Bayesian Learning of Coupled Biogeochemical-Physical Models

Abhinav Gupta[a,b], Pierre F. J. Lermusiaux[a,b,*]

[a] *Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA - 02139*
[b] *Center for Computational Science and Engineering, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA - 02139*

Abstract

Predictive models for marine ecosystems are used for a variety of needs. Due to sparse measurements and limited understanding of the myriad of ocean processes, there is however uncertainty. There is model uncertainty in the parameter values, functional forms with diverse parameterizations, level of complexity needed, and thus in the state fields. We develop a principled Bayesian model learning methodology that allows interpolation in the space of candidate models and discovery of new models, all while estimating state fields and parameter values, as well as the joint probability distributions of all learned quantities. We address the challenges of high-dimensional and multidisciplinary dynamics governed by partial differential equations (PDEs) by using state augmentation and the computationally efficient Gaussian Mixture Model - Dynamically Orthogonal filter. Our innovations include special stochastic parameters to unify candidate models into a single general model and stochastic piecewise function approximations to generate dense candidate model spaces. They allow handling many candidate models, possibly none of which are accurate, and learning elusive unknown functional forms in compatible and embedded models. Our new methodology is generalizable and interpretable and extrapolates out of the space of models to discover new ones. We perform a series of twin experiments based on flows past a seamount coupled with three-to-five component ecosystem models, including flows with chaotic advection. We quantify learning skills, and evaluate convergence and sensitivity to hyper-parameters. Our PDE framework successfully discriminates among model candidates, learns in the absence of prior knowledge by searching in dense function spaces, and updates joint probabilities while capturing non-Gaussian statistics. The parameter values and model formulations that best explain the data are identified.

Keywords: Dynamical systems, Data assimilation, Uncertainty quantification, Gaussian Mixture Models, Dynamically Orthogonal, Model learning, Bayesian, Stochastic PDEs, Ocean and weather prediction

1. Introduction

The ability to predict and understand marine ecosystems is essential for addressing many of the grand challenges faced by humanity, such as climate change, food security, and sustainability. In broad terms, marine ecosystems can be seen as food webs, or flow of food/energy from nutrients to phytoplanktons, to zooplanktons, to fish, and finally recycling back to the nutrients [1; 2]. However, there does not yet exist a single generic model that accurately represents all the components in marine food webs due to the presence of highly complex biological processes with many unknown interactions. Therefore, many approximations are made and only parts of a food web are commonly modeled. The interactions of what is modeled with other portions of the food web then need to be parameterized. In addition, biology is forced by complex nonlinear physics. Most biogeochemical-physical modeling systems thus broadly categorize the nutrients and individual species, representing them as continuous state variable fields, defined as concentrations of

*Corresponding author
Email addresses: guptaa@mit.edu (Abhinav Gupta), pierrel@mit.edu (Pierre F. J. Lermusiaux)

nutrients, biomass, or number of organisms per unit volume of water. The dynamics of these fields consists of reaction terms representing biogeochemical processes such as nutrient uptake, grazing, death, etc., and of forcing by physical processes such as advection, diffusion, and sun light.

A plethora of biogeochemical models have been proposed. They differ in their complexity, or ability to resolve different biological processes. Models of higher complexity have more biological components, functional terms, and parameters. However, process terms and parameters are often poorly known, which hampers the utility of highly complex models [3; 4; 5]. The simplest and most popular models are 3-component nutrient-phytoplankton-zooplankton (NPZ) models [6; 7]. NPZ models are easily understood and serve an important role in ocean research. Including the intermediate state of detritus leads to four component NPZ-Detritus biological models [8]. Intermediate complexity models involve around 7 to 10 components, adding bacteria, nitrate, ammonium, and dissolved organic nitrogen [9], or related state variables [10]. One of the most complex lower-trophic-level marine ecosystem models is the European Regional Seas Ecosystem Model (ERSEM, [11; 12; 13]), originally developed for the North Sea. Many choices of functional forms exist for each of the biological processes [3], leading to application-specific variants of the above models.

Biogeochemical models are commonly developed semi-empirically, leading to uncertainty in their parameters, functional forms, and level of complexity. What is adequate for a particular ocean region may not work elsewhere or may need to be updated due to seasonal or other variabilities [14]. Such model uncertainties transfer to the state variables predicted, complicating model learning by direct comparisons of state variables with in situ data. As a result, when observations are employed to develop models, it is often in an offline mode, fitting parameter values or functional forms to data in controlled experiments. With data assimilation, we could however use observations in a direct Bayesian sense, to jointly learn state variables, parameter values, and discriminate/discover functional forms with quantifiable uncertainty [15]. Most biogeochemical data assimilation [16; 17] can be categorized broadly into two categories. The first is parameter estimation, where model parameters are calibrated by minimizing misfits between model output fields and independent observations [18; 19; 4; 20]. The second is sequential estimation, where observations collected are used to update model states during the forward model integration [10; 21; 22; 23]. However, very few studies deal with the simultaneous estimation of parameters, state variables, and model equations. Doron et al. [24] used a Monte Carlo ensemble of 200 simulations lasting 30-days in the North Atlantic and conducted idealized twin experiments with surface observations of phytoplankton to estimate parameters and states with a Kalman filter-based scheme and state augmentation. Similarly, Jones et al. [25] performed state and parameter estimation in a nonlinear phytoplankton-zooplankton model using two Markov Chain Monte Carlo (MCMC) algorithms in an identical-twin setting. Lately, along with state and parameter estimation, the selection of optimal complexity of biogeochemical models has become a new area of research [15; 4; 26; 27]. Because of the multiscale and intermittent variability of marine ecosystems, there is also a need for generalized and adaptive modeling, where models can adapt during run-time [14; 28; 29; 30].

Several machine learning methods have been developed for the discovery of model equations. The sparse regression-based methods (SINDy; [31; 32]) are promising as they do not require prior knowledge, however, they often require large data sets. Variations of SINDy include weak SINDy to learn PDEs [33], adaptive generation of features to increase the library of models [34], and extensions to Bayesian identification [35]. Deep learning methods have been derived to obtain marine ecosystem closure models [36]. Genetic algorithms [37] and reinforcement learning [38; 39; 40] have been used to search the space of candidate models. However, most of these approaches do not provide uncertainty estimates for the discovered models. Methods have also combined prior knowledge about underlying governing equations for model recovery and refinement. For example, Raissi and Karniadakis [41] used Gaussian processes to learn the values of the parametric response of partially-known nonlinear differential equations. Unfortunately, data and knowledge of governing laws are luxuries in the case of realistic biogeochemical models.

It is clear that a direct principled way of identifying models that best explain the data is much needed. The Bayesian theory and schemes of Lu and Lermusiaux [42; 43] address several of the above needs and drawbacks, using sparse observations for joint Bayesian inference of states and parameters along with probabilistic discrimination among candidate models. However, several questions remain: Could we avoid assimilating observations independently in each candidate model when there are so many models to choose from? And if none of these models are that accurate, could the Bayesian machine find the elusive true formulations? Could

it interpolate within and extrapolate out of known model spaces, while providing accurate joint probability distributions for model states, parameters, and formulations? Could such Bayesian learning be efficient and accurate with high-dimensional and multidisciplinary physical-biogeochemical stochastic PDEs? The overall goal of the present paper is thus to extend and generalize the discrimination-based model learning developed in [42; 43] to allow for interpolation in the space of candidate models and discovery of new models, in an efficient fashion. Our novel learning and discovery of differential models is achieved by introducing special stochastic parameters and stochastic piecewise function approximations. We address the challenges of multidisciplinary dynamics and develop a rigorous PDE Bayesian learning framework using state augmentation and the Gaussian Mixture Model - Dynamically Orthogonal (GMM-DO) filter [44; 45]. The final estimates are notably joint probability distributions for all learned quantities. In Sect. 2, we present the problem statement. In Sect. 3, we develop the general Bayesian learning methodology with special parameters for model learning and discovery. In Sect. 4, we describe the stochastic biogeochemical-physical equations and simulated experiments. In Sect. 5, we apply our methodology to four sets of experiments of varying complexities and learning objectives. Conclusions are provided in Sect. 6.

2. Problem Statement

A single mathematical model that exactly captures all the physical and biological processes occurring in the real world does not yet exist. Hence, there is inherent model uncertainty that manifests in many forms, including: initial and boundary condition uncertainties; unreliable parameter values; multiple competing candidate model functions; unknown functional forms; missing model terms; and, debatable complexity of the model. In this work, we consider discriminating among candidate models, learning among compatible models, and discovering new model formulations. Compatible models are models that can be related to a single dynamical system theoretically and that can also be combined numerically. Compatible models can nonetheless represent different dynamics, e.g., our goals include learning which dynamics are or are not present based on observations.

In general, we consider a stochastic dynamical modeling system defined on a domain $\mathcal{D}$, governing the dynamics of $\boldsymbol{u}(\boldsymbol{x},t;\omega) : \mathbb{R}^n \times [0,T] \to \mathbb{R}^{N_v}$, the stochastic state vector comprising N_v dynamical state variable fields. The realization index ω belongs to a measurable sample space Ω and the model depends on a vector $\boldsymbol{\theta}(\omega)$ of N_θ uncertain parameters. To encompass the majority of scenarios, we write the general form of the uncertain dynamical modeling system as follows,

$$\begin{aligned} \frac{\partial \boldsymbol{u}(\boldsymbol{x},t;\omega)}{\partial t} &= \mathcal{L}[\boldsymbol{u}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{x},t] + \widehat{\mathcal{L}}[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega] + \widetilde{\mathcal{L}}[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega]\,, \\ &\qquad\qquad \boldsymbol{x}\in\mathcal{D},\ t\in[0,T],\ \omega\in\Omega\,, \\ \text{with}\quad & \boldsymbol{u}(\boldsymbol{x},0;\omega) = \boldsymbol{u}_o(\boldsymbol{x};\omega)\,, \\ \text{and}\quad & \mathcal{B}[\boldsymbol{u}(\boldsymbol{x},t;\omega)] = \boldsymbol{b}(\boldsymbol{x},t;\omega),\ \boldsymbol{x}\in\partial\mathcal{D},\ t\in[0,T],\ \omega\in\Omega\,, \end{aligned} \tag{1}$$

where $\boldsymbol{u}_o(\boldsymbol{x};\omega)$, $\mathcal{B}$, and $\boldsymbol{b}(\boldsymbol{x},t;\omega)$ are the stochastic initial conditions, boundary condition operators, and boundary values respectively. The functional form of the first dynamics term $\mathcal{L}[\boldsymbol{u}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{x},t]$ is assumed to be known, but with uncertain parameters. The second term $\widehat{\mathcal{L}}[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega] \in \{\widehat{\mathcal{L}}_1[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega],...,\widehat{\mathcal{L}}_{N_m}[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega]\}$, represents a set of compatible candidate functional forms, where N_m is the number of candidates. For example, for reaction terms, model functions are often from the polynomial, exponential, and/or sinusoidal families, and can be rational or irrational functions. The third term $\widetilde{\mathcal{L}}[\boldsymbol{u}(\boldsymbol{x},t;\omega);\omega]$ has a functional form completely unknown. The stochastic initial and boundary condition formulations can also have uncertain function forms, similar to the dynamical modeling system itself, e.g. they can be known, belonging to a family, or unknown.

In some cases, candidate models have different complexities,

$$\mathcal{M}_i : \begin{cases} \frac{\partial u_1^i(\boldsymbol{x},t;\omega)}{\partial t} = & \mathcal{L}_1^i[u_1^i(\boldsymbol{x},t;\omega),...,u_{N_v(i)}^i(\boldsymbol{x},t;\omega),\boldsymbol{\theta}^i(\omega),\boldsymbol{x},t;\omega] \\ & \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \; i=1,...,N_m \\ \frac{\partial u_{N_v(i)}^i(\boldsymbol{x},t;\omega)}{\partial t} = & \mathcal{L}_{N_v(i)}^i[u_1^i(\boldsymbol{x},t;\omega),...,u_{N_v(i)}^i(\boldsymbol{x},t;\omega),\boldsymbol{\theta}^i(\omega),\boldsymbol{x},t;\omega] \end{cases} \tag{2}$$

where each model, $\mathcal{M}_i$, has $N_v(i)$ number of state variables ($\{u_1^i,...,u_{N_v(i)}^i\}$) from a pool of candidates, and their aggregates. In such situations, the candidate models can often remain compatible with each other, for example low complexity models are embedded in higher complexity ones. We refer to such classes of candidate models as, *compatible-embedded models*. Of course, in general, uncertainty in parameter values, functional forms, and complexities occur simultaneously.

Let $\boldsymbol{U}(t;\omega) \in \mathbb{R}^{N_v N_x}$ denote the spatially discretized state vector of the continuous field $\boldsymbol{u}(\boldsymbol{x},t;\omega)$. where N_x denotes the dimension of the discretized state space. Next, we assume that the observations ($\boldsymbol{\mathcal{Y}}(t;\omega)$) are indirect, noisy, and related to $\boldsymbol{U}(t;\omega)$ according to the linear model from the state to the data space,

$$\boldsymbol{\mathcal{Y}}(t;\omega) = \boldsymbol{H}\boldsymbol{U}(t;\omega) + \boldsymbol{V}(t;\omega), \qquad \boldsymbol{V}(t;\omega) \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{R}) \tag{3}$$

where N_y is the number of available observations; $\boldsymbol{H} \in \mathbb{R}^{N_y \times N_v N_x}$ the observation matrix; and $\boldsymbol{V} \in \mathbb{R}^{N_y}$ a zero-mean, uncorrelated Gaussian measurement noise with covariance matrix $\boldsymbol{R} \in \mathbb{R}^{N_y \times N_y}$. Observations are assumed to be available only at discrete time-instants, t_k for $k = 1, 2, ..., K$.

In summary, our specific objectives are thus two-folds, first to solve the stochastic forward-modeling system (Eqs. 1 & 2), taking into account all the associated uncertainties including compatible, compatible-embedded, and unknown model terms; and second to simultaneously learn, in the Bayesian sense, the state fields, parameters, and model equations based on the observation model (Eq. 3). Our Bayesian learning thus need to evolve the prior and posterior joint probabilities of state fields, parameters, and model formulations, given the observations available and all uncertainties. The overall goal is to accurately represent these probability density functions (pdfs), including the marginal probabilities of known, uncertain, and unknown model formulations. It is only if the observations are sufficiently informative about either the state fields, parameters, and/or model formulations, that the Bayesian machine will identify the true state variables, true parameters, and/or true model. If the observations are not sufficiently informative, the perfect Bayesian machine will not lead to a unique identification, but provide the exact posterior probabilities of the models, parameter values and/or state variable fields.

3. General Bayesian Learning Methodology

In this work, we start from Bayesian learning for rigorous discrimination among candidate models [42; 43]. Each candidate model then evolves the joint pdf of its state variables and parameters, independently from other models, and provides probability distributions that are conditional on the candidate model. When observations are made, both the model-conditional state variables and parameters, and the model pdfs are updated using Bayes' rules [46],

$$\begin{aligned} p_{\boldsymbol{U}|\boldsymbol{\mathcal{Y}},\mathcal{M}}(\boldsymbol{U}|\boldsymbol{y},\mathcal{M}_i) &= \frac{p_{\boldsymbol{\mathcal{Y}}|\boldsymbol{U},\mathcal{M}}(\boldsymbol{y}|\boldsymbol{U},\mathcal{M}_i)}{p_{\boldsymbol{\mathcal{Y}}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_i)} p_{\boldsymbol{U}|\mathcal{M}}(\boldsymbol{U}|\mathcal{M}_i) \;, \quad \forall\, \boldsymbol{U} \in \mathbb{R}^{N_v N_x}, \forall\, i \in \{1,...,N_m\} \;, \\ p_{\mathcal{M}|\boldsymbol{\mathcal{Y}}}(\mathcal{M}_n|\boldsymbol{y}) &= \frac{p_{\boldsymbol{\mathcal{Y}}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_i)}{p_{\boldsymbol{\mathcal{Y}}}(\boldsymbol{y})} p_{\mathcal{M}}(\mathcal{M}_i) \;, \quad \forall\, i \in \{1,...,N_m\} \;, \end{aligned} \tag{4}$$

where $\mathcal{M}_i$ is the i^{th} model candidate and the distributions $p_{\boldsymbol{U}|\mathcal{M}}(\boldsymbol{U}|\mathcal{M}_i)$ and $p_{\boldsymbol{U}|\boldsymbol{\mathcal{Y}},\mathcal{M}}(\boldsymbol{U}|\boldsymbol{y},\mathcal{M}_i)$ are the prior and posterior model-conditional state variable distributions, respectively. The model distribution $p_{\mathcal{M}}(\bullet)$ is the prior probability for each of the candidates being the true model and $p_{\mathcal{M}|\boldsymbol{\mathcal{Y}}}(\bullet|\boldsymbol{y})$ is the corresponding posterior model distribution. This pdf $p_{\mathcal{M}|\boldsymbol{\mathcal{Y}}}(\bullet|\boldsymbol{y})$ allows learning by exact Bayesian discrimination among candidate models. In particular, when observations are not sufficient to achieve unequivocally the ultimate

learning objective, this posterior pdfs will correctly represent the ambiguity including possible multimodal distributions and the effects of biases in the candidate models [42; 43].

The above Bayesian learning evolves each stochastic candidate model separately. To increase efficiency, this can be circumvented, for example, when models are compatible or compatible-embedded. Next, we thus develop new stochastic parameterizations that unify all such candidate models into a single general modeling system. We recast the model learning into new parameter estimation problems, using special stochastic parameters (Sect. 3.1) and stochastic piece-wise function approximation theory (Sect. 3.2). We then evolve the joint probabilities of the state fields, the regular parameters, and these special parameters, using new stochastic DO equations (Sect. 3.3). At each observation time, we perform Bayesian learning using the GMM-DO filter with state augmentation (Sect. 3.3). Our methodology does not need to compute the discrete marginal likelihoods, $p_{\boldsymbol{y}|\mathcal{M}}(\boldsymbol{y}|\mathcal{M}_i)$; instead, it learns in a parameterized continuous model space. We thus extend learning among discrete model formulations to learning within a continuous infinite range of formulations as well as across models of different complexities and into unknown models. In other words, we remain able to discriminate among existing models, but we can now also interpolate in or extrapolate out of the space of models to discover new ones.

3.1. Special Stochastic Parameters: Compatible and Compatible-embedded Models

Let us first consider the case where, when according to prior scientific knowledge, the uncertain model belongs to a set of compatible candidate functional forms ($\hat{\mathcal{L}}[\bullet]$; Eq. 1). In order to recast this learning problem with multiple models into a learning problem with a single model and parameter estimation, the compatible candidate model functions are added to each other but only after being multiplied with novel stochastic parameters. Each of the candidates is thus assigned a special stochastic parameter that can take discrete or continuous values depending on the learning objectives and prior knowledge. For example, binary values would be utilized to discriminate between presence or absence of certain functions, while other values would be utilized to allow some linear interpolation within the space defined by the compatible candidate models. To complete Bayesian learning, when observations are collected, the probability distributions of these uncertain special parameters ($\alpha_k(\omega)$'s, $k = 1, ..., N_m$) are updated and their mean values estimated alongside these of other regular parameters ($\boldsymbol{\theta}(t;\omega)$), using state augmentation. Summarizing, the general model can thus be written as a stochastic linear combination of the candidates,

$$\hat{\mathcal{L}}[\boldsymbol{u}(\boldsymbol{x},t;\omega),t;\omega] = \sum_{k=1}^{N_m} \widehat{\alpha_k(\omega)\mathcal{L}_k}[\boldsymbol{u}(\boldsymbol{x},t;\omega),\boldsymbol{x},t;\omega] \ . \tag{5}$$

where the distribution of the $\alpha_k(\omega)$ is updated at each observation time. This new formulation can thus both help select active candidate functions and identify their linear combinations. It allows interpolating in the space of known candidate functions.

