (49) simulates possible near-coast influences, such as bathymetric changes or pollution. The growth factor parameters G, A_T, and T_ρ are chosen to be 0.8, 0.5, and 0.7, respectively. The decay factor parameters D, A_L, and L_ρ are chosen to be 0.1, 1.8, and 1.0, respectively. The combination of the three growth factor forms (44)–(46) and three decay factor forms (47)–(49) leads to a total of nine possible ADR Eq. (42) for the microorganism. The reaction equations are purposely chosen to lead to relatively subtle variations of the concentration field, as exhibited in Fig. 13. This choice is to showcase that our Bayesian learning can predict accurate posterior model probabilities and correctly identify the true model formulation when the data and dynamics enable it, even in cases where a human may be challenged to do the same.

4.2. Description of the learning experiments

4.2.1. Model formulation

Each of the nine possible sets of stochastic dynamical PDEs (2D flows and microorganism concentration) represents a distinct candidate model for the system. Each candidate is assumed to be

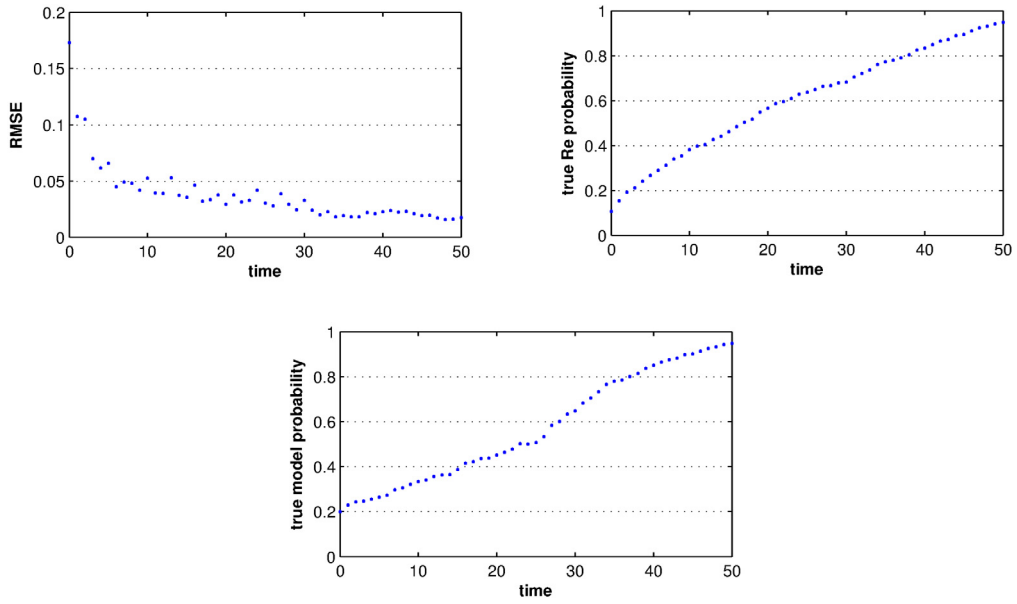

Fig. 11. Time progression of three learning metrics for the stochastic flow past an obstacle system, with observations made from the simulated truth featuring the circular obstacle. Metrics (RMSE of velocity field, probability of the true Re, and probability of the true model obstacle) are plotted at every 1 non-dimensional time, immediately after observation and learning (i.e. posterior values from the GMM-DO filter).

equally likely a priori, i.e. a uniform initial model distribution is used. We now describe the components (2) of these candidates.

$\mathcal{D}_n$ — As in Section 3, the evolution of the fluid velocity field in the stochastic microorganism tracer system is simulated using the non-dimensional incompressible 2D Navier–Stokes PDEs (37)–(39). This velocity forces the non-dimensional form of the general ADR PDE (42) with reaction Eq. (43). With the growth (44)–(46) and decay factors (47)–(49), nine distinct candidate models are possible. The flow parameters are here assumed to be known. Specifically, Re in (37)–(38) is chosen to be 50 while κ in (42) is chosen to be 0.

SG_n — Unlike the stochastic flow past an obstacle, the spatial geometries for the nine candidate models here are identical, as shown in Fig. 12. For numerical implementation, the spatial domain is discretized using a regular rectangular grid of 240 and 120 elements along the r_1 and r_2 directions, respectively ($\Delta r_1 = \Delta r_2 = \frac{1}{20}$). This discretization results in 28,349 r_1-velocity, 28,229 r_2-velocity, and 28,489 microorganism concentrations for each candidate model. The state vector for any of the candidate models is then

$$\boldsymbol{x}(t;\omega) = \begin{bmatrix} \boldsymbol{v_1}(t;\omega) \\ \boldsymbol{v_2}(t;\omega) \\ \boldsymbol{\rho}(t;\omega) \end{bmatrix} ,$$

where $\boldsymbol{\rho}(t;\omega)$ is the spatially-discretized vector of concentration state variables (state augmentation is not needed because all parameters are known). The dimension of the state vector, $N_{\boldsymbol{X}}$, is thus 85,067.

BC_n — For all nine candidate models, deterministic boundary conditions for the velocity and concentration fields are used: Dirichlet for the left domain inflow,

$$\boldsymbol{v} = (1,0)\ ,\ \rho = 0.1 \quad \text{at } r_1 = 0\ , \tag{50}$$

Neumann for the top, bottom, and right domain limits,

$$\frac{\partial \boldsymbol{v}}{\partial r_2} = 0\ ,\ \frac{\partial \rho}{\partial r_2} = 0 \quad \text{at } r_2 = 0 \text{ and } r_2 = 6\ ,$$
$$\frac{\partial \boldsymbol{v}}{\partial r_1} = 0\ ,\ \frac{\partial \rho}{\partial r_1} = 0 \quad \text{at } r_1 = 12\ .$$

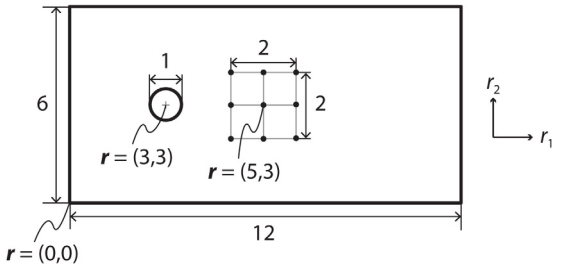

Fig. 12. 2D spatial domain of the stochastic microorganism tracer system. All lengths and coordinates are non-dimensional. The observation array consists of the nine locations at the center of the domain.

and Dirichlet on the surface of the obstacle, imposing no-slip conditions.

IC_n — For each of the nine candidate models, the state vector mean in (28) is initialized as follows. The means of the velocity state variables are obtained by numerically solving the continuity Eq. (39), with the boundary conditions given above. The means of the microorganism concentration states variables are subsequently found by numerically solving (42) for the steady state concentrations corresponding to these mean velocities. Figs. 14, 15, and 16 illustrate the stochastic system at the initial non-dimensional time ($t = 0$). For each candidate model, a covariance matrix is constructed for the velocity and concentration state variables that respects symmetry about the $r_2 = 3$ centerline, using again the boundary-mollified spatial covariance method [59]. Discrete Karhunen–Loève transforms are performed on these covariance matrices to initialize eight pairs of DO vectors and coefficients for each candidate model, for a DO subspace dimension $N_{\mathrm{DO}} = 8$. As for the stochastic flows past an obstacle, the symmetry properties of this procedure lead to unbiased initialization of uncertainty in vortex shedding phase, as exhibited in Figs. 15 and 16.

4.2.2. True solution generation

We illustrate results with two different true solutions, each of which has its own formulation for the microorganism and

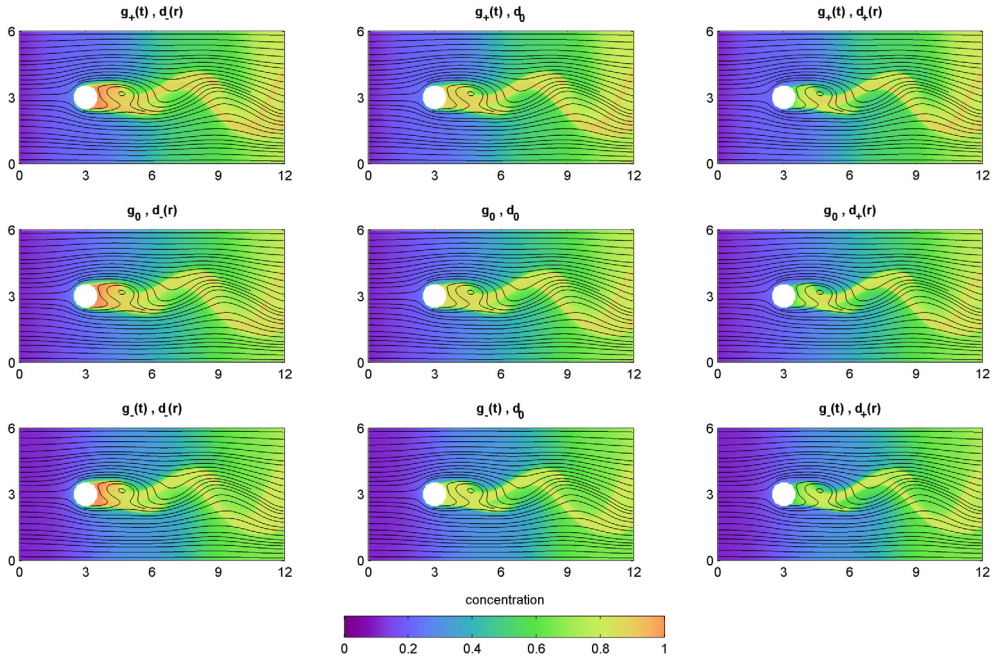

Fig. 13. Microorganism concentrations overlaid on fluid streamlines, for realizations of the stochastic microorganism tracer system featuring nine different microorganism reaction equations at non-dimensional time $t = \pi \times T_\rho \approx 2.2$ (i.e. when periodic temporal effects are either maximal or minimal).