Next, we extend this approach to learn model complexity (Eq. 2). This is achieved by defining new states multiplied with special stochastic parameters, $\boldsymbol{u}'_k = \beta_k(\omega)\boldsymbol{u}_k$, and a general model, $\mathcal{L}'_k$, which encompasses all the candidates in the class of compatible-embedded models,

$$\frac{\partial \boldsymbol{u}'_k(\boldsymbol{x},t;\omega)}{\partial t} = \mathcal{L}'_k[\boldsymbol{u}'_1(\boldsymbol{x},t;\omega), ..., \boldsymbol{u}'_{N_v}(\boldsymbol{x},t;\omega), \boldsymbol{\theta}(t;\omega), \boldsymbol{\beta}(\omega), \boldsymbol{x}, t;\omega], \qquad k = 1, ..., N_v \tag{6}$$

where $N_v = \max\{N_v(i)\}_{i=1}^{N_m}$. By learning these special parameters, we can eliminate certain state variables or aggregate them to form new states, and determine the model of appropriate complexity that best explains the observed data. To illustrate such combinations of compatible-embedded models into a general model, let us consider a case with only two candidate models ($N_m = 2$ in Eq. 2). Let us further assume that the set of states of the first model ($\{u_1, ..., u_{N_v(1)}\}$) are fully contained within the set of states of the second model ($\{u_1, ..., u_{N_v(1)}, ..., u_{N_v(2)}\}$), and the goal is to discriminate between the presence or absence of either of the model. Using a special stochastic parameter $\beta(\omega)$ that is allowed to take only binary values and new states $\boldsymbol{u}'_{N_v(1)+1} = \beta(\omega)\boldsymbol{u}_{N_v(1)+1}$, ..., $\boldsymbol{u}'_{N_v(2)} = \beta(\omega)\boldsymbol{u}_{N_v(2)}$, the general model can be written as (based on Eq. 2

and omitting explicit dependence on $\boldsymbol{x}$, t, & ω for brevity),

$$\begin{aligned}
\frac{\partial u_1}{\partial t} =&(1-\beta)\mathcal{L}^1_1[u_1,...,u_{N_v(1)},\boldsymbol{\theta}^1] + \beta\mathcal{L}^2_1[u_1,...,u_{N_v(1)},u'_{N_v(1)+1},...,u'_{N_v(2)},\boldsymbol{\theta}^2] \ , \\
&\vdots \\
\frac{\partial u_{N_v(1)}}{\partial t} =&(1-\beta)\mathcal{L}^1_{N_v(1)}[u_1,...,u_{N_v(1)},\boldsymbol{\theta}^1] + \beta\mathcal{L}^2_{N_v(1)}[u_1,...,u_{N_v(1)},u'_{N_v(1)+1},...,u'_{N_v(2)},\boldsymbol{\theta}^2] \ , \\
\frac{\partial u'_{N_v(1)+1}}{\partial t} =&\beta\mathcal{L}^2_{N_v(1)+1}[u_1,...,u_{N_v(1)},u'_{N_v(1)+1},...,u'_{N_v(2)},\boldsymbol{\theta}^2] \ , \\
&\vdots \\
\frac{\partial u'_{N_v(2)}}{\partial t} =&\beta\mathcal{L}^2_{N_v(2)}[u_1,...,u_{N_v(1)},u'_{N_v(1)+1},...,u'_{N_v(2)},\boldsymbol{\theta}^2] \ ,
\end{aligned} \tag{7}$$

where $\beta(\omega)=0$ leads to the first candidate model, and $\beta(\omega)=1$ to the second candidate model. In similar fashion, we can derive the general model for cases with more than two candidate models, with states in one model being aggregate of states in other models, etc.

3.2. Stochastic Piecewise Linear Function Approximations: Unknown Models

The above two new uses of special stochastic parameters require a set of candidate functional forms to choose from. However, in some cases, there might be no such prior information / candidates available, hence the unknown part $\widetilde{\mathcal{L}}$ of the model (Eq. 1). These model functions then need to be discovered. We thus propose to parameterize such an unknown function space using stochastic piece-wise continuous functions. In the present work, we consider dense piece-wise linear functions as this representation is both rich and simple, and provides practical approximations of any unknown function. It greatly enhances the functional space in which we can perform our Bayesian search, and enables the discovery of new learned functions.

For brevity, let us only consider the scalar case, where $\widetilde{\mathcal{L}}[u(\boldsymbol{x},t;\omega);\omega]$ is the unknown function (Eq. 1) of a single scalar state variable. Also, it is often the case that prior information about the range of values taken by the state variable is available, $u(\boldsymbol{x},t;\omega)\in[u_L,u_R]$, $\forall\boldsymbol{x}\in\mathcal{D}$ and $t\in[0,T]$. Now, to define a parameterization using continuous piece-wise linear segments, consider the range $\mathcal{H}=[u_L,u_R]$ to be an indexed collection of intervals with non-zero measure $\{I_i=[u^i_L,u^i_R]\}_{0\le i\le N_I}$ forming a partition of $\mathcal{H}$, i.e.,

$$\mathcal{H}=\bigcup_{i=0}^{N_I} I_i \quad \text{and} \quad \mathring{I}_i\cap\mathring{I}_j=\emptyset \quad \text{for } i\neq j \ , \tag{8}$$

and we use N_I+1 points to discretize the range, such that,

$$u_L=u^0_L<u^0_R=u^1_L<...<u^{N_I-1}_R=u^{N_I}_L<u^{N_I}_R=u_R \ . \tag{9}$$

Let $\{\Psi_0,...,\Psi_{N_I+1}\}$ be the linear functions defined on each of these element,

$$\begin{aligned}
\Psi_0(u) &= \begin{cases} \frac{1}{(u^0_R-u_L)}(u^0_R-u) & \text{if } u\in I_0 \ , \\ 0 & \text{otherwise} \end{cases} \\
\Psi_k(u) &= \begin{cases} \frac{1}{(u^{k-1}_R-u^{k-1}_L)}(u-u^{k-1}_L) & \text{if } u\in I_{k-1} \ , \\ \frac{1}{(u^k_R-u^k_L)}(u^k_R-u) & \text{if } u\in I_k \ , \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } k\in\{1,...,N_I\} \ , \\
\Psi_{N_I+1}(u) &= \begin{cases} \frac{1}{(u_R-u^N_L)}(u-u^N_L) & \text{if } u\in I_{N_I} \ , \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{10}$$

and $\gamma_k(\omega)'s,\ k \in 0, ..., N_I + 1$ be $N_I + 2$ *stochastic expansion coefficients* that parameterize the unknown function space by taking a linear combination of the functions defined on each elements. Hence, all together we obtain:

$$\widetilde{\mathcal{L}}[u(\boldsymbol{x}, t; \omega); \omega] = \sum_{k=0}^{N_I+1} \gamma_k(\omega) \Psi_k(u(\boldsymbol{x}, t; \omega)) \ . \tag{11}$$

Thus, estimating the stochastic parameters γ_k's, in turn leads to learning of the unknown model function. The above formulation ensures C^0 continuity in the functional space. The prior distribution of these parameters define the functional space in which the search is performed. By construction, this parameterized space can be made as dense as desired. Throughout this paper, we only consider linear-segments as the basis, however, this formulation can be extended to any other basis, such as higher degree polynomials, etc.

3.3. Bayesian Learning: stochastic DO PDEs, GMM-DO Filter, and Learning Skill

To provide an accurate and informative prior for our new Bayesian learning paradigm with uncertain and unknown nonlinear dynamics and PDEs, we employ Dynamically Orthogonal (DO) equations [47; 48; 49]. The DO equations are optimal reduced order differential equations that evolve, based on the governing nonlinear dynamics, the dominant probabilistic subspace. Their derivation with uncertain parameters is outlined in Appendix A and in Sect. 4.3 for biogeochemical specifics.

For the Bayesian learning at each observation time, the GMM-DO filter [44; 45] is used to perform nonlinear, non-Gaussian updates of the probability distribution of the state variables, as detailed in Appendix B. For the joint Bayesian learning of state variables and parameters, we combine the GMM-DO filter with state augmentation [50; 43] (Appendix C). Our novel schemes allow for efficient simultaneous Bayesian estimation of state variable fields, parameters, and model equations themselves, all while using a single modeling system. They recast the learning of compatible and compatible-embedded models into special parameter estimations and, to allow discovery of formulations, parameterize the space of unknown model functions using piece-wise linear continuous functions. For the former, the learning occurs within the space of candidate models while for the latter, it occurs outside of that space and into the space of unknown model functions, hence providing the capability for model discovery. Importantly, this discovery is interpretable as it is in the form of piece-wise continuous functions. In addition, all of our Bayesian estimations provide much more than maximum likelihood estimates: they predict and update the complete joint probability distribution of states, parameters, and models. When observations are not sufficiently informative to learn and eliminate all but one model, parameter value, or state variable field, our Bayesian learning estimates the correct multi-modal pdfs. Our learning can indeed represent ambiguity, e.g. multiple options are possible, or even equifinality [51], e.g. a set of model estimates have the same likelihood. It can also signal the presence of bias in competing model formulations. Such capabilities will be showcased in Sect. (5).

To evaluate the learning skill, we first compare the mean fields and parameters with the noisy observations, using several error metrics. We also analyze the evolution of the pdfs of fields and parameters, as well as the convergence of these pdfs with stochastic resolution. The definitions and notation used for the hyper-parameters used in the DO methodology and the GMM-DO filter are provided in Table D.2.

4. Biogeochemical-Physical Equations and Simulated Experiments Setup

In this section, we describe the specifics of our simulated Bayesian learning experiments. We start with the biogeochemical differential equations, their coupling with the physics PDEs, and the stochastic DO decomposition with uncertain and unknown terms. This is followed by details of the modeling domain, numerical methods, initialization of the stochastic simulations, true solution generation, simulated observations, and learning metrics.

4.1. Biogeochemical Models

The biogeochemical differential equations that we employ are adapted from [29; 15] and references therein, and from Newberger et. al., [52]. They meet the criterion of being compatible with each other, with low

complexity models being embedded in higher complexity models (compatible-embedded-models). We will utilize three reaction models: the three-component NPZ model, i.e., nutrients (N), phytoplankton (P), and zooplankton (Z); the four-component NPZD model, i.e., N, P, Z and Detritus (D); and, the five-component NNPZD model, i.e., ammonia (NH_4), nitrate (NO_3), P, Z, and D. The NPZ model is given by,

$$\begin{aligned}
\frac{dN}{dt} &= -G\frac{PN}{N+K_u} + \Xi P + \Gamma Z + R_m\gamma Z(1-\exp^{-\Lambda P}) \,, \\
\frac{dP}{dt} &= G\frac{PN}{N+K_u} - \Xi P - R_m Z(1-\exp^{-\Lambda P}) \,, \\
\frac{dZ}{dt} &= R_m(1-\gamma)Z(1-\exp^{-\Lambda P}) - \Gamma Z \,,
\end{aligned} \tag{12}$$

with G representing the optical model,

$$G = V_m\frac{\alpha I}{(V_m^2+\alpha^2 I^2)^{1/2}} \,, \quad \text{and} \quad I(z) = I_0\exp^{k_w z} \,, \tag{13}$$

where z is depth and $I(z)$ models the availability of sunlight for photo-chemical reactions. The parameters in Eqs. 12 & 13 are: k_w, light attenuation by sea water; α, initial slope of the P-I curve; I_0, surface photo-synthetically available radiation; V_m, phytoplankton maximum uptake rate; K_u, half-saturation constant for phytoplankton uptake of nutrients; Ξ, phytoplankton specific mortality rate; R_m, zooplankton maximum grazing rate; Λ, Ivlev grazing constant; γ, fraction of zooplankton grazing egested; and Γ, zooplankton specific excretion/mortality rate. In this NPZ model (Eq. 12), the nutrient uptake by phytoplankton is governed by a Michaelis-Menten formulation, which amounts to a linear uptake relationship at low nutrient concentrations that saturates to a constant at high concentrations. The grazing of phytoplankton by zoo-plankton follows a similar behavior: their growth rate becomes independent of P in case of abundance, but proportional to available P when resources are scarce; hence, zooplankton grazing is modeled by an Ivlev function. The death rates of both P and Z are linear, and a portion of zooplankton grazing in the form of excretion goes directly to nutrients.

For the NPZD model, the only change is in the addition of detritus, which is the intermediate state before dead plankton get converted to nutrients,

$$\begin{aligned}
\frac{dN}{dt} &= -G\frac{PN}{N+K_u} + \Phi D + \Gamma Z \,, \\
\frac{dD}{dt} &= R_m\gamma Z(1-\exp^{-\Lambda P}) + \Xi P - \Phi D \,.
\end{aligned} \tag{14}$$

However, for the NNPZD model, the nutrients are divided into ammonia and nitrates, which are the two most important forms of nitrogen in the ocean [1; 2]. This helps to capture new processes such as phytoplankton cells preferentially taking up ammonia over nitrate because the presence of ammonia inhibits the activity of the enzyme nitrate reductase essential for the uptake kinetics, the pool of ammonia coming from remineralization of detritus, and part of this ammonia pool getting oxidized to become a source of nitrate referred to as nitrification, etc. [1; 2; 10]. The NNPZD model is given by,

$$\begin{aligned}
\frac{dNO_3}{dt} &= \Omega NH_4 - G\left[\frac{NO_3}{NO_3+K_u}\exp^{-\Psi_I NH_4}\right]\widehat{P} \,, \\
\frac{dNH_4}{dt} &= -\Omega NH_4 + \Phi D + \Gamma Z - G\left[\frac{NH_4}{NH_4+K_u}\right]\widehat{P} \,, \\
\frac{dP}{dt} &= G\left[\frac{NO_3}{NO_3+K_u}\exp^{-\Psi_I NH_4} + \frac{NH_4}{NH_4+K_u}\right]\widehat{P} - \Xi P - R_m Z(1-\exp^{-\Lambda P}) \,, \\
\frac{dZ}{dt} &= R_m(1-\gamma)Z(1-\exp^{-\Lambda P}) - \Gamma Z \,, \\
\frac{dD}{dt} &= R_m\gamma Z(1-\exp^{-\Lambda P}) + \Xi P - \Phi D \,.
\end{aligned} \tag{15}$$

The above three models aim to capture the lower-trophic-level (LTL) interactions in the ocean ecosystem. They are the Lagrangian or ordinary differential equation (ODE) versions of these models. For realistic ocean field simulations, the above rate of change are material derivatives of dynamic tracers that are coupled with the physics using advection-diffusion-reaction PDEs. Of course, these models are not directly applicable in every ocean region without parameter tuning or modifying the functional form of the reaction terms. Regional diversity is one of the reasons for parameter and functional form (model) uncertainties.

4.2. Coupling with the Physics

In biogeochemical-physical models, the physics is provided by solving PDEs for the conservation of mass and momentum (Navier-Stokes), internal energy, and salt, e.g., the ocean primitive equations [53; 54]. These models often contain parameterizations to represent subgrid-scale processes [55; 56]. In the present work, we employ the incompressible nonhydrostatic Reynolds-averaged Navier-Stokes (RANS) PDEs [57],

$$\begin{aligned} \nabla.\boldsymbol{u}(\boldsymbol{x},t) &= 0, \qquad \boldsymbol{x}\in\mathcal{D}\ , \\ \frac{\partial \boldsymbol{u}(\boldsymbol{x},t)}{\partial t} + \nabla.(\boldsymbol{u}(\boldsymbol{x},t)\boldsymbol{u}(\boldsymbol{x},t)) &= -\nabla p(\boldsymbol{x},t) + \nu_E\nabla^2\boldsymbol{u}(\boldsymbol{x},t)\ , \qquad \boldsymbol{x}\in\mathcal{D}\ , \end{aligned} \tag{16}$$

where $\boldsymbol{u}(\boldsymbol{x},t)$ is the velocity field, $p(\boldsymbol{x},t)$ the pressure field, and ν_E the turbulent eddy viscosity.

The Lagrangian biogeochemical models (Sect. 4.1) are coupled with the physics using stochastic advection-diffusion-reaction (ADR) PDEs. For N_ϕ stochastic biogeochemical tracers, $\phi^i(\boldsymbol{x},t;\omega)$'s, we obtain,

$$\frac{\partial \phi^i(\boldsymbol{x},t;\omega)}{\partial t} + \underbrace{\nabla.(\boldsymbol{u}(\boldsymbol{x},t)\phi^i(\boldsymbol{x},t;\omega))}_{\text{Advection}} - \underbrace{\mathcal{K}_E\nabla^2\phi^i(\boldsymbol{x},t;\omega)}_{\text{Diffusion}} = \underbrace{S^{\phi^i}(\phi^1,...,\phi^{N_\phi},\theta^1(\omega),...,\theta^{N_\theta}(\omega),\boldsymbol{x},t;\omega)}_{\text{Reaction}}\ , \quad \forall i=\{1,...,N_\phi\}, \tag{17}$$

where $\boldsymbol{u}(\boldsymbol{x},t)$ is the deterministic velocity field governed by (16), $\mathcal{K}_E$ is the eddy diffusivity, $S^{\phi^i}(\phi^1,...,\phi^{N_\phi}, \theta^1(\omega),...,\theta^{N_\theta}(\omega),\boldsymbol{x},t;\omega)$ are the reaction terms defined by the right-hand-side of the ODEs of Sect. 4.1, and the $\theta^l(\omega)$'s, $l=\{1,...,N_\theta\}$, are the uncertain biogeochemical parameters. Biogeochemical reactions are nonlinear in nature, hence, the PDEs (17) form a set of strongly nonlinear, stiff, and coupled PDEs.

4.3. Biogeochemical-Physical Stochastic Dynamically-Orthogonal PDEs

To solve the system of Eqs. (16 & 17) efficiently, we now develop the DO equations for the stochastic ADR PDEs (17) with model and parameter uncertainty. We first separate the reactions into known, uncertain, and unknown terms, and write (17) in vector form,

$$\begin{aligned} \frac{\partial \boldsymbol{\phi}(\boldsymbol{x},t;\omega)}{\partial t} + \nabla.(\boldsymbol{u}(\boldsymbol{x},t)\boldsymbol{\phi}(\boldsymbol{x},t;\omega)) - \mathcal{K}_E\nabla^2\boldsymbol{\phi}(\boldsymbol{x},t;\omega) &= \boldsymbol{S}^{\boldsymbol{\phi}}(\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{\beta}(\omega),\boldsymbol{x},t;\omega) \\ &\quad + \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}(\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{\alpha}(\omega),\boldsymbol{\beta}(\omega),\boldsymbol{x},t;\omega) + \widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}(\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\gamma}(\omega),\boldsymbol{x},t;\omega)\ , \end{aligned} \tag{18}$$

where $\boldsymbol{\phi} = \left[\widehat{\phi}^i\right]_{i=1}^{N_\phi}$. The functional form of the first reaction term $\boldsymbol{S}^{\boldsymbol{\phi}}(\bullet) = \left[S^{\phi^i}(\bullet)\right]_{i=1}^{N_\phi}$ is assumed to be known, however it contains N_θ uncertain regular parameters $\boldsymbol{\theta}(\omega) = \left[\widehat{\theta}^k\right]_{k=1}^{N_\theta}$. The second term $\widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet) = \left[\widehat{S}^{\phi^i}(\bullet)\right]_{i=1}^{N_\phi}$ is uncertain: it belongs to a family of candidate functions, parameterized using N_α special stochastic parameters $\boldsymbol{\alpha}(\omega) = \left[\widehat{\alpha}^k\right]_{k=1}^{N_\alpha}$, and may contain uncertain regular parameters $\boldsymbol{\theta}(\omega)$. The candidate models of different complexities are combined using N_β special stochastic parameters $\boldsymbol{\beta}(\omega) = \left[\widehat{\beta}^k\right]_{k=1}^{N_\beta}$. The $\beta_k(\omega)$'s multiplied with the original biological tracer fields (as described in Sect. 3.1) are absorbed into ϕ_i's and not explicitly shown; however, $\beta_k(\omega)$'s usually appear on the right-hand-side (RHS) in $\boldsymbol{S}^{\boldsymbol{\phi}}(\bullet)$ and $\widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet)$. The third term $\widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet) = \left[\widetilde{S}^{\phi^i}(\bullet)\right]_{i=1}^{N_\phi}$ has a functional form completely unknown, and is parameterized using

N_γ stochastic expansion coefficients $\boldsymbol{\gamma}(\omega) = \big[\widehat{\gamma}^k\big]_{k=1}^{N_\gamma}$. The DO decomposition for the biogeochemical fields into mean $\bar{\boldsymbol{\phi}}$, modes $\tilde{\boldsymbol{\phi}}_i$, and stochastic coefficients Y_i, is given by,

$$\boldsymbol{\phi}(\boldsymbol{x},t;\omega) = \bar{\boldsymbol{\phi}}(\boldsymbol{x},t) + \sum_{i=1}^{N_s} \widetilde{\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t) Y_i(t;\omega)}\,. \tag{19}$$

The uncertain regular and special stochastic parameters are split into means and deviations, $\boldsymbol{\theta}(\omega) = \bar{\boldsymbol{\theta}} + \boldsymbol{\mathfrak{D}}^\theta(\omega)$, $\boldsymbol{\alpha}(\omega) = \bar{\boldsymbol{\alpha}} + \boldsymbol{\mathfrak{D}}^\alpha(\omega)$, and $\boldsymbol{\beta}(\omega) = \bar{\boldsymbol{\beta}} + \boldsymbol{\mathfrak{D}}^\beta(\omega)$. For the nonlinear reaction terms in $\boldsymbol{S}^{\boldsymbol{\phi}}(\bullet)$ and $\widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet)$, as for the nonlinear path planning optimal propulsion term [58; 59], we utilize a local Taylor series expansion around the statistical means, $\bar{\boldsymbol{\phi}}(\boldsymbol{x},t)$, $\bar{\boldsymbol{\theta}}$, $\bar{\boldsymbol{\alpha}}$, and $\bar{\boldsymbol{\beta}}$, to locally represent the nonlinear stochastic effects in the reaction equations as nonlinear mean terms plus stochastic deviations. As we will exemplify, for most uncertainties, such stochastic approximation is efficient for Bayesian learning as it maintains the significant computational advantages of DO with respect to the other methods [60]. Finally, for maximum accuracy, we evaluate the $\widetilde{\boldsymbol{S}}[\bullet]$ terms for every state realization in a Monte-Carlo fashion. Details on DO schemes are provided in Appendix A. Next, we directly provide the DO equations for the mean, modes, and stochastic coefficients (omitting function arguments and using i, j, n, and m as summation indices),

$$\begin{aligned}
\frac{\partial \bar{\boldsymbol{\phi}}}{\partial t} &= -\nabla.(\boldsymbol{u}\bar{\boldsymbol{\phi}}) + \mathcal{K}_E\nabla^2\bar{\boldsymbol{\phi}} + \boldsymbol{S}^{\boldsymbol{\phi}}\big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}\big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \mathbb{E}[\widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}]\,,\\
\frac{\partial \tilde{\boldsymbol{\phi}}_i}{\partial t} &= \boldsymbol{Q}_i - \sum_{j=1}^{N_s} \widetilde{\langle \boldsymbol{Q}_i, \tilde{\boldsymbol{\phi}}_j\rangle \tilde{\boldsymbol{\phi}}_j}\,,\\
\frac{dY_i}{dt} &= \sum_{m=1}^{N_s} \langle \boldsymbol{F}_m, \tilde{\boldsymbol{\phi}}_i\rangle Y_m + \sum_{m=1}^{N_\theta} \widetilde{\left\langle \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \theta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle} \boldsymbol{\mathfrak{D}}^\theta_m + \sum_{m=1}^{N_\beta} \widetilde{\left\langle \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \beta}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle} \boldsymbol{\mathfrak{D}}^\beta_m\\
&\quad + \sum_{m=1}^{N_\theta} \widetilde{\left\langle \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \theta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle} \boldsymbol{\mathfrak{D}}^\theta_m + \sum_{m=1}^{N_\alpha} \widetilde{\left\langle \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \alpha_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle} \boldsymbol{\mathfrak{D}}^\alpha_m + \sum_{m=1}^{N_\beta} \widetilde{\left\langle \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \beta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle} \boldsymbol{\mathfrak{D}}^\beta_m\\
&\quad + \widetilde{\left\langle \widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}} - \mathbb{E}[\widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}], \tilde{\boldsymbol{\phi}}_i \right\rangle},
\end{aligned} \tag{20}$$

where,

$$\begin{aligned}
\boldsymbol{Q}_i &= -\nabla.(\boldsymbol{u}\tilde{\boldsymbol{\phi}}_i) + \mathcal{K}_E\nabla^2\tilde{\boldsymbol{\phi}}_i + \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_i + \sum_{j=1}^{N_s}\sum_{n=1}^{N_\theta} \widetilde{C^{-1}_{Y_iY_j} C_{\boldsymbol{\mathfrak{D}}^\theta_n Y_j}} \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \theta_n}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}\\
&\quad + \sum_{j=1}^{N_s}\sum_{n=1}^{N_\beta} \widetilde{C^{-1}_{Y_iY_j} C_{\boldsymbol{\mathfrak{D}}^\beta_n Y_j}} \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \beta_n}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_i + \sum_{j=1}^{N_s}\sum_{n=1}^{N_\theta} \widetilde{C^{-1}_{Y_iY_j} C_{\boldsymbol{\mathfrak{D}}^\theta_n Y_j}} \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \theta_n}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}\\
&\quad + \sum_{j=1}^{N_s}\sum_{n=1}^{N_\alpha} \widetilde{C^{-1}_{Y_iY_j} C_{\boldsymbol{\mathfrak{D}}^\alpha_n Y_j}} \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \alpha_n}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \sum_{j=1}^{N_s}\sum_{n=1}^{N_\beta} \widetilde{C^{-1}_{Y_iY_j} C_{\boldsymbol{\mathfrak{D}}^\beta_n Y_j}} \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \beta_n}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \sum_{j=1}^{N_s} \widetilde{C^{-1}_{Y_iY_j} \mathbb{E}[Y_j \widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}]}\,,\\
\boldsymbol{F}_m &= -\nabla.(\boldsymbol{u}\tilde{\boldsymbol{\phi}}_m) + \mathcal{K}_E\nabla^2\tilde{\boldsymbol{\phi}}_m + \frac{\partial \boldsymbol{S}^{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_m + \frac{\partial \widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_m\,,
\end{aligned} \tag{21}$$

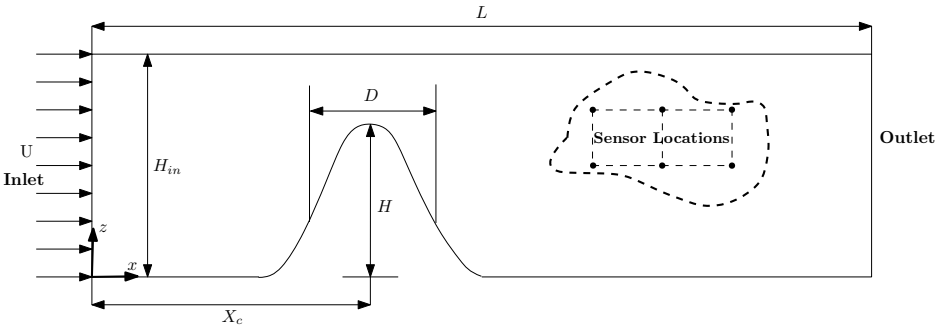


Figure 1: Two-dimensional spatial domain of the flow past a seamount. The seamount is defined by $He^{-(x-X_c)^2/D^2}$, where D is the characteristic width, H the height, and X_c the distance between the inlet and the center of the seamount. Observations are collected downstream of the seamount (see example sensor locations inset), with the exact observation locations depending on the particular experiment.

with $C_{\bullet,\bullet}$ representing cross-covariances, $\mathbb{E}[\bullet]$ expectations, and $\langle \bullet, \bullet \rangle$ spatial inner-products operators.

4.4. Modeling Domain and Boundary Conditions

Our modeling domain is inspired from the Stellwagen Bank at the edge of Massachusetts Bay, which is a whale feeding ground [61; 62; 10; 63; 64; 65; 66]. The experimental setup consists of a two-dimensional domain with a seamount representing an idealized sill (Fig. 1). The mean flow occurs from left to right in the positive x-direction over the seamount. Such flows can create upwelling of nutrients, leading to phytoplankton blooms, zooplankton responses, and nutrient uptake and recycling.

A horizontal length scale of $D \approx 1\ km$ is chosen for the seamount, while the vertical height scale is $H \approx 50\ m$. The overall transverse height of the domain is $H_{in} = 100\ m$. The longitudinal length of the domain is $L = 20\ km$, with center of the seamount at $X_c = 7.5\ km$.

Further, we only consider deterministic boundary conditions (BCs) models. The inlet at the left boundary has Dirichlet BCs for velocity, and zero Neumann for biological tracers,

$$u = U,\ v = 0 \qquad \text{and } \frac{\partial \boldsymbol{\phi}^i}{\partial x} = 0, \qquad \text{at } x = 0\ ,\ \forall i \in \{1, ..., N_\phi\}\ . \tag{22}$$

On the top and bottom boundary, free slip for velocity and again zero Neumann for tracers are applied,

$$\frac{\partial u}{\partial z} = 0,\ v = 0 \text{ and } \frac{\partial \boldsymbol{\phi}^i}{\partial z} = 0, \qquad \text{at } z = 0\ \&\ h\ ,\ \forall i \in \{1, ..., N_\phi\}\ . \tag{23}$$

At the outlet on the right boundary, we have open BCs with zero Neumann for all the state variables,

$$\frac{\partial u}{\partial x} = 0,\ \frac{\partial v}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial \boldsymbol{\phi}^i}{\partial x} = 0, \qquad \text{at } x = L\ ,\ \forall i \in \{1, ..., N_\phi\}\ . \tag{24}$$

Finally, on the obstacle surface, no-slip for velocity and zero Neumann for tracers are used,

$$u = 0,\ v = 0 \qquad \text{and} \qquad \frac{\partial \boldsymbol{\phi}^i}{\partial x} = \frac{\partial \boldsymbol{\phi}^i}{\partial z} = 0, \qquad \text{at } z = He^{-(x-X_c)^2/D^2}\ ,\ \forall i \in \{1, ..., N_\phi\}\ . \tag{25}$$

4.5. Numerical Schemes

The velocity and pressure fields are governed by the incompressible nonhydrostatic RANS PDEs (16). The stochastic biogeochemical fields are coupled with this dynamic RANS flow and governed by a dynamic reduced-order representation of the original stochastic ADR PDEs (17), the DO ADR PDEs we derived (Eqs. 20 & 21). In all experiments Sect. 4.5, we solve the deterministic RANS-biogeochemical PDEs for the true solution as well as the RANS-biogeochemical DO equations for the predicted pdfs using our modular finite-volume framework [67]. The physical domain (Sect. 4.4) is discretized using a uniform finite-volume

staggered C-grid, for both the flow and stochastic biogeochemical fields. The size of finite volumes in each $x-$ and $z-$ direction is equal to $\Delta x = \frac{1}{15}$ and $\Delta z = \frac{1}{15}$ (non-dimensional) respectively, thus, a grid-size of 300×30. Advection is computed explicitly, using a total variation diminishing (TVD) scheme with a monotonized flux limiter [68]. Diffusion is treated implicitly, with a second-order central difference scheme. All the reaction terms are computed explicitly. To handle the complex boundaries with the structured Cartesian grid, a ghost cell immersed boundary method is adopted for accurate enforcement of the boundary conditions [69]. For time-marching of the PDEs (RANS, DO mean, and DO modes), we use a first-order forward Euler method, while for the stochastic DO coefficient ODEs, we use a four-stage Runge-Kutta scheme. A non-dimensional time-step of $\Delta t = \frac{1}{240}$ is used in all the experiments. It is also ensured that we satisfy the Courant-Friedrichs-Lewy (CFL) condition at all times. We refer to ([70]) and ([71]) for more details on the numerical schemes we employ.

4.6. Balanced Initialization: Parameters, State Variable Fields, and Probabilities

The values of the parameters for the physics are chosen such that the flow emulates some coastal ocean dynamics. The dimensional barotropic velocity at the inlet is chosen to be $U \approx 10^{-2}$ to 10^{-1} m/s. The subgridscale eddy-viscosity is $\nu_E \approx 0.01$ to 0.5 m^2/s. Considering the vertical length scale of $H \approx 50$ m for the seamount, we obtain an eddy-viscosity Reynolds number of $Re = \frac{UH}{\nu_E} \approx 1$ to 500. Further, we do not consider any wind-forcing explicitly. For the initial velocity, we use a divergence free velocity field that satisfies the inlet and outlet boundary conditions, and so mass conservation in the given domain. The pressure field is initialized to be zero throughout the domain.

The biological parameters are either deterministic or stochastic. The values of the deterministic parameters are kept fixed for every realization, while the stochastic parameters are sampled from their respective probability distributions or joint distributions, if available. The stochastic parameters are further divided into two categories, regular and special, where the former were originally present in the biogeochemical models and have biological meanings associated with them, while the later are introduced for unification of candidate models and parameterization of unknown functions. The values of biological parameters used in the main experiments are given in Table 1. Probability distributions of all the stochastic parameters are assumed to be uniform and independent of each other, unless otherwise specified. In the experiments presented in this paper, advection-reaction dominates and the eddy-diffusivity for the biological tracers can be taken as negligible, $\mathcal{K}_E \approx 0$, such that the eddy-diffusivity Peclet number $Pe = \frac{UH}{\mathcal{K}_E} \to \infty$. Other experiments (not shown) were also successful however with non-negligible diffusivity, e.g. [43]. In all our simulations, a biological time-scale of the order of 1 *day* is used for all non-dimensionalization purposes.

Following [72; 10; 15; 30], in all the subsequent experiments, biogeochemical fields are initialized in dynamical balance, in accord with their stochastic model PDEs (17) and their parameter values. Specifically, the initial concentration fields for every sampled realization is obtained by finding an equilibrium solution corresponding to its sampled parameter values. These equilibrium fields are found by solving the ODE nonlinear biogeochemical models of Sect. 4.1 at all depths. Equilibrium is reached when temporal variations become negligible, or the system reaches a limit cycle. Further, we also impose the total biomass, $\sum_{i=1}^{N_\phi} \phi^i(z;\omega) = T_{bio}(z)$, to be conserved, with T_{bio} to be linearly increasing from 10 $mmol\ N\ m^{-3}$ at the surface to 30 $mmol\ N\ m^{-3}$ at the depth of 100 m, for all the biogeochemical models. This depth-dependent equilibrium solution for each of the biogeochemical state variables is used to initialize the corresponding fields in space, with the seamount masked at every x location. We also ensure that none of the realizations of the stochastic parameters lead to nonphysical equilibrium solutions, such as negative tracer values. The value of 30 $mmol\ N\ m^{-3}$ is used to non-dimensionalize all the biogeochemical fields and parameter values. For the non-dimensionalization of parameters, when needed, we additionally use a length-scale of 50 m (the height H of the seamount) and a time-scale of 1 *day*.

To initialize the DO decomposition of the biogeochemical fields, after generating the initial fields for each realization, we compute their statistical average and use it to initialize the mean biogeochemical fields. To initialize the DO modes and stochastic coefficients, we take the singular value decomposition (SVD) of the ensemble of mean-removed concatenated fields, keeping the dominant singular values and vectors. We account for the differences in the magnitude of the variability of individual biogeochemical tracers before taking the SVD, by appropriate normalization based on their standard deviations (Appendix A).

4.7. True Solution Generation

In the present work, twin experiments [73; 74; 75; 76] are conducted and the observations are extracted from a simulated truth. To obtain the simulated truth fields for each experiment, a set of parameters and initial state fields are sampled. Starting from these initial conditions, the Navier-Stokes PDEs (16) and the deterministic version of the ADR PDEs (17) with the true biogeochemical model are numerically integrated. The result is the simulated truth solution. In each experiment, all the remaining deterministic parameters, modeling domain, and numerical schemes are as these of the stochastic simulation using the DO equations.

4.8. Observations and Inference

Sparse observations are taken from the simulated true solution (Sect. 4.7). In each experiment, one of the biological tracer fields is observed at 6 to 9 locations (Fig. 1). The observation locations are kept in or near the euphotic zone because deeper depths have limited biological variability. The observation schedule is also experiment dependent, however, is it is not more frequent than once every non-dimensional time.

The observation error standard deviation matrix ($\sqrt{\boldsymbol{R}}$ in Eq. 3) is assumed diagonal. The linear observation matrix $\boldsymbol{H}$ (Eq. 3) is specified such that it predicts the concentration of the observed tracer field at the observation locations by interpolating the concatenated state fields at the observation locations.

Further, the hyper-parameters related to the DO equations and the GMM-DO filter were chosen based on numerical tests and experience [45; 77; 42; 78], for each of the experiments. For the DO equations, the number of modes, number of Monte-Carlo coefficient samples, time-step, etc., were selected so as to be sufficient to capture the dominant uncertainty and evolving probability distribution for each of the state vector fields, parameters, and model equations themselves. For Bayesian learning with the GMM-DO filter, the expectation-maximization (EM) algorithm [79] and Bayesian Information Criterion (BIC) [80; 81] were employed to select the optimal number of GMM components at each data time. Typical BIC-optimized values for N_{GMM} were found to be 10 for the present experiments.

4.9. Learning Metrics

We evaluate the performance of our Bayesian learning framework by comparing the learned solution with the true solution from which noisy observations were collected and by examining the posterior joint state-parameter-model probability distributions. For the former solution evaluations, we compare the true fields to the DO mean fields, and the true parameter values to the most probable DO pdf values of the parameters. To quantify performance, we examine the evolution of the Root Mean Square Error (RMSE) of the biogeochemical tracer fields, the uncertain regular ($\boldsymbol{\theta}(\omega)$) and special ($\boldsymbol{\alpha}(\omega)$ & $\boldsymbol{\beta}(\omega)$) parameters, and/or the stochastic expansion coefficients ($\boldsymbol{\gamma}(\omega)$). The RMSE between a evolved stochastic state field/parameter estimate $\phi(\boldsymbol{x}, t; \omega)$ and its corresponding true field/parameter $\phi^{true}(\boldsymbol{x}, t)$, is given by, $\sqrt{\frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \mathbb{E}[(\phi(\boldsymbol{x}, t; \omega) - \phi^{true}(\boldsymbol{x}, t))^2] d\boldsymbol{x}}$. The square of RMSE hence consists of two contributions [82], one is the square of the L_2 distance between the mean of the variable in the stochastic run and the simulated truth, while the other is the variance of the variable. In every experiment, the RMSE values of each variable are normalized by the corresponding RMSE value just before the first assimilation step. For the latter pdf evaluations, we analyze the evolution of the posterior pdfs of the stochastic DO coefficients, and of the regular and special stochastic parameters. For example, for the DO coefficient realizations, we employ 2-D scatter plots. For the stochastic parameters, we use marginals and kernel-density fits. We also evaluate the convergence of pdf estimates with stochastic resolution, i.e. increasing/decreasing stochastic numerical parameters (N_s, N_r, etc.), see Sect. 3.3.

5. Application Results and Discussion

In order to demonstrate the capabilities of our Bayesian learning we utilize four sets of twin experiments with different coupled biogeochemical-physical dynamics and learning objectives, and perform simultaneous Bayesian estimation of state variables, parameters, and model equations, using observations that are sparse in both space and time. To quantify performance, we evaluate several learning metrics, emphasizing the sharpness of the inference and the accuracy of probability distributions. For each of the four sets of experiments, we conduct multiple studies so as to evaluate the sensitivity to hyper-parameters. However, for each set, we present detailed results for only one experiment and summarize the other results.

5.1. Experiments 1: Discriminating among candidate functional forms

Biologically, mortality is a linear rate process. The mortality terms of phytoplankton and zooplankton however commonly act as "closure" parameterizations in models because as they allow for recycling of nutrients directly from plankton. As a result, due to the missing intermediate states in the recycling model, the zooplankton mortality and recycling processes are often modeled nonlinearly, with a concentration-dependent loss rate [3]. In this first set of experiments, we use the NPZ model with uncertainty introduced by the ambiguity in the presence or absence of a quadratic zooplankton mortality function, along with the uncertainty in the value of the Ivlev grazing parameter (Λ). The uncertainty in the initial biogeochemical conditions is set in balance with the uncertain parameters and model equations, as explained in Sect. 4.6. The learning objective is to simultaneously learn all the biological states, regular parameter Λ, and functional form of zooplankton mortality using a special stochastic parameter, by assimilating sparse observations.

The right-hand-side of NPZ model (Eq. 12) with the quadratic zooplankton mortality is given by,

$$\begin{aligned} S^N &= -G\frac{PN}{N+K_u} + \Xi P + \Gamma Z + \underbrace{\alpha(\omega)(\tilde{\Gamma} Z^2)}_{\text{Quad. Z Mort.}} + R_m \gamma Z(1-\exp^{-\Lambda(\omega)P}) \\ S^P &= G\frac{PN}{N+K_u} - \Xi P - R_m Z(1-\exp^{-\Lambda(\omega)P}) \\ S^Z &= R_m(1-\gamma)Z(1-\exp^{-\Lambda(\omega)P}) - \Gamma Z - \underbrace{\alpha(\omega)(\tilde{\Gamma} Z^2)}_{\text{Quad. Z Mort.}} \quad . \end{aligned} \tag{26}$$

The stochastic parameters are explicitly shown using the realization index (ω), and the ambiguous quadratic mortality term is pointed out. The special stochastic parameter, $\alpha(\omega)$, is restricted to binary values, i.e., either 0 or 1, corresponding to the absence or presence of the quadratic mortality term, respectively. $\Lambda(\omega)$ is sampled from a uniform probability distribution between the non-dimensional values of 3 and 6, and $\alpha(\omega)$ is assumed to have an initial 50%-50% probability of being 0 or 1. The stochastic ADR PDEs with the above stochastic NPZ reactions (Eq. 26) are coupled with the RANS flow PDEs, and solved with the DO methodology (Sects. 4.3–4.5). The other known model parameters related to the physical-biogeochemical model as well as the hyper-parameters for the DO equations are provided in Table 1.

True solution generation: The true solution from which observations are extracted, corresponds to the non-dimensional values, 3.6 for Λ, and 1 for α, i.e., the quadratic mortality term present. The true state fields are initialized and evolved as described in Sect. 4.7. *Observations and learning parameters:* The observations are sparse in both space and time, and consist of zooplankton measurements at six locations downstream of the seamount, only at every two non-dimensional times, starting at $t = 5$. The data shown in Fig. 2 is all that the Bayesian learning framework gets to assimilate over the course of the experiment. Other hyper-parameters related to the GMM-DO filtering are provided in Table 1. *Learning metrics:* As time advances, the sparse data are assimilated using the Bayesian GMM-DO filter in the augmented state space. We compare the true fields and paramaters to their DO estimates (mean and most probable values). To quantify performance, we examine the evolution of the normalized RMSEs (Sect. 4.9) for the N, P, and Z fields, and for the $\Lambda(\omega)$ and $\alpha(\omega)$ parameters, as well as the pdfs of the stochastic parameters, DO coefficients, and biological states.