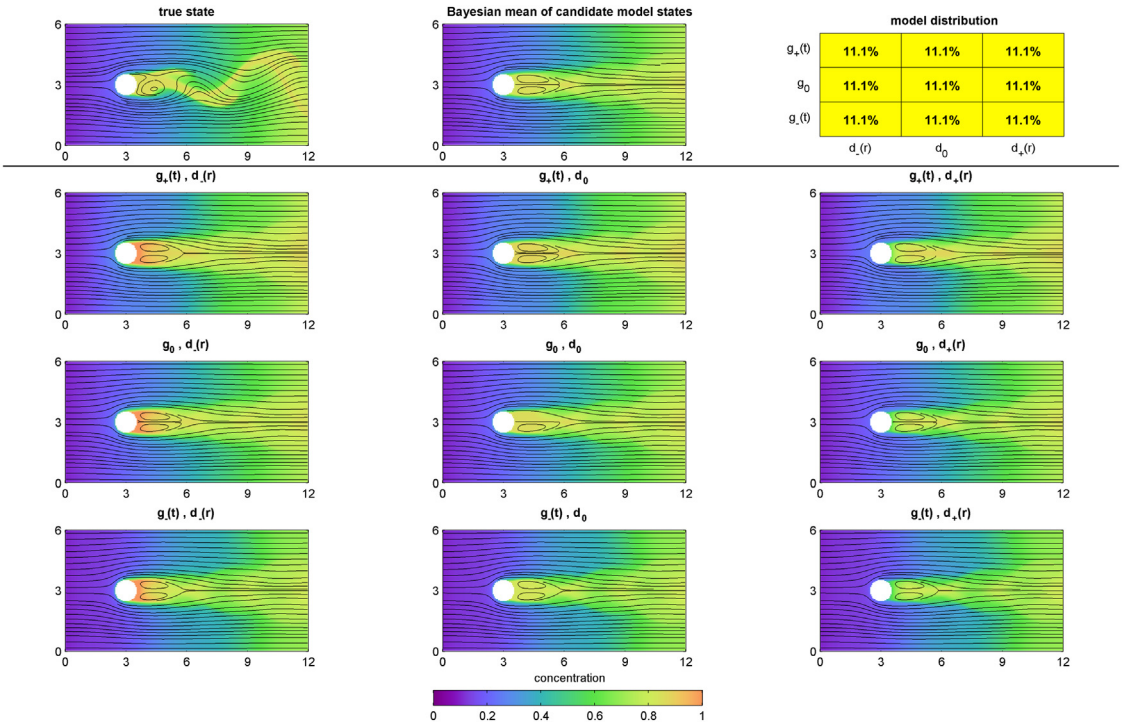

Fig. 14. The stochastic microorganism tracer system at non-dimensional time $t = 0$. Top row: Deterministic simulated truth state fields (left) with constant growth g_0 and decay d_0 factors, from which observations are made (see Fig. 2). Note the active vortex shedding and microorganism response in these truth fields. Initial Bayesian mean of the candidate model state fields (middle) and uniform model distribution illustration (right). Bottom three rows: Initial model-conditional velocity (streamline) and concentration field means for each of the nine candidate models, arranged in an array corresponding to the model distribution illustration in the top right corner.

represents the 'truth' for the corresponding learning experiments. The first (at $t = 0$ in Fig. 14) features the constant formulations for growth factor, g_0, and decay factor, d_0, thus a spatially and temporally invariant microorganism reaction equation. The second features the growth factor formulation

$$g'(\boldsymbol{r}, t) = g'(t) = \left(1 + (1/2)A_T \sin(t/T_\rho)\right) G \tag{51}$$

and decay factor formulation

$$d'(\boldsymbol{r}, t) = d'(\boldsymbol{r}) = \left(1 - (1/2)A_L \exp\left(-(\|\boldsymbol{r} - \boldsymbol{r}_0\| / L_\rho)^2\right)\right) D\,, \tag{52}$$

representing a reaction equation with intermediate spatial and temporal variance that, importantly, is *not explicitly included* in the set of candidate models formulated for the system.

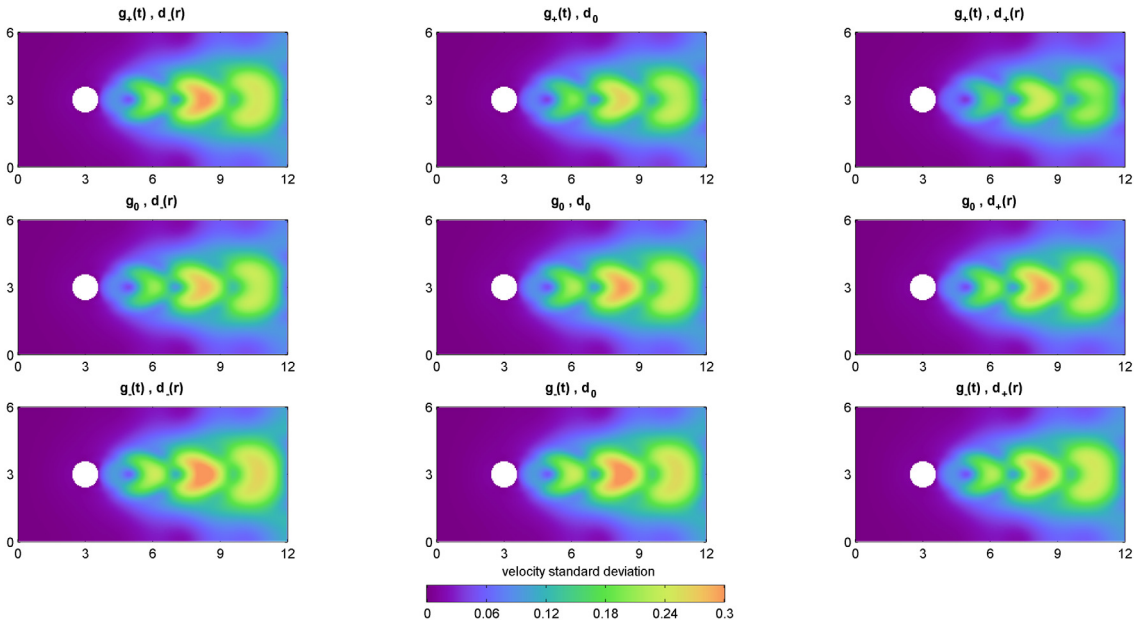

Fig. 15. As Fig. 14, but showing model-conditional uncertainties at non-dimensional time $t = 0$. Specifically, model-conditional velocity field standard deviations for each of the nine candidate models are displayed, arranged in an array corresponding to the model distribution illustration in the top right corner of Fig. 14.

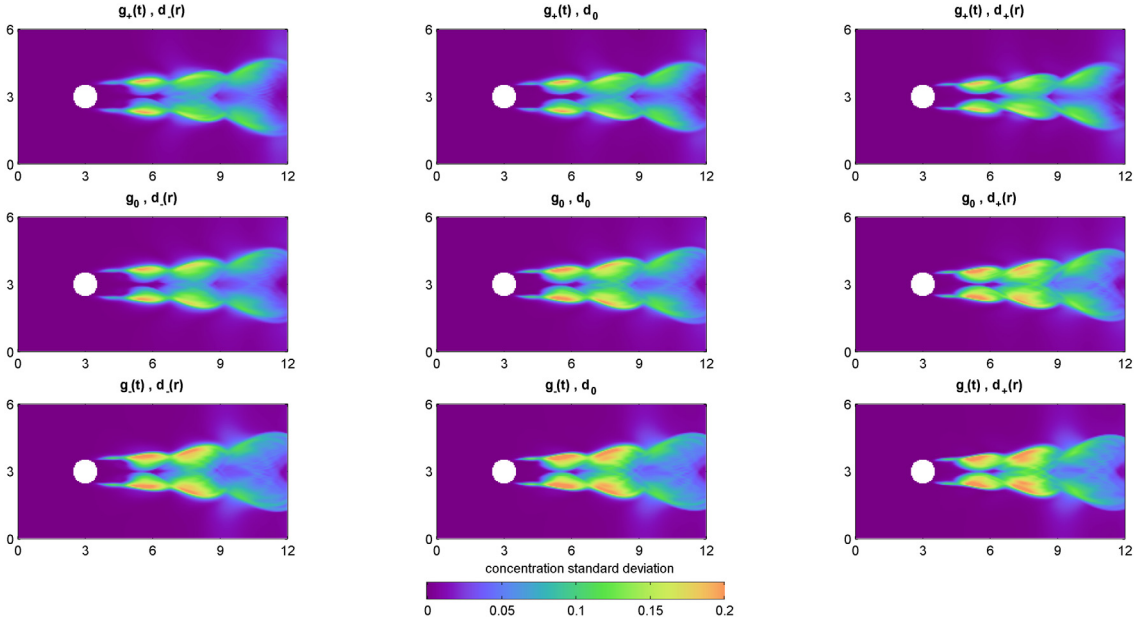

Fig. 16. As Fig. 15, but for the model-conditional concentration field standard deviations.

For each truth, initial velocity fields satisfying the continuity Eq. (39) are first constructed, with arbitrary asymmetries introduced to induce vortex shedding. Microorganism concentrations are initialized at 0.1. The evolution of the velocity, pressure, and concentration fields is computed by numerically solving the Navier–Stokes Eqs. (37)–(39) and microorganism Eq. (42), using our second-order finite-volume scheme [149], with a regular rectangular grid of 240 and 120 elements along the r_1 and r_2 directions, respectively, and a non-dimensional time-step of $\Delta t = \frac{1}{120}$ up to a final non-dimensional time $T = 80$. These deterministic simulations are two possible realizations of the stochastic microorganism tracer system, each with a true microorganism reaction equation, true dynamical velocity field, and true dynamical concentration field that are to be jointly learned. They are used in independent experiments whose results are illustrated in Section 4.3.

4.2.3. Observations and learning parameters

Every 1 non-dimensional time unit (i.e. every 120 time-steps), noisy observations of microorganism concentration are made at nine locations (Fig. 12), for a total of $N_{\boldsymbol{Y}} = 9$ scalar data at every data time. A single non-dimensional time unit for this system corresponds to a phase shift of about $\frac{\pi}{3}$ when Re $= 50$. Unbiased Gaussian noise is applied to the observations with a standard deviation equal to approximately 5% of the maximum microorganism concentration, which are then used by the GMM-DO filter to perform joint model-conditional inference of the velocity and concentration state variables, for each of the candidate models. The observation matrix $\mathbf{H}$ in (23) selects the concentration state

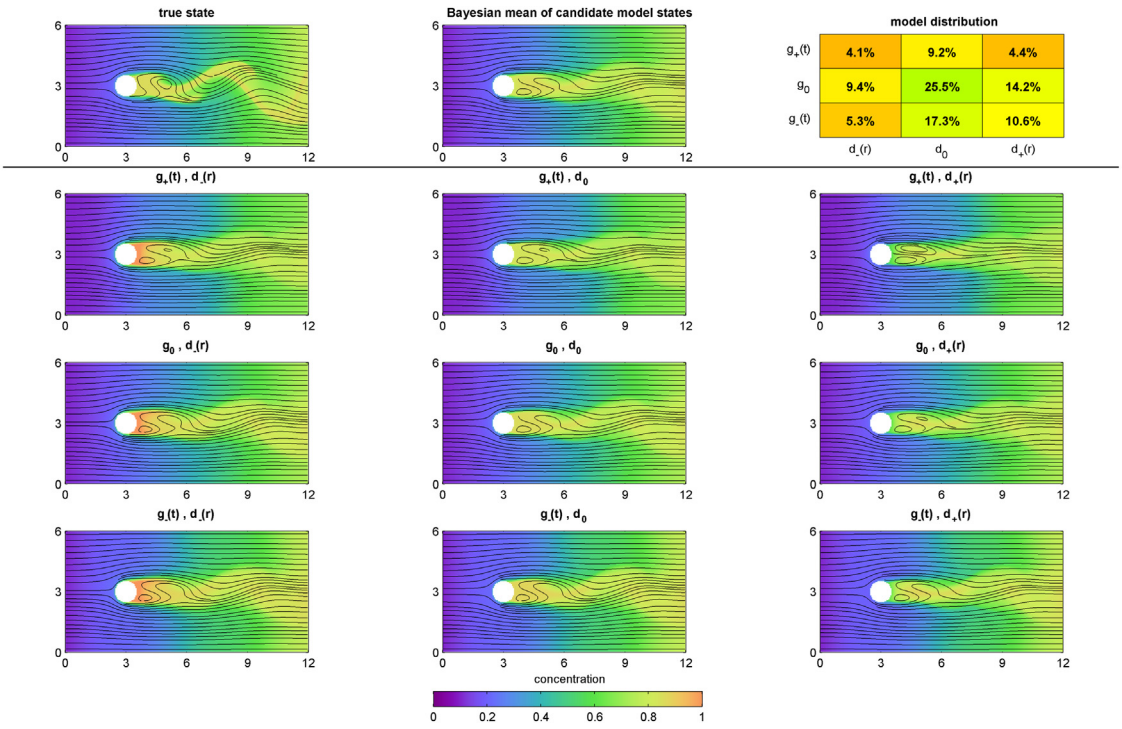

Fig. 17. As Fig. 14, but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

Table 2
Numerical properties for the stochastic microorganism tracer system.

Property	$N_{\mathcal{M}}$	$N_{\boldsymbol{X}}$	N_{Θ}	$N_{\boldsymbol{Y}}$	N_{DO}	N_{MC}	Δr_1	Δr_2	Δt	T
Value	9	85,067	0	9	8	10^4	1/20	1/20	1/120	80

variables at the nine data locations and the covariance matrix $\mathbf{R}$ is diagonal of elements equal to the variance of the observation noise.

As in [118], the BIC and EM algorithm are used to select the optimal number of GMM components N_{GMM} at every data time. Typical BIC-optimized values for N_{GMM} were found to lie between 40 and 60 for these stochastic microorganism experiments. Marginal likelihoods for the candidate models are calculated using (35) after every GMM-DO filtering step and the model distribution is updated using (7).

4.2.4. Numerical method

For each of the candidate models, the state vector mean, DO vectors, and DO coefficients are governed by the DO Eqs. (14)–(16) for the Navier–Stokes Eqs. (37)–(39) and microorganism tracer Eq. (42), and their respective boundary conditions [88]. They are numerically integrated using the finite-volume methodology developed in [123]. For the DO decomposition of the non-polynomial nonlinearities of the microorganism reaction terms (43), a local linearization is employed [162,163]. These locally linearized terms are derived in [88]. A non-dimensional time-step of $\Delta t = \frac{1}{120}$ is used. A total of $N_{\mathrm{MC}} = 10^4$ Monte Carlo samples is used for the stochastic evolution of the DO coefficients. The numerical properties are summarized in Table 2.

4.3. Learning results

We now present the results of learning experiments, first using true models that are one of the candidate models and second using a true model that is not. In each case, for learning metrics, we employ the Bayesian state variable mean RMSE and the probability of the true model or of the models near the unknown truth (i.e. near in model space). Since the fluid velocity and microorganism concentration state variables are all normalized to be of order 1, they are included in RMSE calculations without weighting. A RMSE approaching 0 and a true model probability approaching 1 again indicate successful state and model learning, respectively. Of course, our GMM-DO Bayesian learning aims to predict the correct posterior probabilities, indicating successful learning only when the exact posteriors also indicate so.

4.3.1. Constant growth and decay: Truth is one of the candidate models

The first experiment considers the simulated truth defined by the deterministic simulation featuring the constant growth g_0 and decay d_0 factors. The goal is to learn the state of the velocity and microorganism concentration fields jointly with the formulation of the reaction equation when noisy observations are made from this simulated truth. The evolution of this GMM-DO Bayesian learning is illustrated in Figs. 14 to 18.

Figs. 14, 15, and 16 illustrated the system at $t = 0$, before observations are made. The fluid velocity and microorganism concentration fields of the simulated truth are in a time-periodic state. Regular asymmetric vortex shedding is occurring downstream of the obstacle, with a period of approximately 6.0 non-dimensional time units. The model-conditional velocity and concentration field means for all nine candidate models are symmetric about the $r_2 = 3$ centerline, indicating no bias towards any particular vortex shedding phase. The model distribution, i.e. $p_{\mathcal{M}}(\bullet)$ in (5), is uniform. The model-conditional velocity and concentration standard deviations are also symmetric about the $r_2 = 3$ centerline. Crucially, these initial model-conditional uncertainties are not small, the largest standard deviations are close to the total variability.

Fig. 17 illustrates the system after 10 non-dimensional time units and 10 observation episodes. Each of the candidate models performs 10 Bayesian GMM-DO assimilation (details not shown). The model-conditional microorganism concentration field means

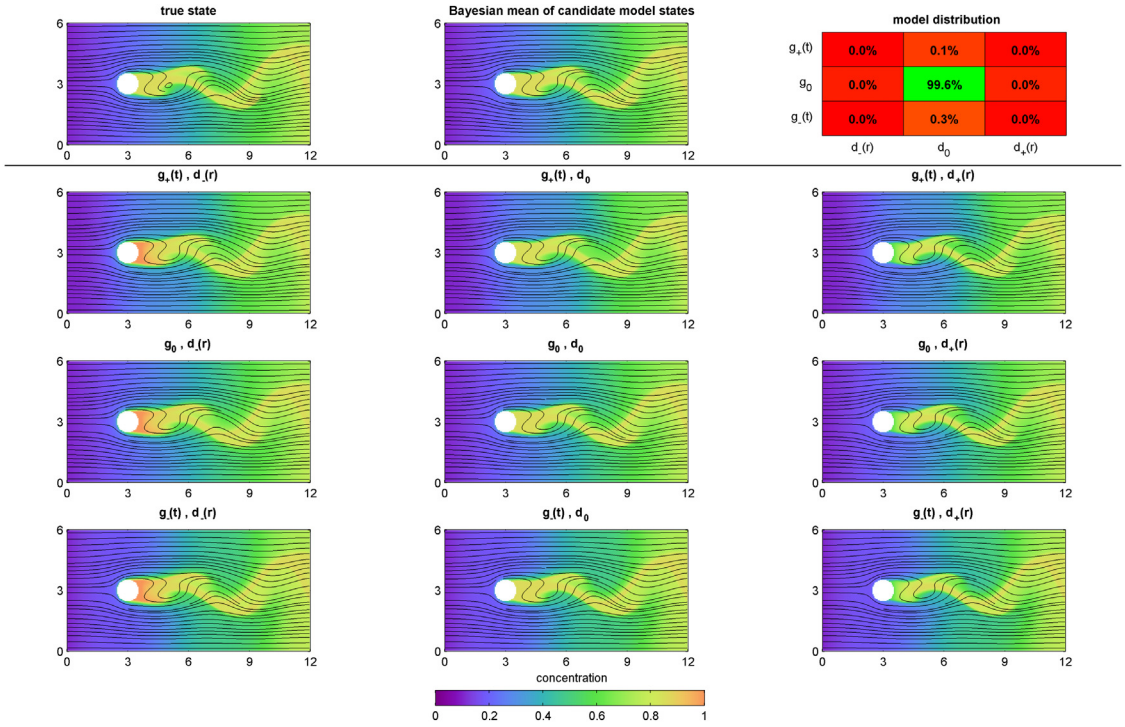

Fig. 18. As Fig. 14, but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

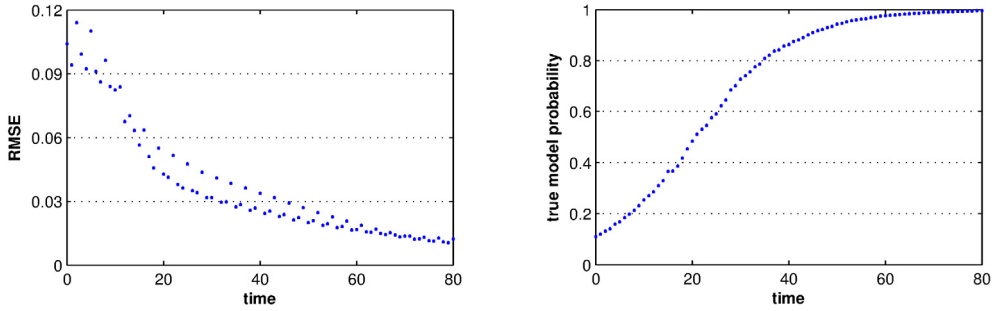

Fig. 19. Time progression of learning metrics for the stochastic microorganism tracer system, with observations made from the simulated truth featuring constant growth g_0 and decay d_0 factors. Metrics (joint RMSE of velocity and microorganism fields as well as the probability of the true model) are plotted at every 1 non-dimensional time, immediately after observation and learning (i.e. posterior values from the GMM-DO filter).

for all nine candidate models are beginning to align in phase with the true concentration field. Despite the absence of velocity observations, the conditional velocity field means for the candidates are also beginning to align with the true velocity field, as a result of the joint Bayesian inference of microorganism concentration and fluid velocity. Inference of the vortex shedding phase behind the obstacle however is far from complete, as evidenced by the blurred features in the Bayesian mean of these conditional concentration field means, when compared to those in the true field. The true model of the system is favored by the model distribution and, importantly, 'adjacent' models – i.e. models featuring smaller deviations in their growth and decay factor forms – have higher probabilities than distant models.