Table 1: Values of the various domain-related, biological, physical, and hyper- parameters used in the four sets of experiments. $H = 50\ m$, $max\{T_{bio}(z)\} = 30\ mmol\ N\ m^{-3}$, and time-scale of 1 day, are the characteristic scales used for non-dimensionalization.

Parameters	**Exp. 1**	**Exp. 2**	**Exp. 3**	**Exp. 4**
Biogeochemical model	NPZ	NPZ & NPZD	NPZ	NNPZD
Biological Parameters				
Light attenuation due to sea water, k_w (m^{-1})	0.067	0.067	0.067	0.067
Initial slope of the P-I curve, α $((W\ m^{-2}\ day)^{-1})$	0.025	0.025	0.025	0.025
Surface photosynthetically available radiation, I_o $(W\ m^{-2})$	158.075	158.075	158.075	158.075
Phytoplankton maximum uptake rate, V_m (day^{-1})	1.5	1.5	1.5	1.5

Half-saturation for phytoplankton uptake of nutrients, K_u^* $(mmol\ N\ m^{-3})$	1	1	1	1
NH_4 inhibition parameter, Ψ_I $((mmol\ N\ m^{-3})^{-1})$	–	–	–	1.46
NH_4 oxidation coefficient, Ω (day^{-1})	–	–	–	0.25
Phytoplankton specific mortality rate, Ξ (day^{-1})	0.1	0.1	0.1	unif(0.01, 0.08)
Zooplankton specific excretion and mortality rate, Γ (day^{-1})	0.145	0.145	0.145	unif(0.125, 0.150)
Presence/absence of quadratic zooplankton term, α	unif{0, 1}	unif{0, 1}	–	unif{0, 1}
Quadratic zooplankton specific excretion and mortality rate, $\tilde{\Gamma}$ (day^{-1})	0.2	0.2	0.2	0.2
Zooplankton maximum grazing rate, R_m (day^{-1})	0.52	0.52	0.52	unif(0.52, 0.72)
Ivlev constant, Λ $((mmol\ N\ m^{-3})^{-1})$	unif(0.1, 0.2)	unif(0.1, 0.2)	0.13	unif(0.052, 0.072)
Fraction of zooplankton grazing egested, γ	0.3	0.3	0.2	0.3
Detritus decomposition rate, Φ (day^{-1})	1.03	1.03	1.03	1.03
Diffusion constants – horizontal & vertical, $(\mathcal{K}_E)$	0	0	0	0
Modeling Domain				
Height of the seamount, H (m)	50	50	50	50
Characteristic width of the seamount, D (km)	1	1	1	1
Distance between inlet and center of seamount, X_c (km)	7.5	7.5	7.5	7.5
Domain height, H_{in} (m)	100	100	100	100
Domain length, L (km)	20	20	20	20
Physical Parameters				
Inverse of Eddy-viscosity Reynolds nb., (Λ_{Re})	1	1	1	1/500
DO Parameters				
Number of Modes, N_s	20	40	20	15
Number of Monte-Carlo samples, N_r	10,000	10,000	1,000	10,000
GMM-DO Parameters				
State being observed	Z	Z	N	P
Observation error standard deviation, $(\sqrt{\boldsymbol{R}})$	0.05	0.05	0.035	0.04
Size of Observation vector, N_y	6	6	8	9
Observation start time (non-dim.)	5	5	1	2
Time interval between assimilations (non-dim.)	2	2	2	1
Observation end time (non-dim.)	25	25	25	25

5.1.1. Learning results

Figure 3 shows the initial state and parameters of the system (at $t = 0$), while Fig. 4 shows the evolved prior state and parameters of the system at $t = 5$ (i.e. just before the 1st observational episode). There are significant differences between the true and prior DO mean fields of the biogeochemical tracers. During these first five non-dimensional time units, a phytoplankton bloom develops just downstream (top-right) of the seamount: upwelling of nutrients above the seamount within the euphotic zone feeds the growth in phytoplankton biomass in the wake.

In Fig. 5, we illustrate the evolving statistics of the stochastic dynamical system from $t = 0$ to $t = 5$ just before data assimilation. We show fields of the phytoplankton standard deviation and dominant three DO modes (Panels 5a & 5b). The standard deviation fields clearly highlight the significant uncertainty around the phytoplankton subsurface maxima and bloom, reaching 30 percent of the mean field maxima. The uncertain subsurface maxima and bloom also clearly affect the DO modes. In Panels 5c & 5d, we show the joint distribution of the top four stochastic coefficients, along with the prior GMM fits using 10 components (Panel 5d). We use the BIC to find the optimal number of components required [44]. The joint distributions demonstrate the highly non-Gaussian nature of the stochastic DO coefficients, which the DO equations are able to evolve, and the GMM-DO filter is able account for. The strong parametric uncertainties is reflected by the thin 2D joint coefficient distributions. In addition, the realizations of the stochastic coefficients are clearly divided into two groups, corresponding to the presence or absence of the quadratic mortality term.

At $t = 5$, the first sparse data is assimilated. Fig. 6 shows the posterior mean fields, prior and posterior parametric distributions, and the normalized RMSE values for the mean fields and two stochastic parameters.

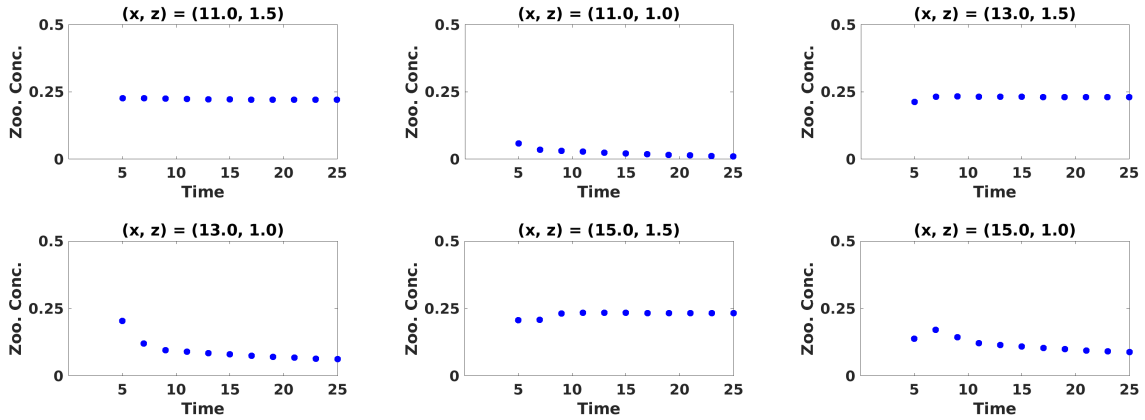


Figure 2: Experiments-1. Time-series of zooplankton concentration data collected at six observation locations (with coordinates given in the respective titles).

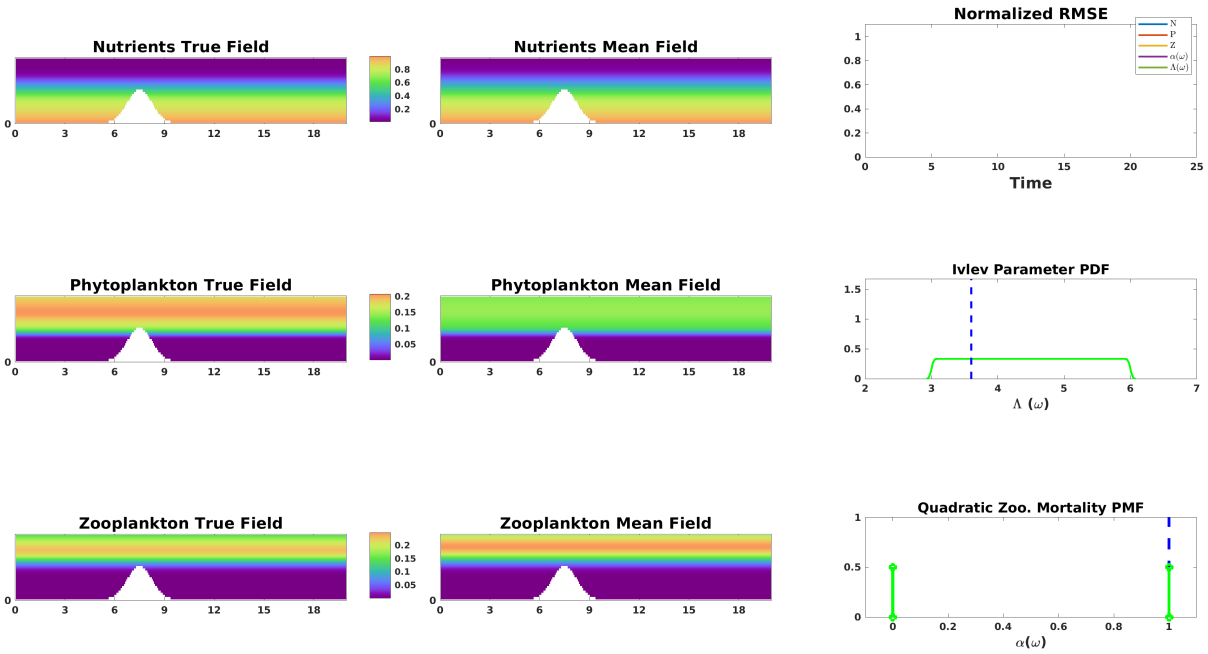


Figure 3: Experiments-1: State of the true and estimate NPZ fields and parameters at $t = 0$ (i.e. initial conditions). The first two columns consist of the non-dimensionalized true (left) and estimate mean (right) tracer fields of N, P and Z. In the third column, the top panel shows the variation of normalized root-mean-square-error (RMSE) with time for the stochastic state variables and parameters. The next two panels contain the pdf of the non-dimensional $\Lambda(\omega)$ and $a(\omega)$ (to learn presence or absence of quadratic zooplankton mortality), with their true unknown values marked with blue dotted lines. The velocity field is deterministic with $Re = 1$.

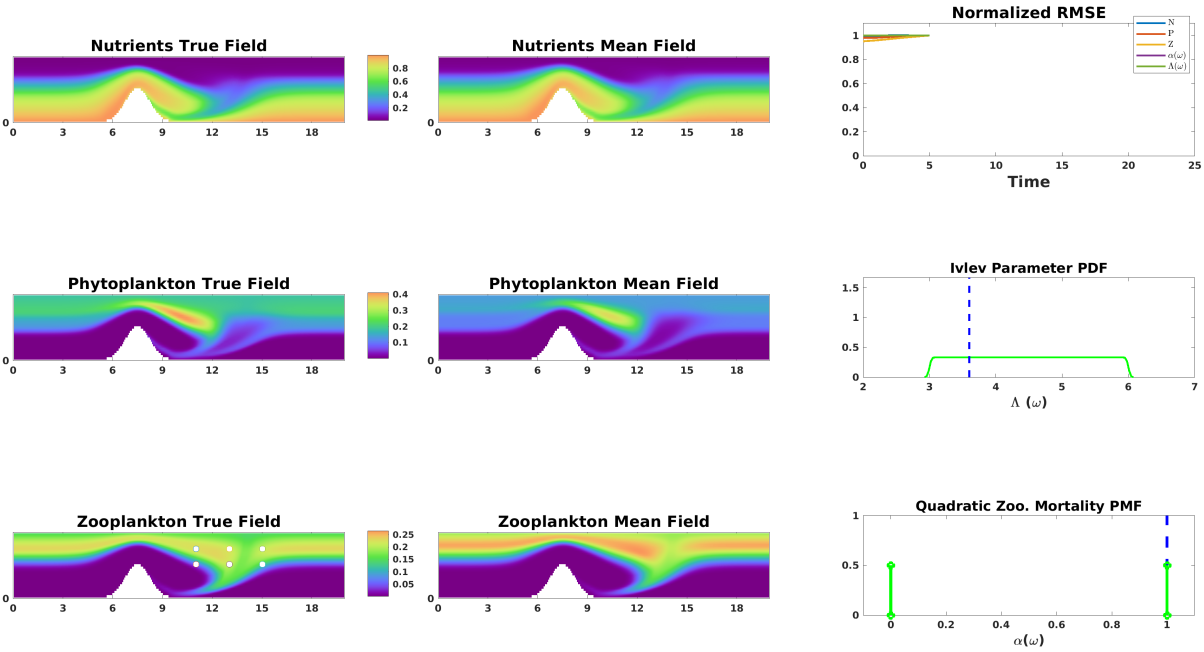


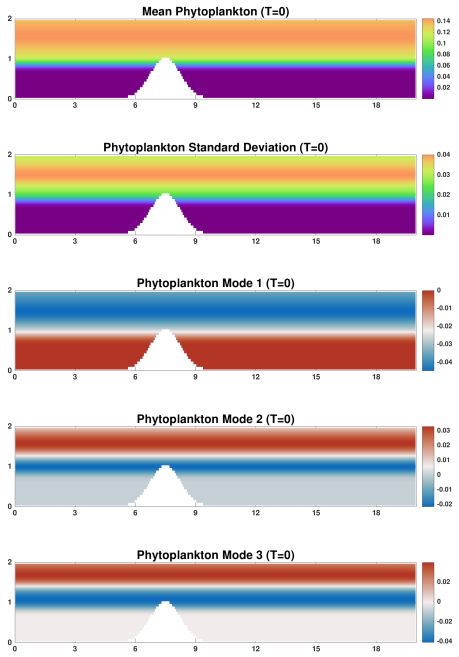
Figure 4: Experiments-1: As Fig. 3, but for the prior fields and parameters at $t = 5$ (i.e. just before the 1st assimilation). Additionally, the white circles on the zooplankton true field mark the six observation locations.

By just observing zooplankton at six locations, the GMM-DO filter simultaneously update all the biological fields and parameters. This is evident from the mean fields getting aligned with the true fields and quantified by the RMSE reductions of about 20 to 30 percent. Also visible is the slight change in the pdf for $\Lambda(\omega)$ and a higher probability value for $\alpha(\omega)$ being one. The six data are so far much more informative about the mortality term than about the Ivlev parameter.

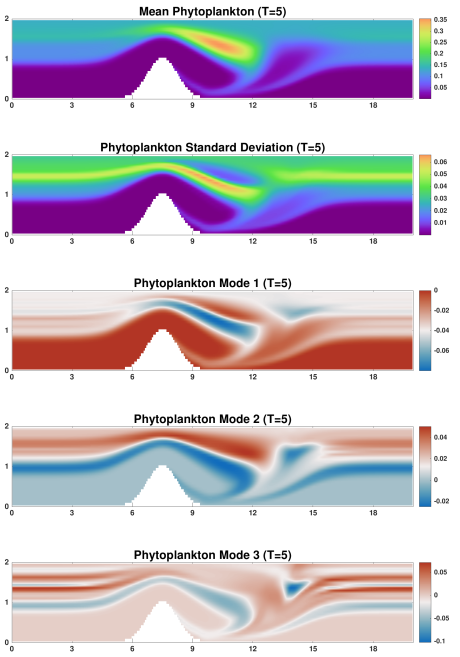
Next, in Fig. 7, we illustrate the same posterior mean fields, prior and posterior parameters, and normalized RMSE values, but at $t = 15$, i.e., at the sixth data assimilation. The flow is fully developed with the biogeochemical fields well learned, as quantified by the normalized RMSEs at about 40 percent. The GMM-DO filter unambiguously detects the presence of quadratic mortality of Z. The pdf of $\Lambda(\omega)$ is also accumulated around its true value, but is multi-modal, indicating nonlinearities and remaining ambiguity.

Finally, at $t = 25$, after 11 assimilation events, the same quantities are shown in Fig. 8. All the biogeochemical mean and true fields match with each other with RMSEs around 20 percent or less. The probability of the presence of the quadratic mortality term is now almost one, while the $\Lambda(\omega)$ pdf has a clear peak near 3.6 with a couple other much lower biased peaks around it. In general, the presence of lower peaks in pdfs of parameters indicate alternative combinations that could explain the data, and also the ability of the GMM-DO filter to capture non-Gaussian pdfs. The learning is also evident from the sustained decrease in the normalized RMSEs at every assimilation step for all the biogeochemical fields and parameters.

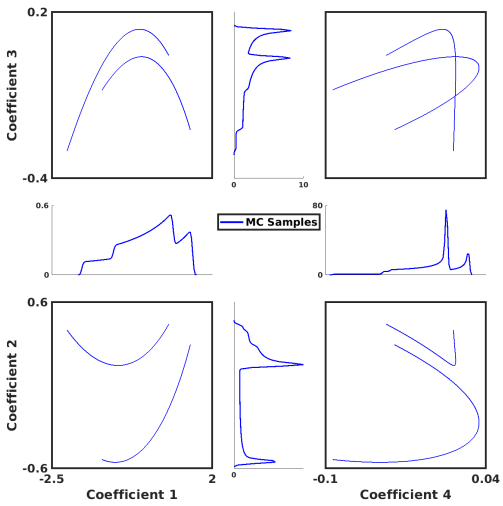
Many similar experiments were completed, changing various hyperparameters related to the GMM-DO filter, such as the biological variable being observed, observation locations, frequency, start-time, etc. Observations from simulated truths with different combinations of $\Lambda(\omega)$ and $\alpha(\omega)$ were also used. We found that the biological variable being observed has an impact on the sharpness of the inference or learnability of the given learning objectives. For example, observing N led to the learning of two distinct combinations of $\Lambda(\omega)$ & $\alpha(\omega)$, 3.1 & 0, and 3.6 & 1, respectively with nearly equal amount of confidence [83]. Decreasing the amount of observation data, or increasing the value of the observation error standard deviation led to larger uncertainty in the learned states and parameters. We also 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. This convergence of the pdfs of the



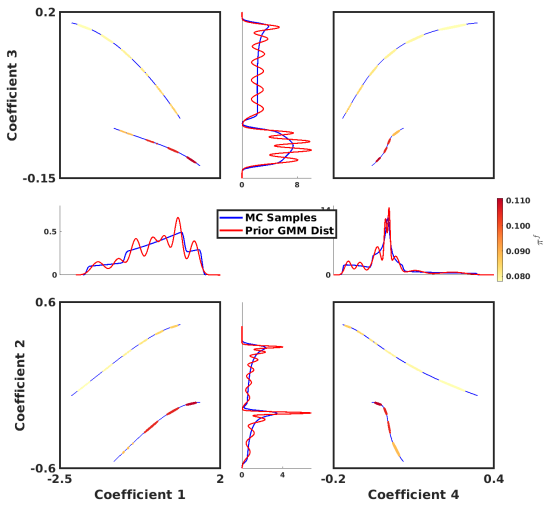
(a) Phytoplankton mean, standard deviation and top three DO modes, at $t = 0$



(b) Phytoplankton mean, standard deviation and top three DO modes, at $t = 5$ (prior)



(c) Joint distributions and respective marginals of the top four stochastic DO coefficients, at $t = 0$



(d) Joint distributions and respective marginals of the top four stochastic DO coefficients, along with the GMM fit, at $t = 5$ (prior)

Figure 5: Experiments-1: Statistics for the initial ($t = 0$) and prior ($t = 5$, just before the 1st assimilation) states of the stochastic NPZ ADR dynamical system.

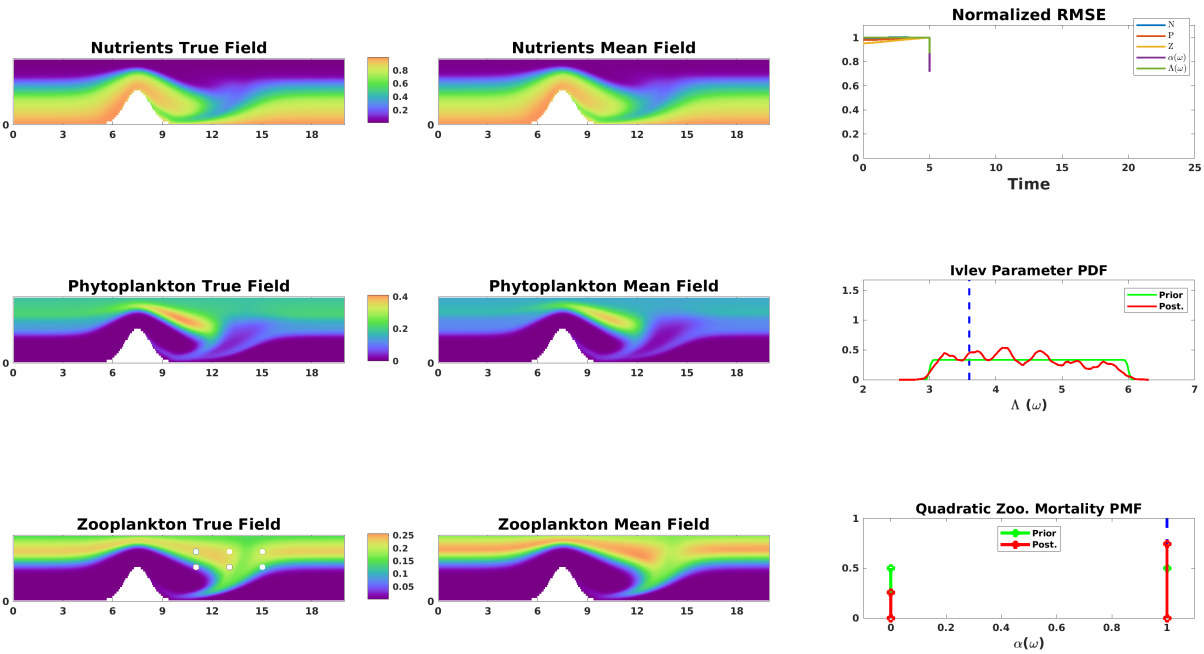


Figure 6: Experiments-1: As Figs. 3 & 4, but for posterior fields and parameters at $t = 5$ (i.e. just after the 1st assimilation).