Fig. 18 illustrates the system after 80 non-dimensional time units and 80 observation episodes. Agreement between the true velocity and concentration fields and the Bayesian means of the inferred velocity and concentration fields has improved significantly. The true model is now also the only plausible model for the system, as indicated by the highly peaked model distribution for the unknown but true reaction model.

Fig. 19 illustrates the time progression of the two learning metrics mentioned above. The RMSE approaches zero and the true model probability approaches one, both indicating successful learning.

4.3.2. Other learning experiments: Truth is one of the candidate models

Many similar learning experiments were performed with simulated truths featuring other combinations of the three time-dependent growth factor formulations (44)–(46) and three spatially-variable decay factor formulations (47)–(49). In these experiments, however, one of the candidate models still had the same formulation as the true model. The results (not shown) were analogous to those illustrated in Figs. 14–18. Of note, simulated truths featuring growth and decay factor combinations on the periphery of the model space (e.g. (44) and (49)) exhibited faster learning rates than those featuring combinations near the center of the model space (e.g. (45) and (48)). This is due to the fact that the former have fewer 'adjacent' candidate models than the latter and are hence more dynamically distinct within the set of

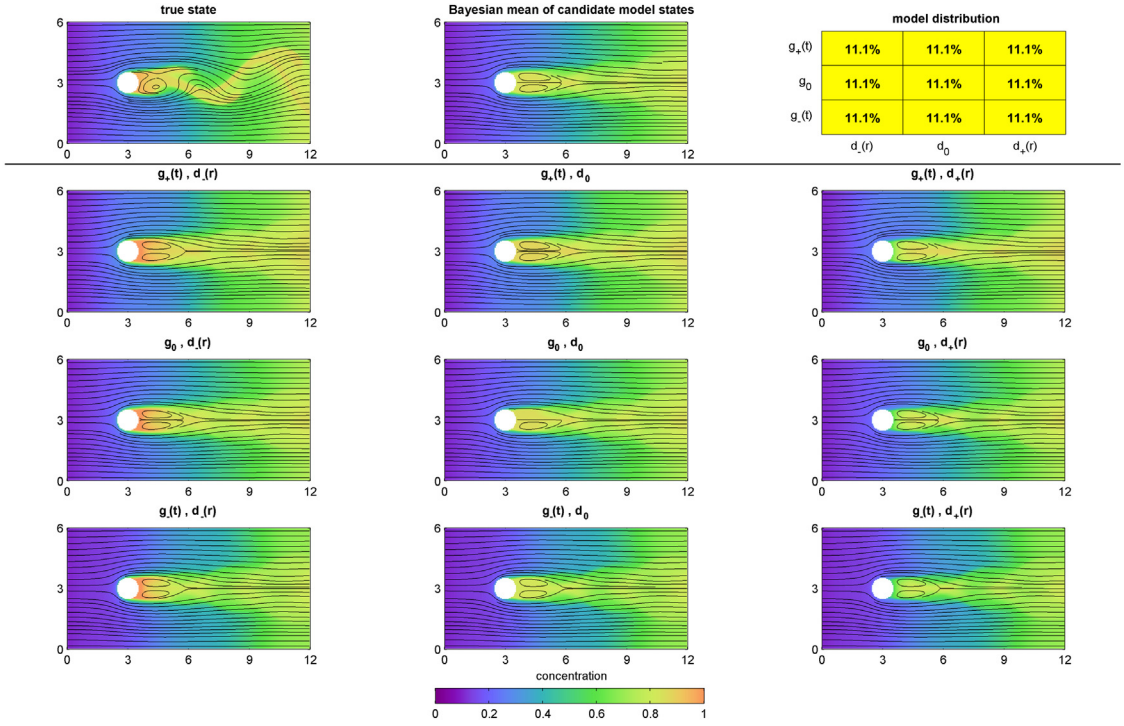

Fig. 20. As Fig. 14, but for the experiment where the simulated truth uses the growth factor $g'(t)$ and decay factor $d'(\boldsymbol{r})$ formulations, and is not part of the nine candidate models.

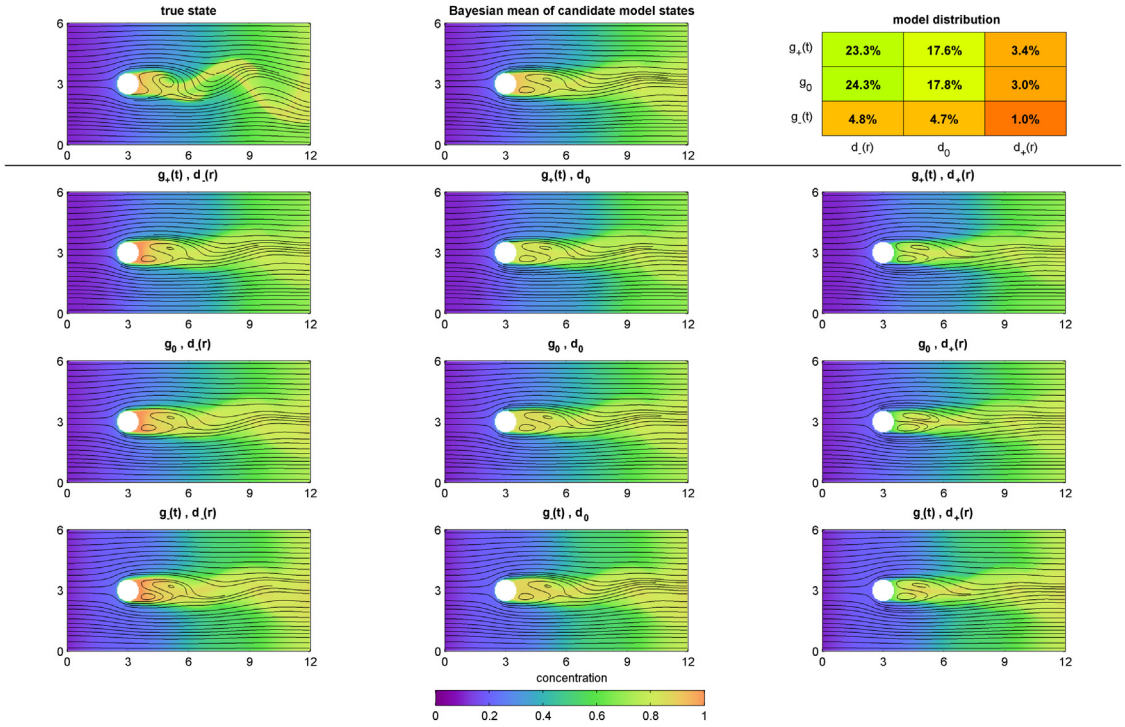

Fig. 21. As Fig. 20, but at non-dimensional time $t = 10$ (i.e. after 10 observation episodes).

candidate models. Finally, we also confirmed the convergence of our GMM-DO Bayesian posteriors by repeating several learning experiments with an increasing number of modes and coefficients (not shown), until the results converged.

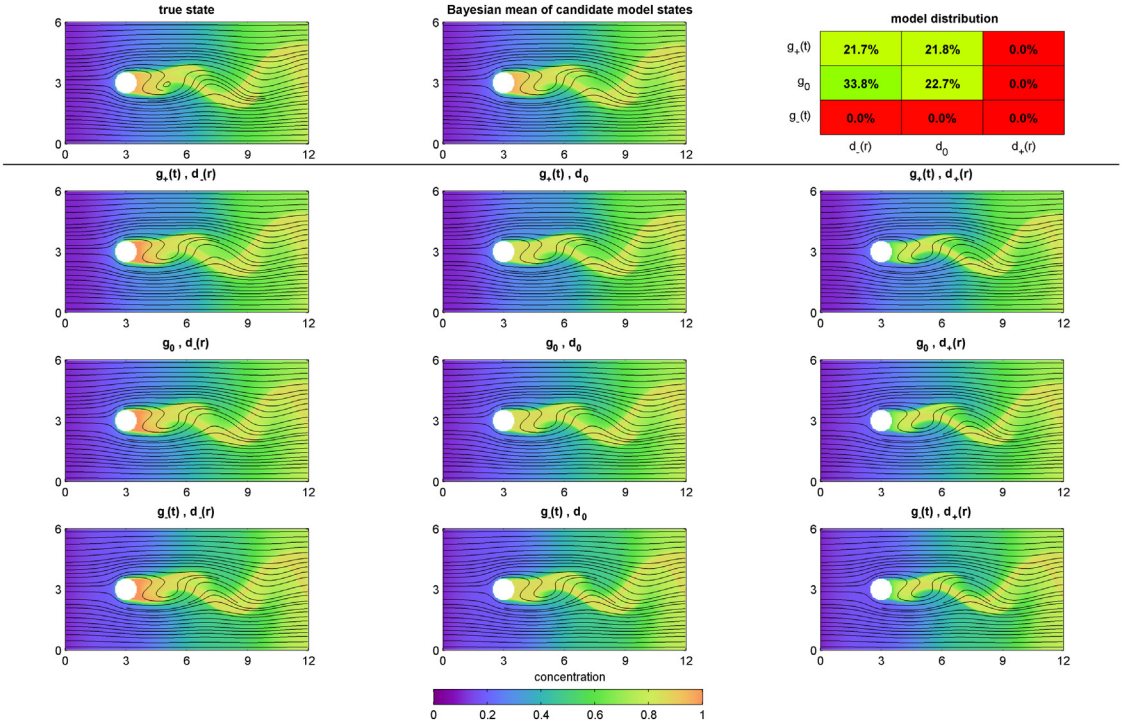

Fig. 22. As Fig. 20, but at non-dimensional time $t = 80$ (i.e. after 80 observation episodes).

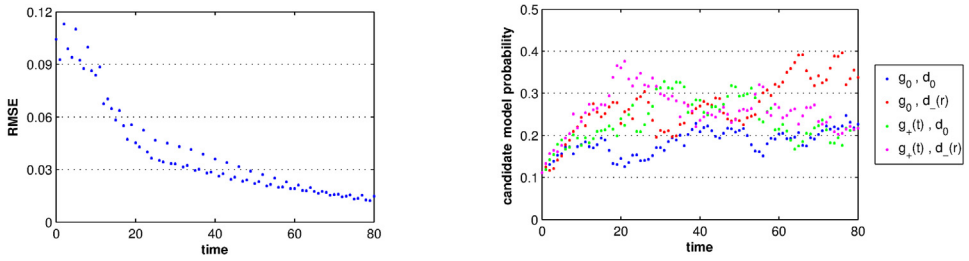

Fig. 23. As Fig. 19, but for the experiment where the truth uses the growth $g'(t)$ and decay $d'(\boldsymbol{r})$ factor formulations. Since the true model is not explicitly represented within the set of candidate models, the probabilities of the four candidates most dynamically similar to the true model are plotted in place of the true model probability.

4.3.3. Time-dependent growth and spatially-variable decay: Truth is not one of the candidate models

The final experiment considers the simulated truth defined by the deterministic simulation featuring the intermediate time-dependent growth and spatially-variable decay factor formulations $g'(t)$ (51) and $d'(\boldsymbol{r})$ (52), respectively. This combination of growth and decay is not explicitly represented as one of the nine candidate models for the system. It is only implicitly accounted for as a linear intermediary between the four candidate models with the two growth factors $g_+(t)$ and g_0 and the two decay factors d_0 and $d_-(\boldsymbol{r})$. The goal of this experiment is to find out if the formulation of the reaction terms (their probabilities) can still be learned jointly with the velocity and microorganism fields from noisy observations made from the simulated truth. The evolution of the GMM-DO Bayesian learning is illustrated in Figs. 20 to 22.

Fig. 20 illustrates the system at $t = 0$, before observations are made. The fluid velocity and microorganism concentration fields of the simulated truth are again in a time-periodic state. As in other experiments, the initial model-conditional velocity and microorganism means for all nine candidate models are symmetric about the $r_2 = 3$ centerline and the model distribution is uniform.

Fig. 21 illustrates the system after 10 non-dimensional time units and 10 observation episodes. The model-conditional microorganism means for all nine candidate models are beginning to exhibit slight asymmetries, indicating partial inference. The model distribution is clearly favoring the candidate models that simultaneously feature one of the two growth factors $g_+(t)$ and g_0 and one of the two decay factors d_0 and $d_-(\boldsymbol{r})$ — i.e. the four candidate models that encircle the true model. The probabilities of the other candidate models are falling quickly.

Fig. 22 illustrates the system after 80 non-dimensional time units and 80 observation episodes. Agreement between the true velocity and microorganism fields and the Bayesian means of the inferred velocity and microorganism fields has improved significantly. The model distribution now indicates that the only plausible models are the four candidate models that encircle the true model. No clear favorite among the four encircling candidate models has emerged though, strongly suggesting that the true model is an intermediary.

Fig. 23 shows the evolution of learning metrics, specifically the joint RMSE of the state variables' Bayesian means as well as the

Table 3
Notation compendium.

General		
$N_{\boldsymbol{X}}$	$\in \mathbb{N}$	Dimension of state vector
$\boldsymbol{X}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	State vector
$\boldsymbol{x}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	State vector realization
$N_{\boldsymbol{\Theta}}$	$\in \mathbb{N}$	Dimension of parameter vector
$\boldsymbol{\Theta}$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	Parameter vector
$\boldsymbol{\theta}$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	Parameter vector realization
$\mathcal{M}$		Stochastic dynamical model
$N_{\mathcal{M}}$	$\in \mathbb{N}$	Number of candidate models
n	$\in \{1, \ldots, N_{\mathcal{M}}\}$	Candidate model index
$\mathcal{M}_n$		nth candidate model
$\mathcal{D}_n$		Stochastic dynamical equations of nth candidate model
SG_n		Spatial geometry of nth candidate model
BC_n		Boundary conditions of nth candidate model
IC_n		Initial conditions of nth candidate model
$N_{\boldsymbol{Y}}$	$\in \mathbb{N}$	Dimension of observation vector
$\boldsymbol{Y}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	Observation vector
$\boldsymbol{y}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	Observation vector realization
DO evolution equations		
$\bar{\boldsymbol{x}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	State vector mean
N_{DO}	$\in \mathbb{N}$	Dimension of stochastic subspace
i	$\in \{1, \ldots, N_{\mathrm{DO}}\}$	DO vector index
$\boldsymbol{\mathcal{X}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\mathrm{DO}}}$	Matrix of DO vectors
$\tilde{\boldsymbol{x}}_i$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	ith DO vector
$\boldsymbol{\Phi}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	DO coefficient vector
$\boldsymbol{\phi}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	DO coefficient vector realization
ϕ_i	$\in \mathbb{R}$	ith DO coefficient
N_{MC}	$\in \mathbb{N}$	Number of Monte Carlo samples
k	$\in \{1, \ldots, N_{\mathrm{MC}}\}$	Monte Carlo sample index
$\boldsymbol{x}_k$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	kth state vector sample
$\boldsymbol{\theta}_k$	$\in \mathbb{R}^{N_{\boldsymbol{\Theta}}}$	kth DO coefficient vector sample
GMM-DO filter		
$\mathbf{H}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{X}}}$	Linear observation matrix
$\mathbf{R}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\boldsymbol{Y}}}$	Observation covariance matrix
N_{GMM}	$\in \mathbb{N}$	Number of GMM components
j	$\in \{1, \ldots, N_{\mathrm{GMM}}\}$	GMM component index
$\pi_{\boldsymbol{X},j}$	$\in \mathbb{R}$	jth (component) weight of prior state GMM
$\boldsymbol{\mu}_{\boldsymbol{X},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	jth mean vector of prior state GMM
$\boldsymbol{\Sigma}_{\boldsymbol{X},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}}$	jth covariance matrix of prior state GMM
$\pi_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}$	jth weight of prior coefficient GMM
$\boldsymbol{\mu}_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	jth mean vector of prior coefficient GMM
$\boldsymbol{\Sigma}_{\boldsymbol{\Phi},j}$	$\in \mathbb{R}^{N_{\mathrm{DO}} \times N_{\mathrm{DO}}}$	jth covariance matrix of prior coefficient GMM
$\pi_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}$	jth weight of posterior state GMM
$\boldsymbol{\mu}_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}}$	jth mean vector of posterior state GMM
$\boldsymbol{\Sigma}_{\boldsymbol{X}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{X}}}$	jth covariance matrix of posterior state GMM
$\mathbf{K}_j$	$\in \mathbb{R}^{N_{\boldsymbol{X}} \times N_{\boldsymbol{Y}}}$	jth gain matrix
$\pi_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}$	jth weight of posterior coefficient GMM
$\boldsymbol{\mu}_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	jth mean vector of posterior coefficient GMM
$\boldsymbol{\Sigma}_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\mathrm{DO}} \times N_{\mathrm{DO}}}$	jth covariance matrix of posterior coefficient GMM
$\tilde{\boldsymbol{y}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}}}$	Transformed observation vector realization
$\tilde{\mathbf{H}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times N_{\mathrm{DO}}}$	Transformed observation matrix
$\tilde{\mathbf{K}}_j$	$\in \mathbb{R}^{N_{\mathrm{DO}} \times N_{\boldsymbol{Y}}}$	jth transformed gain matrix
$\boldsymbol{\mu}'_{\boldsymbol{\Phi}\|\boldsymbol{Y},j}$	$\in \mathbb{R}^{N_{\mathrm{DO}}}$	jth intermediate mean vector
Reduced-dimension state augmentation		
$\boldsymbol{X}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	Augmented state vector
$\boldsymbol{x}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	Augmented state vector realization
$\mathcal{M}_{\boldsymbol{\Theta}}$		Augmented stochastic dynamical model
$\bar{\boldsymbol{x}}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	Augmented state vector mean
$\tilde{\boldsymbol{x}}_{\boldsymbol{\Theta},i}$	$\in \mathbb{R}^{N_{\boldsymbol{X}}+N_{\mathrm{DO}}}$	ith augmented DO vector
$\mathbf{H}_{\boldsymbol{\Theta}}$	$\in \mathbb{R}^{N_{\boldsymbol{Y}} \times (N_{\boldsymbol{X}}+N_{\mathrm{DO}})}$	Augmented linear observation matrix
Operators, functions, and indicators		
$\mathrm{E}[\bullet]$		Expectation operator
$*$		Convolution operator
$\mathcal{L}(\boldsymbol{y} \mid \bullet)$		Observation likelihood function
$\mathcal{N}(\bullet;\, \boldsymbol{\mu},\, \boldsymbol{\Sigma})$		Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$
$(\cdot)^-$		Prior
$(\cdot)^+$		Posterior

probabilities of the four candidate models most dynamically similar to the true model. Surprisingly, the RMSE still approaches 0, even though the true model is not explicitly represented by a candidate model. Our GMM-DO Bayesian model learning (Section 2)

is thus capable of combining different model field estimates according to their probabilities to obtain a Bayesian mean with high field accuracy. The top four model probabilities all float about 0.25 (uniform distribution), with none showing persistent dominance over the others. In fact, the true model is a linear combination of these four models. The non-monotonic nature of the candidate model probability time-series is attributable to the experiment's various asynchronous quasi-periodic features: vortex shedding downstream of the obstacle (with a period of about 6.0 non-dimensional time units), time-dependent growth factor formulation (with time constant $T_\rho = 0.7$), and temporally-discrete observations (with a frequency of 1 non-dimensional time unit).

The above results indicate that interpolation within the functional space of model formulations is possible. To further refine the Bayesian learning, the inference would need to explicitly account for the possibility of intermediary models represented by varied combinations of growth and decay factors, e.g. as a "super combination" of numerically compatible models as developed in [147].

5. Conclusions

We developed a new methodology for Bayesian learning of stochastic dynamical models and applied it successfully to high-dimensional nonlinear systems. Our GMM-DO Bayesian learning accurately evolves the prior and posterior probabilities of the fields and parameters for each candidate model, given the observations available and their uncertainties, and analytically computes the marginal likelihood of each model, enabling principled Bayesian updates of the model distribution.

We exemplified our results in a range of experiments for two high-dimensional, nonlinear stochastic systems: a flow past an uncertain obstacle and a microorganism concentration under uncertain advection–diffusion–reaction. Even though the uncertainties of the state variable fields were purposely chosen large when compared to the differences among model candidates, our Bayesian GMM-DO learning remained capable of identifying the most probable model formulations, even in cases where a human may be challenged to do the same. We confirmed the convergence of our GMM-GO posterior distributions with the rank of the DO decomposition. We also showed that model biases led to multi-modal posterior distributions that were captured by our methodology. We employed learning metrics to quantify the progress in estimating the true state variables, parameters, and model formulation. When the observations were sufficiently informative about the learning objectives, we found that our posterior model probabilities, analytically marginalized over state variables and parameters, correctly identified either the true model or the most plausible models. The latter indicates that interpolation within the functional space of model formulations is possible.