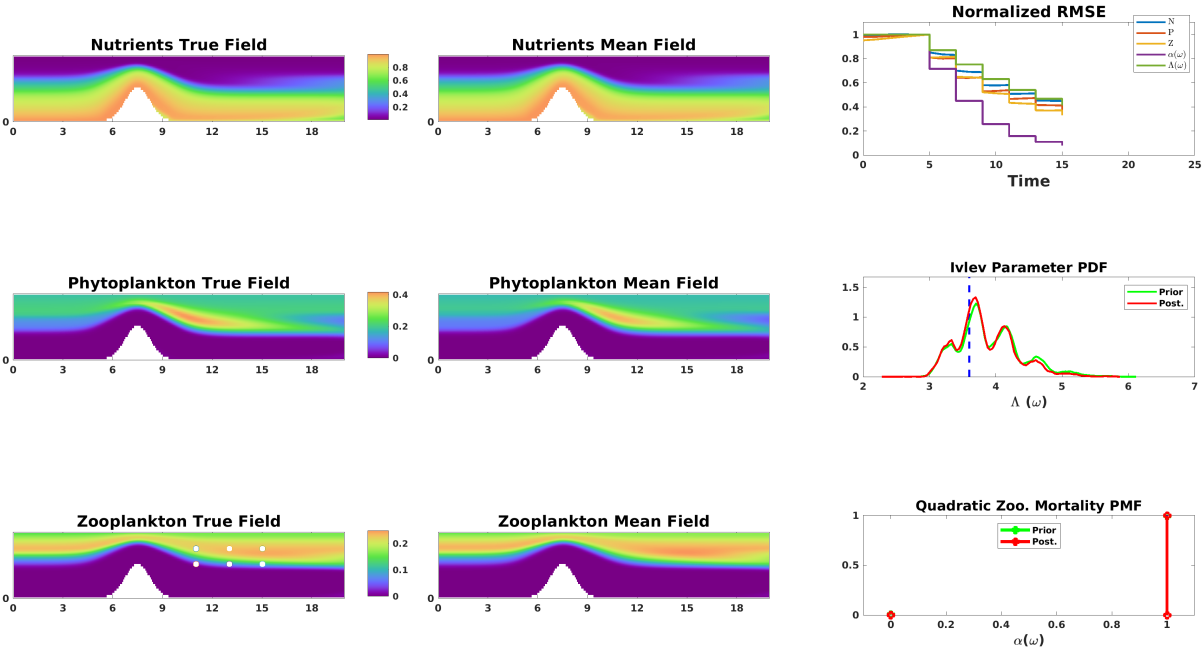


Figure 7: Experiments-1: As Figs. 3 & 4 but for posterior fields and parameters at $t = 15$ (i.e. just after the 6th assimilation).

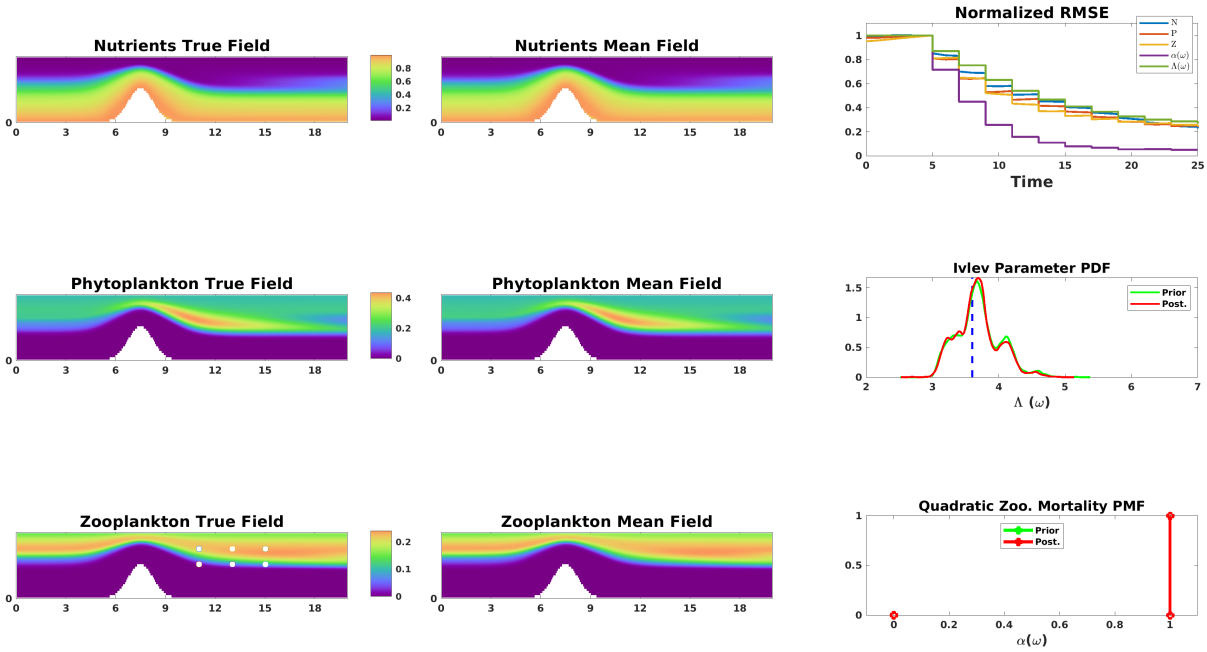


Figure 8: Experiments-1: As Figs. 3 & 4 but for posterior fields and parameters at $t = 25$ (i.e. just after the 11th assimilation).

parameters and DO coefficients, and of the DO modes and mean, indicates that our Bayesian GMM-DO filter provides accurate pdf estimates, and thus shows what has been learned without or with some ambiguity remaining. For the latter case, the multi-model posterior pdfs show that additional observations are needed to sharpen the inference further.

5.2. Experiments 2: Discriminating among models of different complexities

In the second set of experiments, the primary goal is to learn the complexity of the biogeochemical model, e.g., its state variables, along with the biogeochemical fields and Ivlev grazing parameter. Two candidates hierarchical model classes, NPZ and NPZD, are considered possible. To represent them with a single modeling system, we embed the former into the latter using our special stochastic parameter, $\beta(\omega)$. We multiply the detritus state variable (D) and other appropriate terms with $\beta(\omega)$, such that, the value of 1 derives the NPZD model, while the value of 0 derives the NPZ model (see Eq. 7). Thus, the RHS of the general stochastic model which encompasses both NPZ and NPZD models is given by,

$$\begin{aligned} S^N &= -G\frac{PN}{N+K_u} + \Phi D' + \Gamma Z + (1-\beta(\omega))\Xi P \\ S^P &= G\frac{PN}{N+K_u} - \Xi P - R_m Z(1-\exp^{-\Lambda(\omega)P}) \\ S^Z &= R_m(1-\beta(\omega)\gamma)Z(1-\exp^{-\Lambda(\omega)P}) - \Gamma Z \\ S^{D'} &= \beta(\omega)R_m\gamma Z(1-\exp^{-\Lambda(\omega)P}) + \beta(\omega)\Xi P - \Phi D' \\ D' &= \beta(\omega)D\ , \end{aligned} \tag{27}$$

where D' is the modified detritus state. Once again, $\Lambda(\omega)$ is sampled from a uniform probability distribution between the non-dimensional values of 3 and 6, and $\beta(\omega)$ is assumed to have 50%-50% probability of being 0 or 1. The stochastic ADR PDEs with the stochastic NPZD′ reactions (Eq. 27) are coupled with the RANS flow PDEs, and solved with the DO methodology (Sects. 4.3–4.5). The other known physical-biogeochemical parameters as well as the hyper-parameters for the DO equations are given in Table 1.

True solution generation: The true solution corresponds to the NPZ model with a non-dimensional value of 3.6 for the Λ parameter. The state fields are initialized and evolved as described in Sect. 4.7. *Observations and learning parameters:* The observations are sparse in both space and time, and again consist of zooplankton measurements at six locations downstream of the seamount, only at every two non-dimensional times, starting at $t = 5$. Other hyper-parameters related to the GMM-DO filtering are provided in Table 1. *Learning metrics:* As time advances and sparse data are assimilated, we compare the true fields and parameters to their DO estimates. To quantify performance, we examine the evolution of the normalized RMSEs of state fields and parameters, pdfs of stochastic parameters, and variances of DO coefficients.

5.2.1. Learning results

Figure 9 shows the state and parameters of the system at $t = 5$, just before the first observational episode. The most distinctive difference is between the true and mean detritus fields. Since the true model is NPZ, the true detritus field is equal to zero, while the mean detritus field is non-zero because half of the realizations correspond to the NPZD model. The RMSEs of all the variables exactly equal 1, because their respective values just before the first assimilation were used for normalization. The pdf of $\Lambda(\omega)$ is uniform in the main range, and $\beta(\omega)$ has 0.5 probability of being 0 or 1. The variances of the top five modes show a rapid decay with mode number, with the top two variances orders of magnitude larger. The variances of modes 3 and 4 differ initially but become similar over time, indicating a potential cross-over at $t = 5$.

In Fig. 10, we directly show the state of the system at time $t = 25$, after eleven GMM-DO data assimilation (six zooplankton values every two non-dimensional times). We find that our Bayesian learning framework is able to learn the true model to be NPZ, along with the posterior pdf of $\Lambda(\omega)$ concentrated around the true value of 3.6. The mean fields also match the true fields, especially the detritus mean field becoming very close to 0 at all the spatial locations. The RMSEs for all the variables decrease over time, up to about $t = 15$. At that time, the RMSE for the phytoplankton field increases due to a mismatch in the strength of the bloom, thus showing that the zooplankton data are not sufficiently informative for the same. The pdf of $\Lambda(\omega)$ features multiple peaks and thus still indicates that competing hypotheses remain for different pairs of parameter values; this was already the case in the intermediate assimilation steps (not shown). The evolution of the variances of the top five modes shows that these variances can increase and cross-over, for example, lower modes become more important as learning progresses. As the bloom develops, more complex nonlinear dynamics are activated, leading to the growth of some uncertainty modes. Results show that our Bayesian filter captures this as well as biases and non-Gaussian behaviors in the pdfs.

We performed other experiments with parameter sensitivity studies similar to those of Experiments-1; similar trends were found.

5.3. Experiments 3: Learning unknown functional forms

In our third set of experiments, the primary goal is to learn the functional form of the zooplankton mortality without any prior knowledge of candidate forms, along with the uncertain biological tracer fields. We utilize stochastic piece-wise linear functions to parameterize a large set of possible functional forms within a specified range, as explained in Sect. 3.2. Such a parameterization encompasses many different classes of functions, for example, polynomial, exponential, logarithmic, sinusoidal, etc. The right-hand-side of the stochastic NPZ model with the unknown function is given by,

$$\begin{aligned} S^N &= -G\frac{PN}{N+K_u} + \Xi P + \Gamma Z + \underbrace{F(Z;\omega)}_{\text{Unknown Function}} + R_m\gamma Z(1-\exp^{-\Lambda P}) \\ S^P &= G\frac{PN}{N+K_u} - \Xi P - R_m Z(1-\exp^{-\Lambda P}) \\ S^Z &= R_m(1-\gamma)Z(1-\exp^{-\Lambda P}) - \Gamma Z - \underbrace{F(Z;\omega)}_{\text{Unknown Function}} \end{aligned} \tag{28}$$

From prior knowledge [52], the non-dimensional value of zooplankton is assumed non negative and its maximum value to be 0.3. Thus, $F(Z;\omega)$ is set to be composed of any continuous piece-wise linear segments

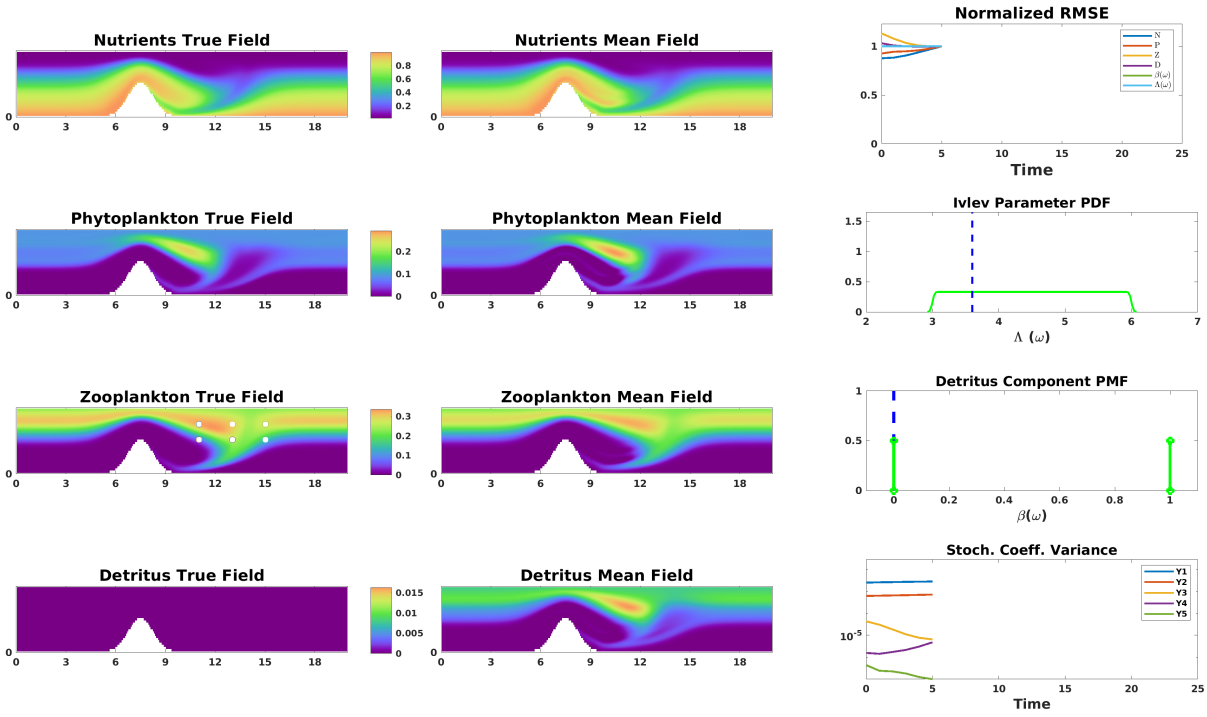


Figure 9: Experiments-2: State of the true and prior estimate NPZD fields and parameters at $t = 5$ (i.e. just before the 1st assimilation). The first two columns consist of the non-dimensionalized true (left) and estimate mean (right) tracer fields of N, P, Z, and D. In the third column, the first panel shows the variation of normalized RMSE with time for all the stochastic state variables and parameters. The next two panels contain the pdf of the non-dimensional $\Lambda(\omega)$ and $\beta(\omega)$ (to learn the complexity, NPZ vs. NPZD), with their true unknown values marked with blue dotted lines. The last panel shows the evolution with time of the variance (log scale) of the top five modes. The velocity field is deterministic with $Re = 1$. Additionally, the white circles on the zooplankton true field mark the six observation locations.

in the interval $Z \in [0, 0.3]$. Dividing this interval $[0, 0.3]$ into 10 equal non-overlapping sections, such that, $0 = Z_L^0 < Z_R^0 = 0.03 = Z_L^1 < ... < Z_R^9 = 0.27 = Z_L^{10} < Z_R^{10} = 0.3$, $F(Z;\omega)$ is thus represented as,

$$F(Z;\omega) = \sum_{k=0}^{11} \hat{\gamma}_k(\omega)\Psi_k(Z) \tag{29}$$

where,

$$\Psi_0(Z) = \begin{cases} \frac{1}{0.03}(0.03 - Z) & \text{if } 0 \le Z \le 0.03\ , \\ 0 & \text{otherwise} \end{cases}$$

$$\Psi_k(Z) = \begin{cases} \frac{1}{(Z_R^{k-1} - Z_L^{k-1})}(Z - Z_L^{k-1}) & \text{if } Z_L^{k-1} \le Z \le Z_R^{k-1}\ , \\ \frac{1}{(Z_R^k - Z_L^k)}(Z_R^k - Z) & \text{if } Z_L^k \le Z \le Z_R^k\ , \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } k \in \{1, ..., 10\}\ , \tag{30}$$

$$\Psi_{11}(Z) = \begin{cases} \frac{1}{0.03}(Z - 0.27) & \text{if } 0.27 \le Z \le 0.3\ , \\ 0 & \text{otherwise} \end{cases}$$

Each set of realizations of $\gamma_k's,\ k \in \{0, ..., 11\}$ are sampled so as to avoid a prior with unnatural highly fluctuating functions. The function range is set within 0 and 0.08; it is non-negative as mortality is negative in the zooplankton equation (28). To initialize the tracer fields, we find equilibrium solutions corresponding

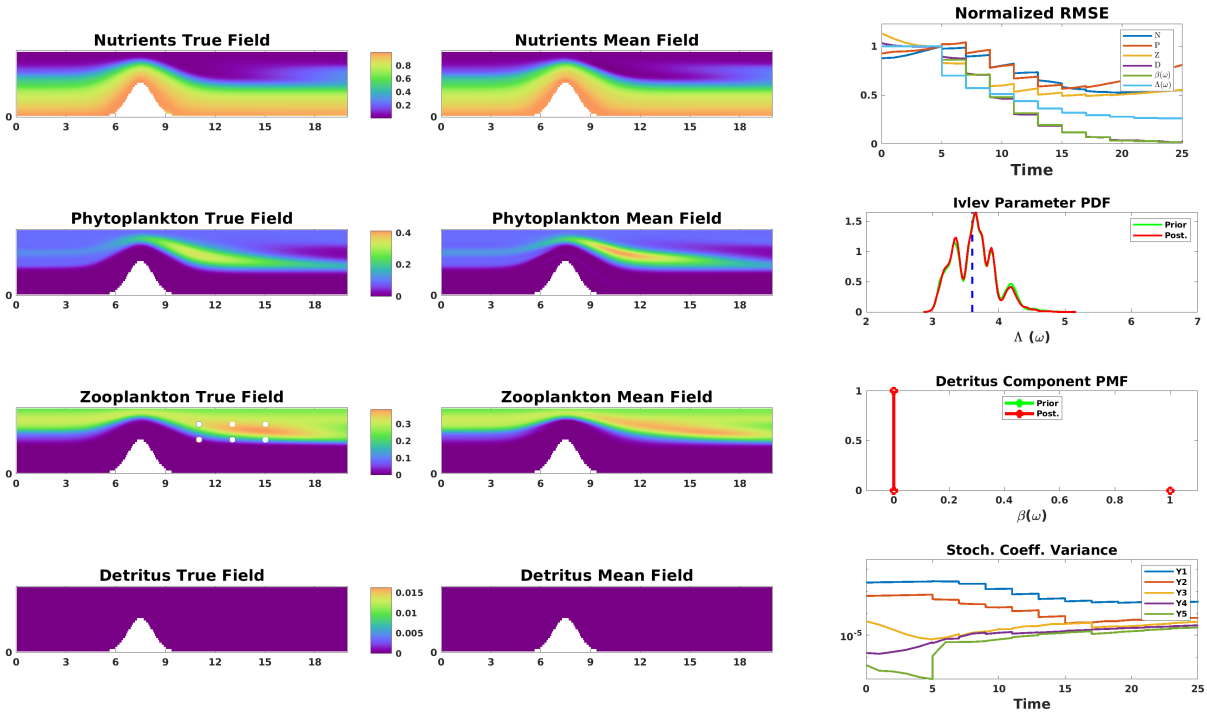


Figure 10: Experiments-2: As Fig. 9 but for posterior fields and parameters at $t = 25$ (i.e. just after the 11th assimilation).

to each realization of the zooplankton mortality function. The stochastic ADR PDEs with the stochastic NPZ reactions (Eq. 28) are coupled with the RANS flow PDEs, and solved with the DO methodology (Sects. 4.3–4.5). Table 1 provides the values of other known model and hyper- parameters. The learning objective of these experiments is to learn $F(Z;\omega)$ by estimating $\gamma_k' s$ along with the biological tracer fields.