The present work is a significant step towards the comprehensive adaptive learning of realistic stochastic dynamical systems [57]. First, our work can be extended to variable model formulations within the context of adaptive modeling [55,57]. Given a computational budget, it is sensible to limit computational resources to high probability regions of the model space. Consequently, as learning proceeds with a finite set of candidate models, low probability candidates can be replaced with new candidates that are in the vicinity of high probability candidates. This adaptive modeling process can be performed either automatically or in conjunction with a human subject matter expert [164]. Second, if candidate models are compatible, they can be combined using parameters and model learning is then performed by parameter estimation [165]. Third, in the present learning, the observation locations were assumed fixed. However, observations can be targeted towards the most informative locations for model learning [164]. Integrating adaptive sampling [57,166] with the present Bayesian model learning methodology is another promising direction.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank our MSEAS group members, especially Mr. Abhinav Gupta and Mr. Aman Jalan. We are grateful to the Office of Naval Research for partial support under grant N00014-11-1-0337 (ATL), N00014-14-1-0476 (LEARNS) and N00014-19-1-2693 (IN-BDA) to the Massachusetts Institute of Technology, USA.

References

[1] J. Honerkamp, Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis, Wiley-VCH, 1993.
[2] H. Andrieu, M. French, W. Krajewski, K. Georgakakos, Stochastic-dynamical rainfall simulation based on weather radar volume scan data, Adv. Water Resour. 26 (5) (2003) 581–593.
[3] M. Anghel, K. Werley, A. Motter, Stochastic model for power grid dynamics, in: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, IEEE, 2007, p. 113.
[4] C. Barrett, H. Hunt III, M. Marathe, S. Ravi, D. Rosenkrantz, R. Stearns, Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems, Theoret. Comput. Sci. (2011).
[5] K. Chhak, A. Moore, The North Atlantic Oscillation as a source of stochastic forcing of the wind-driven ocean circulation, Dyn. Atmos. Oceans 43 (3–4) (2007) 151–170.
[6] E. Epstein, Stochastic dynamic prediction, Tellus 21 (6) (1969) 739–759.
[7] T. Govindan, C. Ibarra-Valdez, J. Ruiz de Chávez, A dynamical stochastic coupled model for financial markets, Physica A 381 (2007) 317–328.
[8] C. Huang, D. Ho, J. Lu, J. Kurths, Partial synchronization in stochastic dynamical networks with switching communication channels, Chaos 22 (2) (2012) 023108.
[9] E. Ionides, C. Bretó, A. King, Inference for nonlinear dynamical systems, Proc. Natl. Acad. Sci. 103 (49) (2006) 18438–18443.
[10] P.F.J. Lermusiaux, C.-S. Chiu, G.G. Gawarkiewicz, P. Abbot, A.R. Robinson, R.N. Miller, P.J. Haley Jr., W.G. Leslie, S.J. Majumdar, A. Pang, F. Lekien, Quantifying uncertainties in ocean predictions, Oceanography 19 (1) (2006) 92–105, http://dx.doi.org/10.5670/oceanog.2006.93.
[11] A.R. Robinson, P.F.J. Lermusiaux, Data assimilation for modeling and predicting coupled physical–biological interactions in the sea, in: A.R. Robinson, J.J. McCarthy, B.J. Rothschild (Eds.), Biological-Physical Interactions in the Sea, in: The Sea, vol. 12, John Wiley and Sons, New York, 2002, pp. 475–536, chapter 12.
[12] A. Robinson, P.F.J. Lermusiaux, N.Q. Sloan III, Data assimilation, in: K.H. Brink, A.R. Robinson (Eds.), The Global Coastal Ocean-Processes and Methods, in: The Sea, vol. 10, John Wiley and Sons, New York, 1998, pp. 541–594, chapter 20.
[13] R. Rodriguez, H. Tuckwell, Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks, Phys. Rev. E 54 (5) (1996) 5585.
[14] D. Draper, Assessment and propagation of model uncertainty, J. R. Stat. Soc. Series B (1995) 45–97.
[15] J. Hoeting, D. Madigan, A. Raftery, C. Volinsky, Bayesian model averaging: A tutorial, Statist. Sci. (1999) 382–401.
[16] R. Kass, A. Raftery, Bayes factors, J. Amer. Statist. Assoc. (1995) 773–795.
[17] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85.
[18] J. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press, 1992.
[19] J. Bongard, H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 104 (2007) 9943–9948.
[20] W. Bridewell, P. Langley, L. Todorovski, S. Dzeroski, Inductive process modeling, Mach. Learn. 71 (2008) 1–32.
[21] B. Kouchmeshky, W. Aquino, J. Bongard, H. Lipson, Co-evolutionary algorithm for structural damage identification using minimal physical testing, Internat. J. Numer. Methods Engrg. 69 (5) (2007) 1085–1107.

[22] C. Park, W. Bridewell, P. Langley, Integrated systems for inducing spatio-temporal process models, in: Proceedings of the Twenty-Fourth AAAI Conf. on Artificial Intelligence, Association for the Advancement of Artificial Intelligence, 2010.

[23] A. Watson, I. Parmee, Identification of fluid systems using genetic programming, in: Proceedings of the Second Online Workshop on Evolutionary Computation (WEC2), 1996, pp. 45–48.

[24] C. Wikle, L. Berliner, A Bayesian tutorial for data assimilation, Physica D 230 (1–2) (2007) 1–16.

[25] C. Wikle, M. Hooten, A general science-based framework for dynamical spatio-temporal models, Test 19 (2010) 417–451.

[26] C. Wikle, L. Berliner, R. Milliff, Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions, Mon. Weather Rev. 131 (2003) 1051–1062.

[27] M. Hooten, D. Larsen, C. Wikle, Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model, Landscape Ecol. 18 (2003) 487–502.

[28] L. Berliner, R. Milliff, C. Wikle, Bayesian hierarchical modeling of air-sea interaction, J. Geophys. Res. 108 (2003) 3104–3121.

[29] M. Hooten, C. Wikle, R. Dorazio, J. Royle, Hierarchical spatiotemporal matrix models for characterizing invasions, Biometrics 63 (2007) 558–567.

[30] R. Milliff, A. Bonazzi, C. Wikle, N. Pinardi, L. Berliner, Ocean ensemble forecasting. Part I: Ensemble mediterranean winds for a Bayesian hierarchical model, Q. J. R. Meteorol. Soc. (2011).

[31] K. Rimstad, H. Omre, Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction, Geophysics 75 (2010) 93–108.

[32] Y. Song, C. Wikle, C. Anderson, S. Lack, Bayesian estimation of stochastic parametrizations in a numerical weather forecasting model, Mon. Weather Rev. 135 (2007) 4045–4059.

[33] M. Choi, V. Chandrasekaran, D. Malioutov, J. Johnson, A. Willsky, Multiscale stochastic modeling for tractable inference and data assimilation, Comput. Methods Appl. Mech. Engrg. 197 (2008) 3492–3515.

[34] M. Choi, V. Chandrasekaran, A. Willsky, Gaussian multiresolution models: Exploiting sparse Markov and covariance structure, IEEE Trans. Signal Process. 58 (2010) 1012–1024.

[35] A. Willsky, Multiresolution Markov models for signal and image processing, Proc. IEEE 90 (8) (2002) 1396–1458.

[36] A. Ihler, S. Kirshner, M. Ghil, A. Robertson, P. Smyth, Graphical models for statistical inference and data assimilation, Physica D 230 (1) (2007) 72–87.

[37] P.F.J. Lermusiaux, P.J. Haley, W.G. Leslie, A. Agarwal, O. Logutov, L.J. Burton, Multiscale physical and biological dynamics in the philippine archipelago: Predictions and processes, Oceanography 24 (1) (2011) 70–89, http://dx.doi.org/10.5670/oceanog.2011.05, Special Issue on the Philippine Straits Dynamics Experiment.

[38] Y. Gil, S.A. Pierce, H. Babaie, A. Banerjee, K. Borne, G. Bust, M. Cheatham, I. Ebert-Uphoff, C. Gomes, M. Hill, J. Horel, L. Hsu, J. Kinter, C. Knoblock, D. Krum, V. Kumar, P.F.J. Lermusiaux, Y. Liu, C. North, V. Pankratius, S. Peters, B. Plale, A. Pope, S. Ravela, J. Restrepo, A. Ridley, H. Samet, S. Shekhar, Intelligent systems for geosciences: An essential research agenda, Commun. ACM 62 (1) (2019) 76–84, http://dx.doi.org/10.1145/3192335.

[39] I. Fodor, A survey of dimension reduction techniques, Lawrence Livermore National Laboratory technical report, 2002.

[40] P. Benner, M. Ohlberger, A. Cohen, K. Willcox, Model reduction and approximation: theory and algorithms, SIAM, 2017.

[41] P. Attar, E. Dowell, J. White, J. Thomas, Reduced order nonlinear system identification methodology, AIAA J. 44 (2006) 1895–1904.

[42] T. Braconnier, M. Ferrier, J.-C. Jouhard, M. Montagnac, P. Sagaut, Towards an adaptive POD/SVD surrogate model for aeronautic design, Comput. & Fluids 40 (2011) 195–209.

[43] K. Hall, J. Thomas, E. Dowell, Proper orthogonal decomposition technique for transonic unsteady aerodynamic flows, AIAA J. 38 (10) (2000) 1853–1862.

[44] J. Burkardt, M. Gunzburger, H.-C. Lee, POD and CVT-based reduced-order modeling of Navier-Stokes flows, Comput. Methods Appl. Mech. Engrg. 196 (2006) 337–355.

[45] S. Frolov, A. Baptista, T. Leen, Z. Lu, R. Merwe, Fast data assimilation using a nonlinear Kalman filter and a model surrogate: An application to the Columbia River estuary, Dyn. Atmos. Oceans 48 (2009) 16–45.

[46] R. van der Merwe, T. Leen, Z. Lu, S. Frolov, A. Baptista, Fast neural network surrogates for very high dimensional physics-based models in computational oceanography, Neural Netw. 20 (4) (2007) 462–478.

[47] D. Lucia, P. Beran, W. Silva, Reduced-order modeling: New approaches for computational physics, Prog. Aerosp. Sci. 40 (1) (2004) 51–117.

[48] B. Glaz, L. Liu, P. Friedmann, Reduced-order nonlinear unsteady aerodynamic modeling using a surrogate-based recurrence framework, AIAA J. 48 (10) (2010) 2418.