True solution generation: The true solution contains quadratic zooplankton mortality, with values of the other parameters provided in Table 1. *Observations and learning parameters:* Observations remain sparse in time and space, but here they consist of the nutrient field at 8 spatial locations, starting at $t = 1$ and occurring every two non-dimensional times. In these experiments, we start the assimilation at the earlier $t = 1$ time in order to limit the exploding growth of uncertainty in the system, because each function realization leads to very different biological dynamics, several of which would lead to nonphysical biological states. Other hyper-parameters related to the GMM-DO filtering are provided in Table 1. *Learning metrics:* We compare the true fields and parameters to their DO estimates. To quantify performance, we also examine the evolution of the normalized RMSEs and pdf and realizations of the stochastic piece-wise linear functions.

5.3.1. Learning results

Figure 11 illustrates the prior at $t = 1$. Every realization in the space of the unknown function is assumed to be equilikely. In general, mortality being 0 for $Z = 0$ is common knowledge. Otherwise, it could act as a sink for zooplankton and lead to negative tracer values. However, we let this be discovered by the learning algorithm. The DO biogeochemical mean fields are quite far from the unknown true fields, and the prior function realizations are not similar to the true quadratic mortality.

As the eight N observations are assimilated every two non-dimensional times, nearly all the piece-wise linear function realizations converge to the true quadratic mortality. Results after 13 GMM-DO assimilation in Fig. 12 show this. We find however that the N data are not as informative about mortality function for Z beyond 0.25. This is in part because the maximum value reached in the true Z field is ~ 0.2, which limits the uncertainty reduction in the larger Z regime. The mean fields also converge to the true fields. The normalized RMSEs of all the biogeochemical fields indeed decrease at each assimilation. The learned

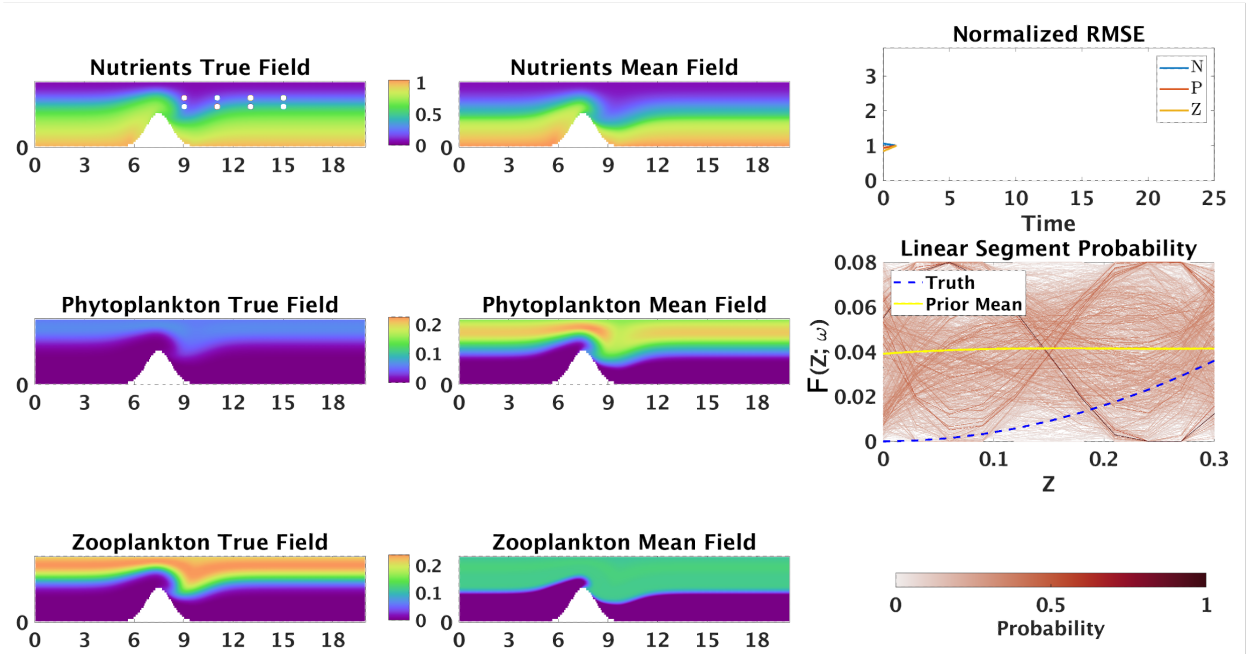


Figure 11: Experiments-3: State of the true and prior estimate NPZ fields and parameters at $t = 1$ (i.e. just before the 1st assimilation). The first two columns consist of the non-dimensionalized true (left) and estimate mean (right) tracer fields of N, P and Z. In the third column, the first panel shows the evolution of normalized RMSE for all the stochastic state variables. The second panel contains all the realizations of the unknown functional form approximated by piece-wise linear segments. The function realizations are colored according to their respective normalized probability density values. The velocity field is deterministic with $Re = 1$. Additionally, the white circles on the nutrient true field mark the 8 observation locations.

phytoplankton mean field however remain a bit higher than true fields, in part because they were much higher initially. It is also because the observed data (here eight N data) are not equally informative about all the learning objectives. As in [84; 85; 82], this is confirmed by mutual information fields (not shown).

Other experiments included studying the effect of incorporating or excluding prior knowledge such as the function value being 0 for $Z = 0$ and using smoothly varying function realizations. For the former, sampling γ_k's independent of each other led to highly fluctuating function realizations which completely impaired the learnability of the unknown function. For the latter, enforcing $\gamma_0 = 0$ sets $F(0;\omega) = 0$ for all realizations improved the convergence among the learned function realizations and the true function. Finally, increasing the number of independent observations (more N data, data for Z or P as well, etc.) also improved the sharpness of our GMM-DO inference: in all examples we show, we highlight cases with sparse observations as seen in real ocean applications.

5.4. Experiments 4: Learning in chaotic dynamics

In the last set of experiments, in order to robustly test our algorithms, the aim is to learn a five-component NNPZD model with a flow of Reynolds number $Re = 500$. At such high Re, vortices start to shed in the wake of the seamount and the flow chaotic. The learning objectives include all 5 biogeochemical fields, the Ivlev grazing parameter (Λ), the phytoplankton-specific mortality rate (Ξ), the zooplankton maximum grazing rate (R_m), the zooplankton specific mortality (Γ), and the presence or absence of the quadratic zooplankton mortality term. The stochastic NNPZD reactions, with all the uncertain parameters explicitly

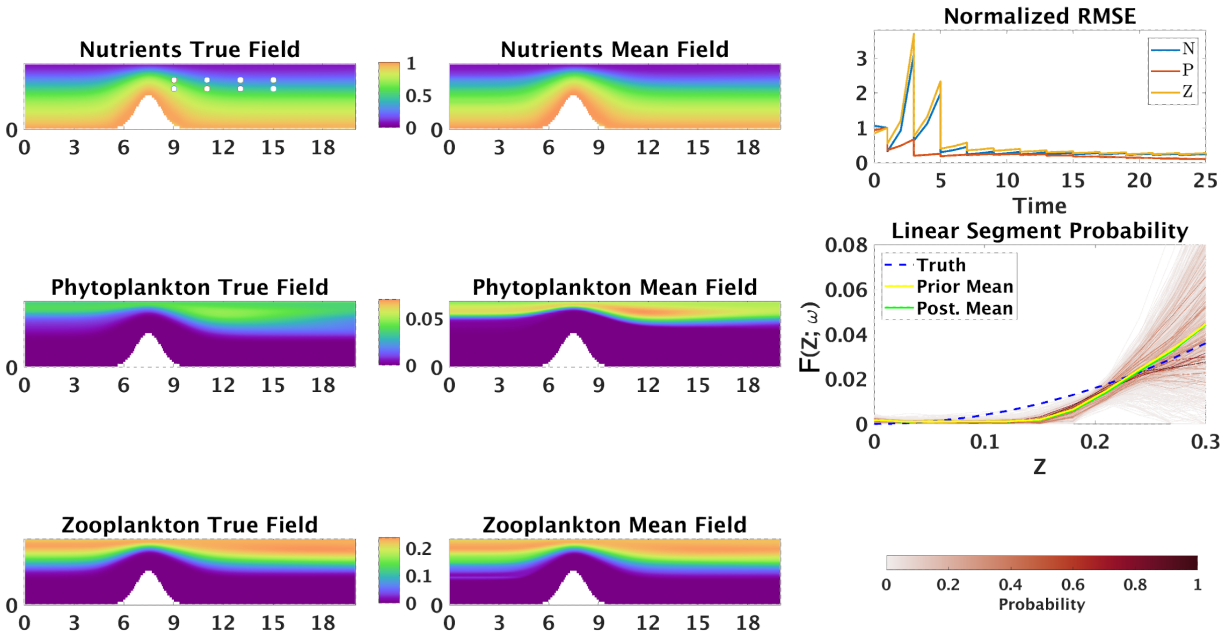


Figure 12: Experiments-3: As Fig. 11 but for posterior fields and function at $t = 25$ (i.e. just after the 13th assimilation).

containing ω as an argument, are given by,

$$
\begin{aligned}
S^{NO_3} &= \Omega NH_4 - G\left[\frac{NO_3}{NO_3 + K_u}\exp^{-\Psi_I NH_4}\right]\widehat{P}\,,\\
S^{NH_4} &= -\Omega NH_4 + \Phi D + \Gamma(\omega)Z + \alpha(\omega) \underbrace{(\Gamma_2 Z^2)}_{\text{Quad. Z Mort.}} - G\left[\frac{NH_4}{NH_4 + K_u}\right]\widehat{P}\,,\\
S^{P} &= G\left[\frac{NO_3}{NO_3 + K_u}\exp^{-\Psi_I NH_4} + \frac{NH_4}{NH_4 + K_u}\right]\widehat{P} - \Xi(\omega)P\,,\\
&\quad - R_m(\omega)Z(1-\exp^{-\Lambda(\omega)P})\,,\\
S^{Z} &= R_m(\omega)(1-\gamma)Z(1-\exp^{-\Lambda(\omega)P}) - \Gamma(\omega)Z + \alpha(\omega) \underbrace{(\Gamma_2 Z^2)}_{\text{Quad. Z Mort.}}\,,\\
S^{D} &= R_m(\omega)\gamma Z(1-\exp^{-\Lambda(\omega)P}) + \Xi(\omega)P - \Phi D\,.
\end{aligned}
\tag{31}
$$

Initially, we assume uniform and independent pdfs for the 4 uncertain regular parameters and equiprobability for the quadratic zooplankton mortality term to be present or absent. The stochastic ADR PDEs with the stochastic NNPZD reactions (31) are coupled with the deterministic RANS flow PDEs, and solved with the DO methodology (Sects. 4.3–4.5). The other known physical-biogeochemical model parameters as well as the hyper-parameters for the DO equations are provided in Table 1.

True solution generation: The true solution from which observations are extracted, corresponds to the non-dimensional values, 1.5 for Λ, 0.04 for Ξ, 0.6 for R_m, 0.14 for Γ, and 0 for α, i.e. the quadratic mortality term absent. The state fields are initialized and evolved as described in Sect. 4.7. *Observations and learning parameters:* Observations remain sparse and univariate, but due to the unstable and fast dynamics of the flow, there is a need for slightly more frequent data than in other experiments. The phytoplankton field is observed at nine locations starting at $t = 2$ and subsequently every one non-dimensional time. In total, we assimilate 24 times, i.e. until $t = 25$. Other hyper-parameters related to the GMM-DO filtering are provided in Table 1. *Learning metrics:* We compare the true fields and parameters to their DO estimates.

To quantify performance, we compute the evolution of the normalized RMSEs for all the 5 biological fields and 5 stochastic parameters. We also analyze the evolution of pdfs of the regular and special stochastic parameters, and the variances of DO coefficients.

5.4.1. Learning results

Figure 13 shows the prior estimates at $t = 2$. The flow has just started to develop. There are significant differences between the true and mean biogeochemical fields. The normalized RMSEs are equal to 1 by construction. The pdfs of all parameters remained as they were initially since no data has been assimilated.

Figure 14 illustrates the posterior estimates at $t = 2$, just after the first assimilation. Large corrections were made to the mean tracer fields (also visible in their RMSEs that decay by about 15 to 25%), and the GMM-DO learning already predicts the absence of quadratic zooplankton term. These first 9 P observations are not as informative however about the other parameters (their RMSEs only decay by about 4% to 8%).

Figure 15 shows the estimates at $t = 25$, after 24 GMM-DO assimilation steps. In addition to the mean fields, our augmented filter has been learning the 4 regular parameters. Their posterior pdfs have become Gaussian which has occurred in intermediate assimilation steps (not shown). We also show the evolution of variance of the top 3 modes. We find that the total variance on average either decreases or remains similar, while that of individual modes in general decreases but may also increase in accord with the stochastic dynamics. The velocity field being chaotic renders the learning more challenging in this experiment but our framework can still meet all the learning objectives, even with sparse and univariate data.

Other experiments were performed. As expected, they demonstrated sensitivity to the schedule, type, and quantity of observations. With only nine sparse and univariate data, starting them after the chaos sets in, or sampling even less frequently than every one non-dimensional time, led to posterior pdf of some stochastic parameters that were not concentrated around their respective true values. Similar results were found even when less than nine data were collected. Adding other observation types improved the learning. For other sensitivity studies, trends similar to other experiments were found.

6. Conclusions

Biogeochemical-physical models for the ocean are inherently uncertain due to the inability to capture all the complex marine interactions and processes with a single mathematical model. Uncertainty manifests itself in many different forms including the initial conditions, boundary conditions, parameters, parameterizations, state variables, and the model complexity and equations themselves. In this work, we develop a principled Bayesian model learning methodology that interpolates in the space of candidate models and discovers model formulations, all while estimating state variable fields and parameter values, as well as the joint probability distributions of all learned quantities. It employs the GMM-DO filter and state augmentation to predict and update pdfs of high-dimensional and multidisciplinary ecosystem dynamics governed by PDEs. Using sparse observations and Bayes' law, the complete joint probabilities of biogeochemical-physical fields and parameters, and of known, uncertain, and unknown model formulations are updated. Non-Gaussian statistics, ambiguity, and biases are captured. The parameter values and the model formulations that best explain the data are identified. A first crucial innovation are special stochastic parameters that unify compatible candidate models, possibly of different complexities, into a single general stochastic PDEs system. A second are stochastic piecewise function approximations that generate dense candidate model spaces. Our new methodology is generalizable and interpretable, and provides marginal pdfs for all learned model quantities. At the cost of single stochastic model simulation with parameter estimation, it seamlessly and rigorously discriminates among many existing models, possibly none of which are accurate, but also extrapolates out of the space of models to discover new ones.

The performance of our Bayesian learning framework was evaluated using a series of twin experiments based on flows past a seamount with compatible and embedded PDEs for the three-component NPZ model (nutrients (N), phytoplankton (P), and zooplankton (Z)), four-component NPZD model (N, P, Z and Detritus (D)), and five-component NNPZD model (ammonia (NH_4), nitrate (NO_3), P, Z, and D). In the first set of experiments, we use the NPZ model with uncertain initial conditions, unknown Ivlev grazing parameter value, and ambiguity in the presence or absence of the quadratic zooplankton mortality term. Our new

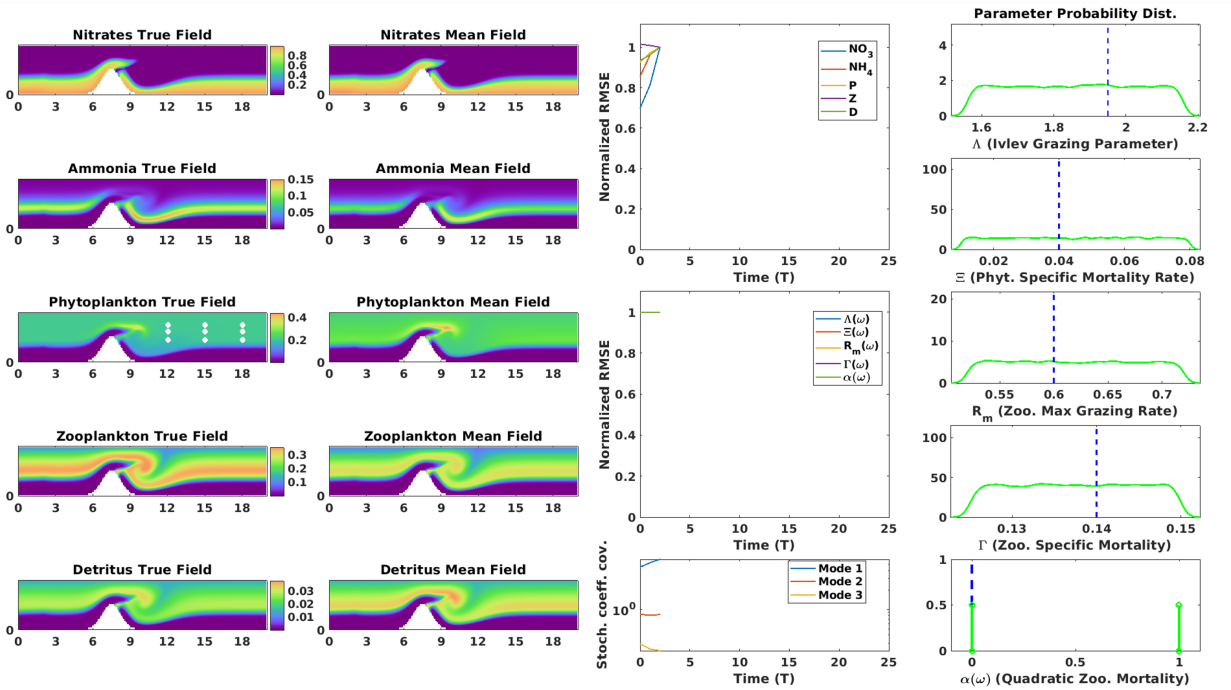


Figure 13: Experiments-4: State of the true and prior estimate NNPZD fields and parameters at $t = 2$ (i.e. just before the 1st assimilation). The first two columns consist of the non-dimensionalized true (left) and estimate mean (right) fields of NO_3, NH_4, P, Z, and D. In the third column, the first two panels show the evolution of the normalized RMSEs for the 5 state variables and 5 parameters. The third panel shows the evolution of variance of the top 3 DO modes. In the fourth column, the panels contain the pdf of the non-dimensional $\Lambda(\omega)$, $\Xi(\omega)$, $R_m(\omega)$, $\Gamma(\omega)$, and $\alpha(\omega)$ (learns the presence or absence of quadratic zooplankton mortality), with their true unknown values marked with blue dotted lines. The velocity field is deterministic with $Re = 500$. Additionally, the white circles on the phytoplankton true field mark the 9 observation locations.

Bayesian learning simultaneously estimated the state variables, Ivlev parameter, and unknown functional form, just using sparse Z observations in space and time (six data points every two non-dimensional times). The posterior pdf of the parameter contained secondary peaks, indicating that alternative combinations of parameter values could explain the observed data. This showcased the ability of our framework to capture non-Gaussian statistics including ambiguity and biases. In the second set of experiments, assimilating just eight N data every two non-dimensional times, we demonstrated the ability to learn the complexity of the model. We identified the true model within NPZ and NPZD, along with the uncertain fields and Ivlev grazing parameter. In the third set of experiments, we assumed no prior knowledge about the functional form of zooplankton mortality and generated a dense function space using stochastic piece-wise linear approximations. Assimilating just eight N data every two non-dimensional times, our framework then searched in this rich functional space, estimated the fields and regular parameter values, and was shown capable of discovering the mortality function. The last set of experiments involved learning the complex NNPZD model in an unsteady chaotic deterministic flow with vortex shedding. The NNPZD model had uncertainty in all the tracer fields, four parameters, and in the presence or absence of the zooplankton mortality term. All of the learning objectives were achieved simultaneously, using only nine P data point every non-dimensional time. In all cases, we quantified the learning skill, and evaluated convergence and the sensitivity to hyper-parameters.