[49] B. Peherstorfer, K. Willcox, Dynamic data-driven reduced-order models, Comput. Methods Appl. Mech. Engrg. 291 (2015) 21–41.

[50] D. Huynh, D. Knezevic, J. Peterson, A. Patera, High-fidelity real-time simulation on deployed platforms, Comput. & Fluids 43 (1) (2011) 74–81.

[51] K. Logemann, J. Backhaus, I. Harms, SNAC: A statistical emulator of the north-east Atlantic circulation, Ocean Model. 7 (1) (2004) 97–110.

[52] K. Strounine, S. Kravtsov, D. Kondrashov, M. Ghil, Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance, Physica D 239 (3) (2010) 145–166.

[53] P.F.J. Lermusiaux, A.R. Robinson, Data assimilation via error subspace statistical estimation, part I: Theory and schemes, Mon. Weather Rev. 127 (7) (1999) 1385–1407, http://dx.doi.org/10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.

[54] P.F.J. Lermusiaux, Data assimilation via error subspace statistical estimation, part II: Mid-atlantic bight shelfbreak front simulations, and ESSE validation, Mon. Weather Rev. 127 (7) (1999) 1408–1432, http://dx.doi.org/10.1175/1520-0493(1999)127<1408:DAVESS>2.0.CO;2.

[55] P.F.J. Lermusiaux, C. Evangelinos, R. Tian, P.J. Haley Jr., J.J. McCarthy, N.M. Patrikalakis, A.R. Robinson, H. Schmidt, Adaptive coupled physical and biogeochemical ocean predictions: A conceptual basis, in: Computational Science - ICCS 2004, in: Lecture Notes in Computer Science, vol. 3038, Springer Berlin Heidelberg, 2004, pp. 685–692, http://dx.doi.org/10.1007/978-3-540-24688-6_89.

[56] R.C. Tian, P.F.J. Lermusiaux, J.J. McCarthy, A.R. Robinson, A generalized prognostic model of marine biogeochemical-ecosystem dynamics: Structure, parameterization and adaptive modeling, Harvard Reports in Physical/Interdisciplinary Ocean Science 67, Dep. of Earth and Planetary Sciences, Harvard U., Cambridge, MA, 2004.

[57] P.F.J. Lermusiaux, Adaptive modeling, adaptive data assimilation and adaptive sampling, Physica D 230 (1) (2007) 172–196, http://dx.doi.org/10.1016/j.physd.2007.02.014.

[58] T.P. Sapsis, Dynamically Orthogonal Field Equations for Stochastic Fluid Flows and Particle Dynamics (Ph.D. thesis), Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, MA, 2011.

[59] T.P. Sapsis, P.F.J. Lermusiaux, Dynamically orthogonal field equations for continuous stochastic dynamical systems, Physica D 238 (23–24) (2009) 2347–2360, http://dx.doi.org/10.1016/j.physd.2009.09.017.

[60] F. Feppon, P.F.J. Lermusiaux, A geometric approach to dynamical model-order reduction, SIAM J. Matrix Anal. Appl. 39 (1) (2018) 510–538, http://dx.doi.org/10.1137/16M1095202.

[61] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 113 (15) (2016) 3932–3937.

[62] S. Rudy, A. Alla, S.L. Brunton, J.N. Kutz, Data-driven identification of parametric partial differential equations, SIAM J. Appl. Dyn. Syst. 18 (2) (2019) 643–660.

[63] D.A. Messenger, D.M. Bortz, Weak SINDy for partial differential equations, 2020, arXiv preprint arXiv:2007.02848.

[64] C.S. Kulkarni, A. Gupta, P.F.J. Lermusiaux, Sparse regression and adaptive feature generation for the discovery of dynamical systems, in: F. Darema, E. Blasch, S. Ravela, A. Aved (Eds.), Dynamic Data Driven Application Systems. DDDAS 2020., in: Lecture Notes in Computer Science, vol. 12312, Springer, Cham, 2020, pp. 208–216, http://dx.doi.org/10.1007/978-3-030-61725-7_25.

[65] R.K. Niven, A. Mohammad-Djafari, L. Cordier, M. Abel, M. Quade, Bayesian identification of dynamical systems, Multidiscip. Digital Publ. Inst. Proceed. 33 (1) (2020) 33.

[66] G.-J. Both, S. Choudhury, P. Sens, R. Kusters, DeepMoD: Deep learning for model discovery in noisy data, J. Comput. Phys. 428 (2021) 109985.

[67] G.-J. Both, R. Kusters, Sparsely constrained neural networks for model discovery of PDEs, 2020, arXiv preprint arXiv:2011.04336.

[68] H. Xu, H. Chang, D. Zhang, Dl-pde: Deep-learning based data-driven discovery of partial differential equations from discrete and noisy data, 2019, arXiv preprint arXiv:1908.04463.

[69] H. Xu, H. Chang, D. Zhang, DLGA-PDE: Discovery of PDEs with incomplete candidate library via combination of deep learning and genetic algorithm, J. Comput. Phys. 418 (2020) 109584.

[70] Z. Chen, Y. Liu, H. Sun, Deep learning of physical laws from scarce data, 2020, arXiv preprint arXiv:2005.03448.

[71] M. Maslyaev, A. Hvatov, A. Kalyuzhnaya, Data-driven partial derivative equations discovery with evolutionary approach, in: International Conference on Computational Science, Springer, 2019, pp. 635–641.

[72] M. Bassenne, A. Lozano-Durán, Computational model discovery with reinforcement learning, 2019, arXiv preprint arXiv:2001.00008.

[73] G. Novati, H.L. de Laroussilhe, P. Koumoutsakos, Automating turbulence modelling by multi-agent reinforcement learning, Nat. Mach. Intell. 3 (1) (2021) 87–96.

[74] Y. Wang, Z. Shen, Z. Long, B. Dong, Learning to discretize: solving 1D scalar conservation laws via deep reinforcement learning, 2019, arXiv preprint arXiv:1905.11079.