These four sets of experiments were complementary, allowing us to showcase the features of our PDE Bayesian learning framework. It successfully discriminates among functional forms and model complexities, and also learns in the absence of prior knowledge by searching in dense function spaces. The next steps include applying this framework to more complex ocean applications. Even though we demonstrate our learning framework using biogeochemical models, it is applicable to many domains with model uncertainty,

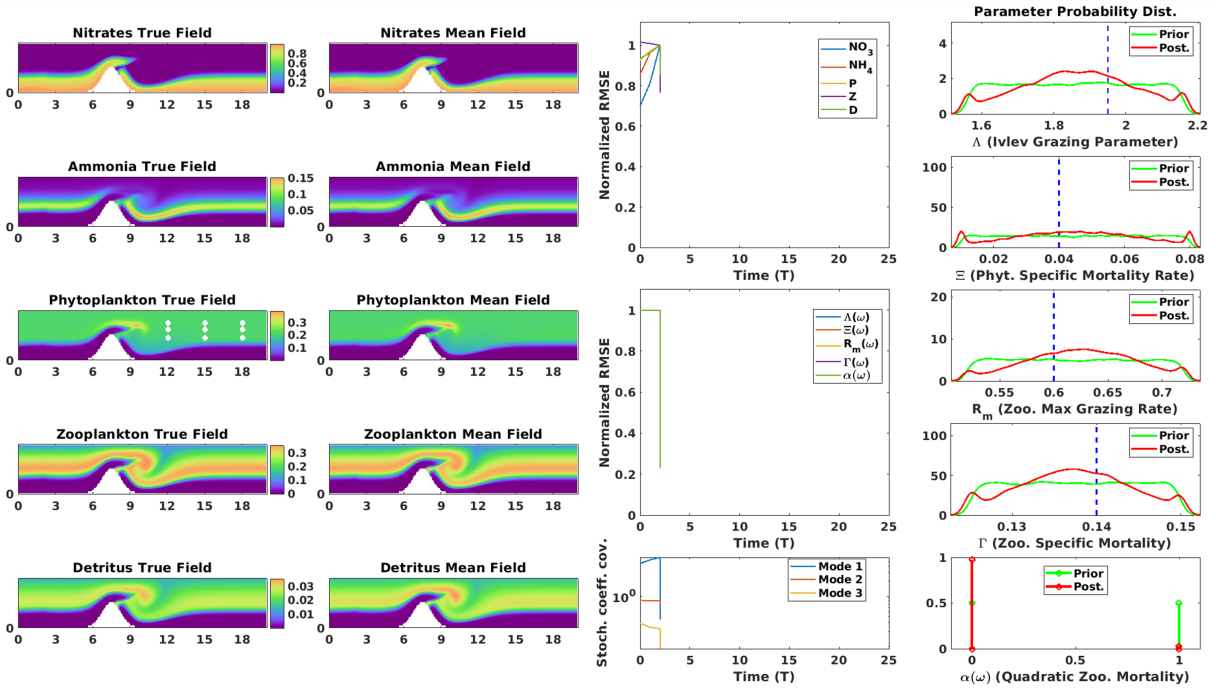


Figure 14: Experiments-4: As Fig. 13, but for posterior fields and parameters at $t = 2$ (i.e. just after the 1st assimilation).

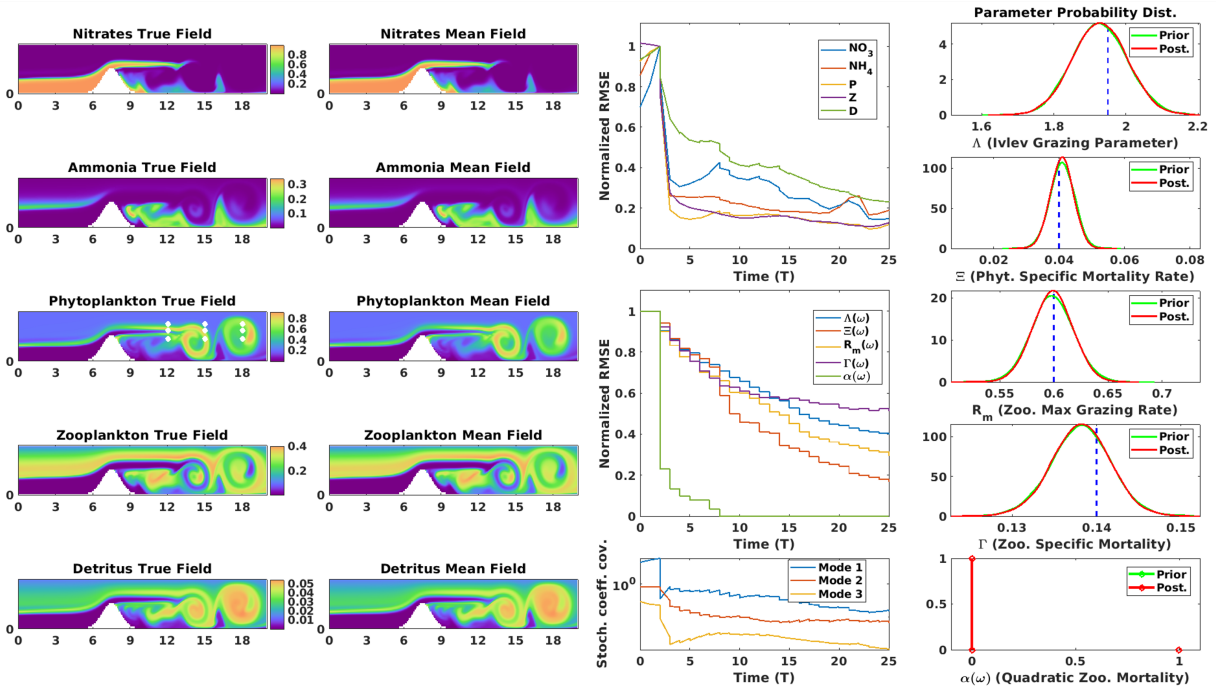


Figure 15: Experiments-4: As Fig. 13, but for posterior fields and parameters at $t = 25$ (i.e. just after the 24th assimilation).

for example, medicine, economics, energy, etc. Our framework can provide scientists not only the ability to choose between competing existing hypotheses but to also discover new ones in an principled manner.

Acknowledgments

We thank the members of our MSEAS group for insightful discussions, including A. Babu for proofreading. We are grateful to the Office of Naval Research for partial support under grants N00014-19-1-2693 (IN-BDA) and N00014-20-1-2023 (MURI ML-SCOPE), and to Sea Grant and NOAA for support under grant NA18OAR4170105 (BIOMAPS), all to the Massachusetts Institute of Technology. We also thank MathWorks and the Mechanical Eng. Dept. at MIT for awarding a competitive 2020-2021 MathWorks Mechanical Eng. Fellowship for A.G.

Appendix A. Dynamically Orthogonal (DO) Equations

In this appendix, we derive the dynamically orthogonal (DO) equations [47; 48; 71; 49] for optimal reduced-order probabilistic evolution of high-dimensional stochastic dynamical systems with regular and special parametric uncertainties for known, uncertain, and unknown model formulations.

The general stochastic nonlinear dynamical system governs the dynamics of $\boldsymbol{\phi}(\boldsymbol{x},t;\omega):\mathbb{R}^n\times[0,T]\rightarrow\mathbb{R}^{N_\phi}$, the stochastic state vector comprising N_ϕ physical-biogeochemical fields defined on a spatial domain $\mathcal{D}$, where ω is the realization index belonging to a measurable sample space Ω. It is is given by,

$$\begin{aligned}\frac{\partial\boldsymbol{\phi}(\boldsymbol{x},t;\omega)}{\partial t} &= \mathcal{L}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{\beta}(\omega),\boldsymbol{x},t;\omega]+\widehat{\mathcal{L}}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\alpha}(\omega),\boldsymbol{x},t;\omega]+\widetilde{\mathcal{L}}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\gamma}(\omega),\boldsymbol{x},t;\omega],\\ &\qquad\qquad \boldsymbol{x}\in\mathcal{D},\ t\in[0,T],\ \omega\in\Omega\,,\\ \text{with}\quad &\boldsymbol{\phi}(\boldsymbol{x},0;\omega)=\boldsymbol{\phi}_o(\boldsymbol{x};\omega)\,,\\ \text{and}\quad &\mathcal{B}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega)]=\boldsymbol{b}(\boldsymbol{x},t;\omega),\ \boldsymbol{x}\in\partial\mathcal{D},\ t\in[0,T],\ \omega\in\Omega\,,\end{aligned} \tag{A.1}$$

where $\boldsymbol{\phi}_o(\boldsymbol{x};\omega)$, $\mathcal{B}$, and $\boldsymbol{b}(\boldsymbol{x},t;\omega)$ are the stochastic initial conditions, boundary condition operators, and boundary values respectively. The functional form of the first dynamics term $\mathcal{L}[\bullet]$ is assumed to be known, however it contains N_θ uncertain regular parameters $\boldsymbol{\theta}(\omega)$. The second term $\widehat{\mathcal{L}}[\bullet]$ is uncertain: it belongs to a family of candidate functions, parameterized using N_α special stochastic parameters $\boldsymbol{\alpha}(\omega)$. $\widehat{\mathcal{L}}[\bullet]$ can also contain uncertain regular parameters $\boldsymbol{\theta}(\omega)$. The candidate models of different complexities are combined using N_β special stochastic parameters $\boldsymbol{\beta}(\omega)$. The $\beta_k(\omega)$'s multiplied with the original state variables (as described in Sect. 3.1) are absorbed into ϕ_i's and not explicitly shown; however, $\beta_k(\omega)$'s can still appear on the right-hand-side (RHS) in $\mathcal{L}[\bullet]$ and $\widehat{\mathcal{L}}[\bullet]$. The third term $\widetilde{\mathcal{L}}[\bullet]$ has a functional form completely unknown, and is parameterized using N_γ stochastic expansion coefficients $\boldsymbol{\gamma}(\omega)$.

The DO methodology employs a generalized, time-dependent Karhunen-Loéve decomposition of $\boldsymbol{\phi}(\boldsymbol{x},t;\omega)$ into a mean, $\bar{\boldsymbol{\phi}}(\boldsymbol{x},t)\in\mathbb{R}^{N_\phi}$, N_s deterministic modes, $\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)\in\mathbb{R}^{N_\phi}$, and stochastic coefficients, $Y_i(t;\omega)\in\mathbb{R}$,

$$\boldsymbol{\phi}(\boldsymbol{x},t;\omega)=\bar{\boldsymbol{\phi}}(\boldsymbol{x},t)+\sum_{i=1}^{N_s}Y_i(t;\omega)\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)\,. \tag{A.2}$$

We define the stochastic subspace $\boldsymbol{V_S}=span\{\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)\}_{i=1}^{N_s}$ as the linear space spanned by the N_s deterministic modes that evolve to capture the dominant uncertainty in $\boldsymbol{V_S}$. In general, the number of modes N_s is orders of magnitude smaller than the dimension of the discretized state variables or of the domain grid N_x, i.e. $N_s\ll N_\phi N_x$. Similarly, uncertain regular and special parameters are split into means and deviations, $\boldsymbol{\theta}(\omega)=\bar{\boldsymbol{\theta}}+\boldsymbol{\mathfrak{D}}^\theta(\omega)$, $\boldsymbol{\alpha}(\omega)=\bar{\boldsymbol{\alpha}}+\boldsymbol{\mathfrak{D}}^\alpha(\omega)$, and $\boldsymbol{\beta}(\omega)=\bar{\boldsymbol{\beta}}+\boldsymbol{\mathfrak{D}}^\beta(\omega)$.

Nonlinear terms on the RHS are handled using local Taylor series expansion around the statistical means of states and parameters. We use first order Taylor series expansion for the $\mathcal{L}[\bullet]$ and $\widehat{\mathcal{L}}[\bullet]$ terms,

$$\begin{aligned}
\mathcal{L}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\theta}(\omega),\boldsymbol{\beta}(\omega),\boldsymbol{x},t;\omega] &\approx \mathcal{L}|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \sum_{i=1}^{N_s} \tilde{\boldsymbol{\phi}}_i Y_i + \sum_{i=1}^{N_\theta} \frac{\partial \mathcal{L}}{\partial \theta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \mathfrak{D}_i^\theta + \sum_{i=1}^{N_\beta} \frac{\partial \mathcal{L}}{\partial \beta}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \mathfrak{D}_i^\beta \,, \\
\widehat{\mathcal{L}}[\boldsymbol{\phi}(\boldsymbol{x},t;\omega),\boldsymbol{\alpha}(\omega),\boldsymbol{x},t;\omega] &\approx \widehat{\mathcal{L}}|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}}}} + \frac{\partial \widehat{\mathcal{L}}}{\partial \boldsymbol{\phi}}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}}}} \sum_{i=1}^{N_s} \tilde{\boldsymbol{\phi}}_i Y_i + \sum_{i=1}^{N_\alpha} \frac{\partial \widehat{\mathcal{L}}}{\partial \alpha_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}}}} \mathfrak{D}_i^\alpha \,.
\end{aligned} \tag{A.3}$$

Using a higher-order polynomial approximation leads to higher accuracy for the DO evolution, but also increases computational costs. For analyses of the scaling of computational costs with the order of polynomial approximation, we refer to [83; 86]. Handling the $\widetilde{\mathcal{L}}[\bullet]$ term is less straightforward because of the need to evaluate the interval in which each state realization value lies at all discrete times and spatial locations in the domain (see Sect. 3.2). Thus, for maximum accuracy, we directly evaluate the $\widetilde{\mathcal{L}}[\bullet]$ terms for every state realization in a Monte-Carlo fashion. To increase efficiency without much loss of accuracy, recent techniques such as dynamic clustering [87; 88; 89] could also be used.

To derive the DO equations, we substitute the KL decomposition (Eq. A.2) into the stochastic system (Eq. A.1). To obtain an efficient closed-form dynamical system, without loss of generality, we impose the DO condition [47]: the rate of change of the stochastic subspace is orthogonal to itself,

$$\frac{d\boldsymbol{V_S}}{dt} \perp \boldsymbol{V_S} \Leftrightarrow \left\langle \frac{\partial \tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)}{\partial t}, \tilde{\boldsymbol{\phi}}_j(\boldsymbol{x},t) \right\rangle = 0 \quad \forall i,j \in \{1,...,N_s\}\,, \tag{A.4}$$

where $\langle \boldsymbol{a},\boldsymbol{b}\rangle = \int_{\mathcal{D}} \sum_i (a^i b^i) d\mathcal{D}$ denotes the spatial inner-product of vectors $\boldsymbol{a} = [a^1, a^2, ...]^T$ and $\boldsymbol{b} = [b^1, b^2, ...]^T$. Note that the DO condition (A.4) also implies the preservation of orthogonality for the basis $\{\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)\}_{i=1}^{N_s}$ themselves [70]. Substituting Eq. (A.2) into Eq. (A.1), and using Eq. (A.4) and the above schemes for nonlinear terms, we derive independent evolution equations for the DO mean, modes, and stochastic coefficients. These are the DO evolution equations (omitting function arguments for brevity),

$$\begin{aligned}
\frac{\partial \bar{\boldsymbol{\phi}}}{\partial t} &= \mathcal{L}|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \widehat{\mathcal{L}}|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \mathbb{E}[\widetilde{\mathcal{L}}] \\
\frac{\partial \tilde{\boldsymbol{\phi}}_i}{\partial t} &= \boldsymbol{Q}_i - \sum_{j=1}^{N_s} \langle \boldsymbol{Q}_i, \tilde{\boldsymbol{\phi}}_j \rangle \tilde{\boldsymbol{\phi}}_j \,, \\
\frac{dY_i}{dt} &= \sum_{m=1}^{N_s} \langle \boldsymbol{F}_m, \tilde{\boldsymbol{\phi}}_i \rangle Y_m + \sum_{m=1}^{N_\theta} \left\langle \frac{\partial \mathcal{L}}{\partial \theta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle \mathfrak{D}_m^\theta + \sum_{m=1}^{N_\beta} \left\langle \frac{\partial \mathcal{L}}{\partial \beta}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle \mathfrak{D}_m^\beta \\
&\quad + \sum_{m=1}^{N_\theta} \left\langle \frac{\partial \widehat{\mathcal{L}}}{\partial \theta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle \mathfrak{D}_m^\theta + \sum_{m=1}^{N_\alpha} \left\langle \frac{\partial \widehat{\mathcal{L}}}{\partial \alpha_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle \mathfrak{D}_m^\alpha + \sum_{m=1}^{N_\beta} \left\langle \frac{\partial \widehat{\mathcal{L}}}{\partial \beta_i}\Big|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}}, \tilde{\boldsymbol{\phi}}_i \right\rangle \mathfrak{D}_m^\beta \\
&\quad + \left\langle \widetilde{\mathcal{L}} - \mathbb{E}[\widetilde{\mathcal{L}}], \tilde{\boldsymbol{\phi}}_i \right\rangle,
\end{aligned} \tag{A.5}$$

where,

$$
\begin{aligned}
\boldsymbol{Q}_i = & \left.\frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_i + \sum_{j=1}^{N_s} \sum_{n=1}^{N_\theta} \widehat{C^{-1}_{Y_i Y_j} C_{\mathfrak{D}^\theta_n Y_j}} \left.\frac{\partial \mathcal{L}}{\partial \theta_n}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \sum_{j=1}^{N_s} \sum_{n=1}^{N_\beta} \widehat{C^{-1}_{Y_i Y_j} C_{\mathfrak{D}^\beta_n Y_j}} \left.\frac{\partial \mathcal{L}}{\partial \beta_n}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \\
& + \left.\frac{\partial \widehat{\mathcal{L}}}{\partial \boldsymbol{\phi}}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_i + \sum_{j=1}^{N_s} \sum_{n=1}^{N_\theta} \widehat{C^{-1}_{Y_i Y_j} C_{\mathfrak{D}^\theta_n Y_j}} \left.\frac{\partial \widehat{\mathcal{L}}}{\partial \theta_n}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \sum_{j=1}^{N_s} \sum_{n=1}^{N_\alpha} \widehat{C^{-1}_{Y_i Y_j} C_{\mathfrak{D}^\alpha_n Y_j}} \left.\frac{\partial \widehat{\mathcal{L}}}{\partial \alpha_n}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \\
& + \sum_{j=1}^{N_s} \sum_{n=1}^{N_\beta} \widehat{C^{-1}_{Y_i Y_j} C_{\mathfrak{D}^\beta_n Y_j}} \left.\frac{\partial \widehat{\mathcal{L}}}{\partial \beta_n}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} + \sum_{j=1}^{N_s} \widehat{C^{-1}_{Y_i Y_j} \mathbb{E}[Y_j \widetilde{\mathcal{L}}]}\,, \\
\boldsymbol{F}_m = & \left.\frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_m + \left.\frac{\partial \widehat{\mathcal{L}}}{\partial \boldsymbol{\phi}}\right|_{\substack{\boldsymbol{\phi}=\bar{\boldsymbol{\phi}},\\ \boldsymbol{\theta}=\bar{\boldsymbol{\theta}},\\ \boldsymbol{\alpha}=\bar{\boldsymbol{\alpha}},\\ \boldsymbol{\beta}=\bar{\boldsymbol{\beta}}}} \tilde{\boldsymbol{\phi}}_m\,,
\end{aligned}
\tag{A.6}
$$

and $\mathbb{E}[\bullet]$ represents the expectation operator, $C^{-1}_{Y_i Y_j}$ the inverse of the cross-covariance between the i^{th} and j^{th} stochastic coefficients, and $C_{Y_i Y_j}$ is given by,

$$
C_{Y_i, Y_j} = \mathbb{E}[Y_i(t;\omega) Y_j(t;\omega)]\,. \tag{A.7}
$$

As discussed in ([83; 86; 69]), the boundary conditions are also obtained by inserting DO decompositions in Eq. (A.1). This yields for the mean fields,

$$
\mathcal{B}[\bar{\boldsymbol{\phi}}(\boldsymbol{x},t)]|_{\boldsymbol{x}\in\partial\mathcal{D}} = \mathbb{E}[\boldsymbol{b}(\boldsymbol{x},t;\omega)]\,, \tag{A.8}
$$

and for the modes fields,

$$
\mathcal{B}[\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t)]|_{\boldsymbol{x}\in\partial\mathcal{D}} = \sum_{j=1}^{N_s} \widehat{\mathbb{E}[Y_j(t;\omega)\boldsymbol{b}(\boldsymbol{x},t;\omega)] C^{-1}_{Y_i Y_j}}\,. \tag{A.9}
$$

Similarly, the initial conditions in Eq. (A.2) are approximated by using the DO decomposition of the initial stochastic fields $\boldsymbol{\phi}_o(\boldsymbol{x};\omega)$.

Finally, the stochastic dynamical system (Eq. A.1) is multivariate and we normalize the spatial inner-product operator using appropriate scaling, so as to account for the different uncertainty magnitudes of state variables [76; 90; 72; 91; 92]. For the present DO modes $\tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t) = [\tilde{\phi}^1_i(\boldsymbol{x},t), ..., \tilde{\phi}^{N_\phi}_i(\boldsymbol{x},t)]$, the normalized spatial inner-product is,

$$
\langle \tilde{\boldsymbol{\phi}}_i(\boldsymbol{x},t), \tilde{\boldsymbol{\phi}}_j(\boldsymbol{x},t) \rangle = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} \widehat{\sum_{k=1}^{N_\phi} \left(\frac{1}{\sigma^2_{nd,k}} \tilde{\phi}^k_i \tilde{\phi}^k_j \right)} d\mathcal{D}\,, \tag{A.10}
$$

where $|\mathcal{D}|$ is the volume (area) of the domain and $\sigma_{nd,k}$ is the expected volume-averaged standard deviations of state variable k. These $\sigma_{nd,k}$'s normalize the relative weights given to state variables in the inner-product.