[75] M. Raissi, G.E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, J. Comput. Phys. 357 (2018) 125–141.
[76] Z. Long, Y. Lu, X. Ma, B. Dong, Pde-net: Learning pdes from data, in: International Conference on Machine Learning, PMLR, 2018, pp. 3208–3216.
[77] Z. Long, Y. Lu, B. Dong, Pde-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network, J. Comput. Phys. 399 (2019) 108925.
[78] R. Swischuk, L. Mainini, B. Peherstorfer, K. Willcox, Projection-based model reduction: Formulations for physics-based machine learning, Comput. & Fluids 179 (2019) 704–717.
[79] M. Qraitem, D. Kularatne, E. Forgoston, M.A. Hsieh, Bridging the gap: Machine learning to resolve improperly modeled dynamics, Physica D 414 (2020) 132736.
[80] A. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski, Using Bayesian model averaging to calibrate forecast ensembles, Mon. Weather Rev. 133 (5) (2005) 1155–1174.
[81] J. Sloughter, T. Gneiting, A. Raftery, Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, J. Amer. Statist. Assoc. 105 (489) (2010) 25–35.
[82] Y. Bar-Shalom, X. Li, T. Kirubarajan, J. Wiley, Estimation with Applications to Tracking and Navigation, Wiley-Interscience, 2001.
[83] P. Maybeck, Stochastic Models, Estimation and Control, Vol. 1, Academic Press, 1979.
[84] O.G. Logutov, A.R. Robinson, Multi-model fusion and error parameter estimation, Q. J. R. Meteorol. Soc. 131 (613) (2005) 3397–3408, http://dx.doi.org/10.1256/qj.05.99.
[85] T. Gneiting, A. Raftery, Weather forecasting with ensemble methods, Science 310 (5746) (2005) 248–249.
[86] T. Krishnamurti, C. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, S. Surendran, Multimodel ensemble forecasts for weather and seasonal climate, J. Clim. 13 (23) (2000) 4196–4216.
[87] M. Rixen, E. Ferreira-Coelho, Operational prediction of acoustic properties in the ocean using multi-model statistics, Ocean Model. 11 (3) (2006) 428–440.
[88] P.G.Y. Lu, Bayesian inference of stochastic dynamical models (Master's thesis), Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts, 2013.
[89] T. Bayes, An essay towards solving a problem in the doctrine of chances. by the late rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S., Phil. Trans. 53 (1763) 370–418.
[90] R. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (1) (1960) 35–45.
[91] R. Kalman, R. Bucy, New results in linear filtering and prediction theory, J. Basic Eng. 83 (1961) 95–108.
[92] M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process. 50 (2) (2002) 174–188.
[93] C. Andrieu, N. De Freitas, A. Doucet, M. Jordan, An introduction to MCMC for machine learning, Mach. Learn. 50 (1) (2003) 5–43.
[94] P. Fearnhead, Computational methods for complex stochastic systems: A review of some alternatives to MCMC, Stat. Comput. 18 (2) (2008) 151–171.
[95] A. Gelb, Applied Optimal Estimation, MIT Press, 1974.
[96] A. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.
[97] D. Simon, Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches, Wiley-Interscience, 2006.
[98] J. Lions, Optimal Control of Systems Governed by Partial Differential Equations, Springer Verlag, 1971.
[99] F. Le Dimet, O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations, Tellus 38A (1986) 97–110.
[100] A. Bennett, Inverse Methods in Physical Oceanography, Cambridge University Press, 1992.
[101] E. Kalnay, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.
[102] W. Lahoz, B. Khattatov, R. Menard, Data Assimilation: Making Sense of Observations, Springer, 2010.
[103] P. Malanotte-Rizzoli, Modern Approaches to Data Assimilation in Ocean Modeling, Elsevier Oceanography Series, 1996.
[104] S. Park, L. Xu, Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications, Springer, 2009.
[105] C. Wunsch, The Ocean Circulation Inverse Problem, Cambridge University Press, 1996.
[106] R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, Contemp. Phys. 49 (2) (2008) 71–104.
[107] F. Auclair, P. Marsaleix, P. De Mey, Space-time structure and dynamics of the forecast error in a coastal circulation model of the Gulf of Lions, Dyn. Atmos. Oceans 36 (2003) 309–346.
[108] D. Dee, A. Da Silva, The choice of variable for atmospheric moisture analysis, Mon. Weather Rev. 131 (2003) 155–171.
[109] P.F.J. Lermusiaux, C.-S. Chiu, A.R. Robinson, Modeling uncertainties in the prediction of the acoustic wavefield in a shelfbreak environment, in: E.-C. Shang, Q. Li, T.F. Gao (Eds.), Proc. of the 5th Int. Conf. on Theoretical and Computational Acoustics, World Scientific Pub. Co., 2002, pp. 191–200, http://dx.doi.org/10.1142/9789812777362_0020.
[110] N.R. Council, Statistics and Physical Oceanography, The National Academies Press, 1993.
[111] A. Doucet, N. de Freitas, N. Gordon, Sequential Monte-Carlo Methods in Practice, Springer-Verlag, 2001.
[112] R. Srinivasan, Importance Sampling: Applications in Communications and Detection, Springer, 2002.
[113] X. Meng, W. Wong, Simulating ratios of normalizing constants via a simple identity: A theoretical exploration, Statist. Sinica 6 (1996) 831–860.
[114] A. Gelman, X. Meng, Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, Statist. Sci. (1998) 163–185.
[115] R. Neal, Annealed importance sampling, Stat. Comput. 11 (2) (2001) 125–139.
[116] N. Chopin, C. Robert, Properties of nested sampling, Biometrika 97 (3) (2010) 741–755.
[117] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. 1 (4) (2006) 833–860.
[118] T. Sondergaard, P.F.J. Lermusiaux, Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme, Mon. Weather Rev. 141 (6) (2013) 1737–1760, http://dx.doi.org/10.1175/MWR-D-11-00295.1.
[119] T. Sondergaard, P.F.J. Lermusiaux, Data assimilation with Gaussian mixture models using the dynamically orthogonal field equations. part II: Applications, Mon. Weather Rev. 141 (6) (2013) 1761–1785, http://dx.doi.org/10.1175/MWR-D-11-00296.1.
[120] T.P. Sapsis, P.F.J. Lermusiaux, Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty, Physica D 241 (1) (2012) 60–76, http://dx.doi.org/10.1016/j.physd.2011.10.001.
[121] F. Feppon, P.F.J. Lermusiaux, Dynamically orthogonal numerical schemes for efficient stochastic advection and Lagrangian transport, SIAM Rev. 60 (3) (2018) 595–625, http://dx.doi.org/10.1137/16M1109394.
[122] M. Loeve, Probability Theory II, Springer, 1978.
[123] M.P. Ueckermann, P.F.J. Lermusiaux, T.P. Sapsis, Numerical schemes for dynamically orthogonal equations of stochastic fluid and ocean flows, J. Comput. Phys. 233 (2013) 272–294, http://dx.doi.org/10.1016/j.jcp.2012.08.041.
[124] J. Bilmes, A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, International Computer Science Institute, 1998.
[125] P. Stoica, Y. Selén, Model-order selection: A review of information criterion rules, IEEE Signal Process. Mag. 21 (4) (2004) 36–47.
[126] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley-Interscience, 2001.
[127] M. Ghil, P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, Adv. Geophys. 33 (1991) 141–266.
[128] D. Subramani, P.F.J. Lermusiaux, Probabilistic ocean predictions with dynamically-orthogonal primitive equations, 2021, in preparation.
[129] P.F.J. Lermusiaux, On the mapping of multivariate geophysical fields: Sensitivities to size, scales, and dynamics, J. Atmos. Oceanic Techn. 19 (10) (2002) 1602–1637, http://dx.doi.org/10.1175/1520-0426(2002)019<1602:OTMOMG>2.0.CO;2.
[130] P.F.J. Lermusiaux, D.G.M. Anderson, C.J. Lozano, On the mapping of multivariate geophysical fields: Error and variability subspace estimates, Quart. J. R. Meteor. Soc. 126 (565) (2000) 1387–1429, http://dx.doi.org/10.1256/smsqj.56509.
[131] K.A. Gkirgkis, Stochastic Ocean Forecasting with the Dynamically Orthogonal Primitive Equations (Master's thesis), Massachusetts Institute of Technology, Mechanical Engineering, Cambridge, Massachusetts, 2021.
[132] S. Dennis, G. Chang, Numerical solutions for steady flow past a circular cylinder at Reynolds numbers up to 100, J. Fluid Mech. 42 (3) (1970) 471–489.
[133] B. Fornberg, A numerical study of steady viscous flow past a circular cylinder, J. Fluid Mech. 98 (4) (1980) 819–855.
[134] B. Fornberg, Steady viscous flow past a circular cylinder up to Reynolds number 600, J. Comput. Phys. 61 (2) (1985) 297–320.
[135] F. Smith, Laminar flow of an incompressible fluid past a bluff body: The separation, reattachment, eddy properties and drag, J. Fluid Mech. 92 (1) (1979) 171–206.
[136] D. Tritton, Experiments on the flow past a circular cylinder at low Reynolds numbers, J. Fluid Mech. 6 (4) (1959) 547–567.
[137] M. Zdravkovich, Flow Around Circular Cylinders Volume 1: Fundamentals, Cambridge University Press, 1997.
[138] P. Kundu, I. Cohen, Fluid Mechanics, Elsevier, 2004.

[139] B. Sumer, J. Fredsøe, Hydrodynamics Around Cylindrical Structures, World Scientific Publishing Company, 2006.
[140] R. Chhabra, A. Soares, J. Ferreira, Steady non-Newtonian flow past a circular cylinder: A numerical study, Acta Mech. 172 (1) (2004), 1–16.
[141] S. Mittal, B. Kumar, Flow past a rotating cylinder, J. Fluid Mech. 476 (2003) 303–334.
[142] H. Blackburn, R. Henderson, A study of two-dimensional flow past an oscillating cylinder, J. Fluid Mech. 385 (1) (1999) 255–286.
[143] C. Jackson, A finite-element study of the onset of vortex shedding in flow past variously shaped bodies, J. Fluid Mech. 182 (1) (1987) 23–45.
[144] J. Lin, Y. Yang, D. Rockwell, Flow past two cylinders in tandem: Instantaneous and averaged flow structure, J. Fluids Struct. 16 (8) (2002) 1059–1071.
[145] B. Fornberg, Steady viscous flow past a sphere at high Reynolds numbers, J. Fluid Mech. 190 (1988) 471–489.
[146] M. Zdravkovich, Flow Around Circular Cylinders Volume 2: Applications, Cambridge University Press, 2002.
[147] P.G.Y. Lu, P.F.J. Lermusiaux, PDE-based Bayesian inference of high-dimensional dynamical models, MSEAS Report 19, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA, 2014.
[148] M. Kopera, Non-dimensionalization of the Navier-Stokes equations, University of Warwick, Centre for Scientific Computing, 2008.
[149] M.P. Ueckermann, P.F.J. Lermusiaux, 2.29 Finite Volume MATLAB Framework Documentation, MSEAS Report 14, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 2012, URL http://mseas.mit.edu/?p=2567.
[150] G.J. Sheard, M.C. Thompson, K. Hourigan, From spheres to circular cylinders: the stability and flow structures of bluff ring wakes, J. Fluid Mech. 492 (2003) 147.
[151] I. Kim, X. Wu, Unified strouhal-Reynolds number relationship for laminar vortex streets generated by different-shaped obstacles, Phys. Rev. E 92 (4) (2015) 043011.
[152] M. Behrenfeld, R. O'Malley, D. Siegel, C. McClain, J. Sarmiento, G. Feldman, A. Milligan, P. Falkowski, R. Letelier, E. Boss, Climate-driven trends in contemporary ocean productivity, Nature 444 (7120) (2006), 752–755.
[153] C. Field, M. Behrenfeld, J. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components, Science 281 (5374) (1998) 237–240.
[154] V. Saba, M. Friedrichs, D. Antoine, R. Armstrong, I. Asanuma, M. Behrenfeld, A. Ciotti, M. Dowell, N. Hoepffner, K. Hyde, et al., An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe, Biogeosciences 8 (2) (2011) 489–503.
[155] W. Fennel, T. Neumann, Introduction to the Modelling of Marine Ecosystems, Elsevier Science, 2004.
[156] E. Hofmann, M. Friedrichs, Predictive modeling for marine ecosystems, Sea 12 (2002) 537–565.
[157] E. Hofmann, C. Lascara, Overview of interdisciplinary modeling for marine ecosystems, Sea 10 (1998) 507–540.
[158] C. Lalli, T. Parsons, Biological Oceanography: An Introduction, Butterworth-Heinemann, 1997.
[159] M.P. Ueckermann, P.F.J. Lermusiaux, High order schemes for 2D unsteady biogeochemical ocean models, Ocean Dyn. 60 (6) (2010) 1415–1445, http://dx.doi.org/10.1007/s10236-010-0351-x.
[160] R. Tian, P. Lermusiaux, J. McCarthy, A. Robinson, A generalized prognostic model of marine biogeochemical-ecosystem dynamics: Structure, parameterization and adaptive modeling, Harvard Reports in Physical/Interdisciplinary Ocean Science #67, 2004.
[161] K.J. Heywood, E.D. Barton, J.H. Simpson, The effects of flow disturbance by an oceanic island, J. Mar. Res. 48 (1) (1990) 55–73.
[162] P. Lermusiaux, M. Ueckermann, T. Lolla, Linearization for the Dynamically Orthogonal Evolution Equations, MIT Multidisciplinary Simulation, Estimation, and Assimilation Systems Group Report, 2013.
[163] D.N. Subramani, P.F.J. Lermusiaux, Energy-optimal path planning by stochastic dynamically orthogonal level-set optimization, Ocean Model. 100 (2016) 57–77, http://dx.doi.org/10.1016/j.ocemod.2016.01.006.
[164] P.F.J. Lermusiaux, D.N. Subramani, J. Lin, C.S. Kulkarni, A. Gupta, A. Dutt, T. Lolla, P.J. Haley Jr., W.H. Ali, C. Mirabito, S. Jana, A future for intelligent autonomous ocean observing systems, J. Mar. Res. 75 (6) (2017) 765–813, http://dx.doi.org/10.1357/002224017823524035, The Sea. Volume 17, The Science of Ocean Prediction, Part 2..
[165] A. Gupta, P.J. Haley, D.N. Subramani, P.F.J. Lermusiaux, Fish modeling and Bayesian learning for the Lakshadweep islands, in: OCEANS 2019 MTS/IEEE SEATTLE, IEEE, Seattle, 2019, pp. 1–10, http://dx.doi.org/10.23919/OCEANS40490.2019.8962892.
[166] K.D. Heaney, G. Gawarkiewicz, T.F. Duda, P.F.J. Lermusiaux, Nonlinear optimization of autonomous undersea vehicle sampling strategies for oceanographic data-assimilation, J. Field Robotics 24 (6) (2007) 437–448, http://dx.doi.org/10.1002/rob.20183.