Appendix B. Gaussian Mixture Model (GMM)-DO Filter

The GMM-DO filter [44; 45] consists of a recursive succession in time of two steps: a forecast DO step (Appendix A) and a Bayesian update step. Using the affine transformation between stochastic coefficients

and state variables (Eq. A.2), the GMM-DO filter obtains the Bayesian update of the state variable distribution through an equivalent update of the stochastic coefficient distribution. The result is an efficient reduced-dimension Bayesian state variable inference [44]. Next, we assume the DO coefficients of the discrete state variables are augmented with the regular and special parameters (see Appendix C).

For the Bayesian update, the GMM-DO filter first represents the prior probability distribution of the stochastic coefficients in the DO subspace using a GMM,

$$p_{\boldsymbol{Y}^f}(Y^f) \approx \sum_{j=1}^{N_{\text{GMM}}} \pi^f_{\boldsymbol{Y},j} \times \mathcal{N}(Y^f; \boldsymbol{\mu}^f_{\boldsymbol{Y},j}, \boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}) \qquad \forall Y^f \in \mathbb{R}^{N_s} , \tag{B.1}$$

where N_{GMM} is the to-be-determined number of GMM components, $\pi^f_{\boldsymbol{Y},j} \in [0,1]$ the j^{th} component weight (also $\sum_{j=1}^{N_{\text{GMM}}} \pi^f_{\boldsymbol{Y},j} = 1$), $\boldsymbol{\mu}^f_{\boldsymbol{Y},j}$ the j^{th} component mean vector, and $\boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}$ the j^{th} 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 for GMMs [79] is used to find maximum likelihood estimate for the GMM parameters $\pi^f_{\boldsymbol{Y},j}$, $\boldsymbol{\mu}^f_{\boldsymbol{Y},j}$ and $\boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}$, while the selection of the number of GMM components (N_{GMM}) is determined by the Bayesian Information Criterion (BIC) [80] by successively fitting GMMs of varying complexity (e.g. GMM = 1, 2, 3, ...) until a minimum of the BIC is obtained.

Using the Gaussian observation model (Eq. 3), the GMM for the prior stochastic coefficients is updated by Bayesian update, using conjugacy [44]. The resulting GMM of the posterior stochastic coefficients is,

$$p_{\boldsymbol{Y}^a}(Y^a) \approx \sum_{j=1}^{N_{\text{GMM}}} \pi^a_{\boldsymbol{Y},j} \times \mathcal{N}(Y^a; \boldsymbol{\mu}^a_{\boldsymbol{Y},j}, \boldsymbol{\Sigma}^a_{\boldsymbol{Y},j}) , \qquad \forall Y^a \in \mathbb{R}^{N_s} , \tag{B.2}$$

where,

$$\begin{aligned}
\pi^a_{\boldsymbol{Y},j} &= \frac{\pi^f_{\boldsymbol{Y},j} \times \mathcal{N}(\tilde{\boldsymbol{y}}; \tilde{\boldsymbol{H}}\boldsymbol{\mu}^f_{\boldsymbol{Y},j}, \tilde{\boldsymbol{H}}\boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}\tilde{\boldsymbol{H}}^T + \boldsymbol{R})}{\sum_{m=1}^{N_{\text{GMM}}} \pi^f_{\boldsymbol{Y},m} \times \mathcal{N}(\tilde{\boldsymbol{y}}; \tilde{\boldsymbol{H}}\boldsymbol{\mu}^f_{\boldsymbol{Y},m}, \tilde{\boldsymbol{H}}\boldsymbol{\Sigma}^f_{\boldsymbol{Y},m}\tilde{\boldsymbol{H}}^T + \boldsymbol{R})} , \qquad \forall j \in \{1, ..., N_{\text{GMM}}\} , \\
\boldsymbol{\mu}^a_{\boldsymbol{Y},j} &= \hat{\boldsymbol{\mu}}^a_{\boldsymbol{Y},j} - \sum_{m=1}^{N_{\text{GMM}}} \pi^a_{\boldsymbol{Y},m} \times \hat{\boldsymbol{\mu}}^a_{\boldsymbol{Y},m} , \qquad \forall j \in \{1, ..., N_{\text{GMM}}\} , \\
\boldsymbol{\Sigma}^a_{\boldsymbol{Y},j} &= (\boldsymbol{I} - \tilde{\boldsymbol{K}}_j\tilde{\boldsymbol{H}})\boldsymbol{\Sigma}^f_{\boldsymbol{Y},j} , \qquad \forall j \in \{1, ..., N_{\text{GMM}}\} ,
\end{aligned} \tag{B.3}$$

with the following definitions,

$$\begin{aligned}
\tilde{\boldsymbol{H}} &= \boldsymbol{H}\tilde{\boldsymbol{\Phi}} , \\
\tilde{\boldsymbol{y}} &= \boldsymbol{y} - \boldsymbol{H}\bar{\boldsymbol{\Phi}}^f , \\
\hat{\boldsymbol{\mu}}^a_{\boldsymbol{Y},j} &= \boldsymbol{\mu}^f_{\boldsymbol{Y},j} + \tilde{\boldsymbol{K}}_j(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{H}}\boldsymbol{\mu}^f_{\boldsymbol{Y},j}) , \qquad \forall j \in \{1, ..., N_{\text{GMM}}\} , \\
\tilde{\boldsymbol{K}}_j &= \boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}\tilde{\boldsymbol{H}}^T(\tilde{\boldsymbol{H}}\boldsymbol{\Sigma}^f_{\boldsymbol{Y},j}\tilde{\boldsymbol{H}}^T + \boldsymbol{R})^{-1} \equiv \tilde{\boldsymbol{\Phi}}^T\boldsymbol{K}_j , \qquad \forall j \in \{1, ..., N_{\text{GMM}}\} .
\end{aligned} \tag{B.4}$$

The posterior GMM state space distribution is obtained from Eq. (B.2) by updating the state vector mean,

$$\bar{\boldsymbol{\Phi}}^a = \bar{\boldsymbol{\Phi}}^f + \tilde{\boldsymbol{\Phi}} \sum_{j=1}^{N_{\text{GMM}}} \pi^a_{\boldsymbol{Y},j} \times \hat{\boldsymbol{\mu}}^a_{\boldsymbol{Y},j} . \tag{B.5}$$

In the GMM-DO update step, no matrices of size larger than $N_\phi N_x \times S \ll (N_\phi N_x)^2$ are manipulated. The GMM-DO filter is thus computationally feasible for high-dimensional multivariate PDE systems (Eq. A.1).

At last, new Monte-Carlo samples are drawn from the posterior GMM (Eq. B.2) and dynamically evolved using the DO evolution Eqs. (A.5) until new observations come in and the filtering process is repeated.

Appendix C. State Augmentation

To simultaneously estimate the uncertain parameters and states, we employ state augmentation [50]. We start by decomposing the stochastic regular parameters ($\boldsymbol{\theta}(\omega) \in \mathbb{R}^{N_\theta}$), special parameters ($\boldsymbol{\alpha}(\omega) \in \mathbb{R}^{N_\alpha}$ and $\boldsymbol{\beta}(\omega) \in \mathbb{R}^{N_\beta}$), and expansion coefficients ($\boldsymbol{\gamma}(\omega) \in \mathbb{R}^{N_\gamma}$) into their means and uncertain parts,

$$\begin{aligned} \boldsymbol{\theta}(\omega) &= \bar{\boldsymbol{\theta}} + \mathfrak{D}^{\boldsymbol{\theta}}(\omega) \,, \\ \boldsymbol{\alpha}(\omega) &= \bar{\boldsymbol{\alpha}} + \mathfrak{D}^{\boldsymbol{\alpha}}(\omega) \,, \\ \boldsymbol{\beta}(\omega) &= \bar{\boldsymbol{\beta}} + \mathfrak{D}^{\boldsymbol{\beta}}(\omega) \,, \\ \boldsymbol{\gamma}(\omega) &= \bar{\boldsymbol{\gamma}} + \mathfrak{D}^{\boldsymbol{\gamma}}(\omega) \,. \end{aligned} \tag{C.1}$$

The augmented state vector can be written as,

$$\boldsymbol{\Phi}_{\text{aug}}(t;\omega) = \begin{bmatrix} \boldsymbol{\theta}(\omega) \\ \boldsymbol{\alpha}(\omega) \\ \boldsymbol{\beta}(\omega) \\ \boldsymbol{\gamma}(\omega) \\ \boldsymbol{\Phi}(t;\omega) \end{bmatrix} \in \mathbb{R}^{N_\phi N_x + N_\theta + N_\alpha + N_\beta + N_\gamma} \,. \tag{C.2}$$

Now, let us write the DO decomposition for this new augmented system. We define a new coefficient matrix in which each parameter uncertainty amounts to an additional scalar stochastic coefficient,

$$\boldsymbol{DY}(t;\omega) = \left[\mathfrak{D}^{\boldsymbol{\theta}}(\omega) | \mathfrak{D}^{\boldsymbol{\alpha}}(\omega) | \mathfrak{D}^{\boldsymbol{\beta}}(\omega) | \mathfrak{D}^{\boldsymbol{\gamma}}(\omega) | \boldsymbol{Y}(t;\omega)\right] \in \mathbb{R}^{N_s + N_\theta + N_\alpha + N_\beta + N_\gamma} \,, \tag{C.3}$$

a new modes matrix with parameters having unit modes,

$$\tilde{\boldsymbol{\Phi}}_{\text{aug}}(t) = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{\Phi}}(t) \end{bmatrix} \in \mathbb{R}^{(N_\phi N_x + N_\theta + N_\alpha + N_\beta + N_\gamma) \times (N_s + N_\theta + N_\alpha + N_\beta + N_\gamma)} \,, \tag{C.4}$$

and a new augmented mean vector,

$$\bar{\boldsymbol{\Phi}}_{\text{aug}}(t) = \begin{bmatrix} \bar{\boldsymbol{\theta}} \\ \bar{\boldsymbol{\alpha}} \\ \bar{\boldsymbol{\beta}} \\ \bar{\boldsymbol{\gamma}} \\ \bar{\boldsymbol{\Phi}}(t) \end{bmatrix} \in \mathbb{R}^{N_\phi N_x + N_\theta + N_\alpha + N_\beta + N_\gamma} \,. \tag{C.5}$$

Thus, the DO decomposition of the augmented state is given by,

$$\begin{aligned} \boldsymbol{\Phi}_{\text{aug}}(t;\omega) &= \bar{\boldsymbol{\Phi}}_{\text{aug}}(t) + \sum_{i=1}^{N_s + N_\theta + N_\alpha + N_\beta + N_\gamma} \tilde{\boldsymbol{\Phi}}_{\text{aug},i}(t) DY_i(t;\omega) \\ &= \bar{\boldsymbol{\Phi}}_{\text{aug}}(t) + \tilde{\boldsymbol{\Phi}}_{\text{aug}}(t) \boldsymbol{DY}(t;\omega) \,. \end{aligned} \tag{C.6}$$

We can also define the augmented observation model as,

$$\begin{aligned} \boldsymbol{\mathcal{Y}} &= \begin{bmatrix} \boldsymbol{0} & \boldsymbol{H} \end{bmatrix} \boldsymbol{\Phi}_{\text{aug}} + \boldsymbol{V}, \qquad \boldsymbol{V} \sim \mathcal{N}(0, \boldsymbol{R}) \\ &= \boldsymbol{H}_{\text{aug}} \boldsymbol{\Phi}_{\text{aug}} + \boldsymbol{V} \,, \end{aligned} \tag{C.7}$$

where $\boldsymbol{H}$ is the original observation matrix, and $\boldsymbol{H}_{\text{aug}} \in \mathbb{R}^{N_y \times (N_\phi N_x + N_\theta + N_\alpha + N_\beta + N_\gamma)}$ the augmented observation matrix, while $\boldsymbol{\Phi}_{\text{aug}}$ is the augmented state ensemble.

We can consider the above augmented state vector as forecast for time t_k, and employ the GMM-DO filter (Appendix B) to obtain joint posterior distributions of all parameters and state variables. The GMM fit is completed jointly for the (normalized) parameter realizations and DO stochastic coefficients realizations of the discrete state variables. If the observations, commonly of state variables, are informative about some parameter values, the pdf of these parameters will be updated by the Bayesian update, jointly with the pdf of the DO stochastic coefficients of the discrete state variables.

Appendix D. Notations

We define the notation used throughout the paper in Table D.2, without repeating definitions already given in Table 1.

Table D.2: Notation Compendium

General		
$\boldsymbol{x}$	$\in \mathbb{R}^n$	Spatial coordinate vector
t	$\in \mathbb{R}$	Temporal coordinate
T	$\in \mathbb{R}$	Total simulation time
$\mathcal{D}$		Simulation domain
$\partial\mathcal{D}$		Simulation domain boundary
ω		Realization index
Ω		Measurable sample space
N_x	$\in \mathbb{N}$	Size of the discretized domain
$\boldsymbol{u}$	$\in \mathbb{R}^{N_v}$ or $\mathbb{R}^2$	General state vector or velocity field
$\boldsymbol{u}_o$	$\in \mathbb{R}^{N_v}$ or $\mathbb{R}^2$	Initial condition of $\boldsymbol{u}$
N_v	$\in \mathbb{N}$	Number of state variables
$N_v(i)$	$\in \mathbb{N}$	Number of state variables in the i^{th} candidate model of different complexity
$\{u_1^i, ..., u_{N_v(i)}^i\}$	$\in \mathbb{R}^{N_v(i)}$	State variables for the i^{th} candidate model of different complexity
$\boldsymbol{U}$	$\in \mathbb{R}^{N_v N_x}$	Discretized state vector of $\boldsymbol{u}$
$\boldsymbol{b}$		Boundary values
$\mathcal{M}$		Candidate model
N_m	$\in \mathbb{N}$	Number of candidate models
$\boldsymbol{\theta}$	$\in \mathbb{R}^{N_\theta}$	Uncertain regular parameters
N_θ	$\in \mathbb{N}$	Number of uncertain regular parameters
$\boldsymbol{\phi}$	$\in \mathbb{R}^{N_\phi}$	Biological tracer fields
$\boldsymbol{\phi}_o$	$\in \mathbb{R}^{N_\phi}$	Initial condition of $\boldsymbol{\phi}$
N_ϕ	$\in \mathbb{N}$	Number of biological tracers
p	$\in \mathbb{R}^2$	Pressure field
u and v	$\in \mathbb{R}$	Horizontal and vertical velocity
x and z	$\in \mathbb{R}$	Horizontal and vertical direction
i, j, m, and n	$\in \mathbb{N}$	Miscellaneous index
Model learning methodology		
$\boldsymbol{\alpha}$	$\in \mathbb{R}^{N_\alpha}$	Special stochastic parameters for combining candidate models with different functional forms
N_α	$\in \mathbb{N}$	Number of special stochastic parameters α_k's
$\boldsymbol{\beta}$	$\in \mathbb{R}^{N_\beta}$	Special stochastic parameters for combining candidate models of different complexities
N_β	$\in \mathbb{N}$	Number of special stochastic parameters β_k's
$\mathcal{H}$		Range of values taken by the state variable
I_i		Interval with non-zero measure
N_I	$\mathbb{N}$	Number of intervals
$\boldsymbol{\gamma}$	$\in \mathbb{R}^{N_\gamma}$	Stochastic expansion coefficient
N_γ	$\in \mathbb{N}$	Number of stochastic expansion coefficient γ_k's
k	$\in \mathbb{N}$	Index for uncertain regular parameters, special stochastic parameters, and stochastic expansion coefficients
DO evolution equations		
N_s	$\in \mathbb{N}$	Number of DO modes
N_r	$\in \mathbb{N}$	Number of Monte-Carlo samples
$\bar{\boldsymbol{\phi}}$	$\in \mathbb{R}^{N_\phi}$	Biological tracer DO mean
$\bar{\boldsymbol{\Phi}}$	$\in \mathbb{R}^{N_\phi N_x}$	Discretized biological tracer DO mean $\bar{\boldsymbol{\phi}}$
$\tilde{\boldsymbol{\phi}}_i$	$\in \mathbb{R}^{N_\phi}$	i^{th} biological tracer DO mode
$\tilde{\boldsymbol{\Phi}}_i$	$\in \mathbb{R}^{N_\phi N_x \times N_s}$	Discretized biological tracer DO modes matrix
Y_i	$\in \mathbb{R}^{N_s}$	i^{th} DO stochastic coefficient
$\boldsymbol{Y}$	$\in \mathbb{R}^{N_r \times N_s}$	DO stochastic coefficient matrix
$\bar{\boldsymbol{\theta}}$, $\bar{\boldsymbol{\alpha}}$, $\bar{\boldsymbol{\beta}}$, and $\bar{\boldsymbol{\gamma}}$	$\in \mathbb{R}^{N_\theta}$, $\mathbb{R}^{N_\alpha}$, $\mathbb{R}^{N_\beta}$, and $\mathbb{R}^{N_\gamma}$ resp.	Mean vectors of uncertain parameters

$\boldsymbol{\mathfrak{D}^{\theta}}$, $\boldsymbol{\mathfrak{D}^{\alpha}}$, $\boldsymbol{\mathfrak{D}^{\beta}}$, and $\boldsymbol{\mathfrak{D}^{\gamma}}$	$\in \mathbb{R}^{N_\theta}$, $\mathbb{R}^{N_\alpha}$, $\mathbb{R}^{N_\beta}$, and $\mathbb{R}^{N_\gamma}$ resp.	Mean removed (deviation) part of uncertain parameters
$\sigma_{nd,\bullet}$	$\in \mathbb{R}$	Weight of different state variables in inner-product computation
GMM-DO filter		
$\boldsymbol{\mathcal{Y}}$	$\in \mathbb{R}^{N_y}$	Observation vector
$\boldsymbol{y}$	$\in \mathbb{R}^{N_y}$	Observation vector realization
N_y	$\in \mathbb{N}$	Number of observations
$\boldsymbol{H}$	$\in \mathbb{R}^{N_y \times N_v N_x}$ or $\mathbb{R}^{N_y \times N_\phi N_x}$	Linear observation matrix
$\boldsymbol{V}$	$\in \mathbb{R}^{N_y}$	Measurement noise vector
$\boldsymbol{R}$	$\in \mathbb{R}^{N_y \times N_y}$	Covariance matrix of measurement noise
N_{GMM}	$\in \mathbb{N}$	Number of Gaussian mixture model (GMM) components
$\pi^{\bullet}_{\bullet,j}$	$\in [0,1]$	j^{th} GMM component weight
$\boldsymbol{\mu}^{\bullet}_{\bullet,j}$	$\in \mathbb{R}^{N_s}$	j^{th} GMM component mean vector
$\boldsymbol{\Sigma}^{\bullet}_{\bullet,j}$	$\in \mathbb{R}^{N_s \times N_s}$	j^{th} GMM component covariance matrix
$\boldsymbol{K}_j$	$\in \mathbb{R}^{N_v N_x \times N_y}$ or $\mathbb{R}^{N_\phi N_x \times N_y}$	j^{th} Kalman gain matrix
$p_{\bullet}(\bullet)$	$\in \mathbb{R}$	Probability distribution value
$\tilde{\boldsymbol{y}}$	$\in \mathbb{R}^{N_y}$	Transformed observation vector realization
$\tilde{\boldsymbol{H}}$	$\in \mathbb{R}^{N_y \times N_s}$	Transformed linear observation matrix
$\tilde{\boldsymbol{K}}_j$	$\in \mathbb{R}^{N_s \times N_y}$	Transformed j^{th} Kalman gain matrix
$\hat{\boldsymbol{\mu}}^{\bullet}_{\bullet,j}$	$\in \mathbb{R}^{N_s}$	Intermediate j^{th} GMM component mean vector
Reduced dimension state augmentation		
$\boldsymbol{\Phi}_{\text{aug}}$	$\in \mathbb{R}^{\substack{N_\phi N_x+N_\theta+N_\alpha \\ +N_\beta+N_\gamma}}$	Augmented discretized state vector
$\bar{\boldsymbol{\Phi}}_{\text{aug}}$	$\in \mathbb{R}^{\substack{N_\phi N_x+N_\theta+N_\alpha \\ +N_\beta+N_\gamma}}$	Augmented discretized DO mean vector
$\tilde{\boldsymbol{\Phi}}_{\text{aug}}$	$\in \mathbb{R}^{\substack{N_\phi N_x+N_\theta+N_\alpha \\ +N_\beta+N_\gamma}}$	Augmented discretized DO modes matrix
$\boldsymbol{DY}$	$\in \mathbb{R}^{\substack{N_s+N_\theta+N_\alpha \\ +N_\beta+N_\gamma}}$	Augmented DO stochastic coefficient matrix
$\boldsymbol{H}_{\text{aug}}$	$\in \mathbb{R}^{\substack{N_y \times \\ (N_\phi N_x+N_\theta+N_\alpha+N_\beta+N_\gamma)}}$	Augmented linear observation matrix
$\boldsymbol{I}$	$\in \mathbb{R}^{\substack{(N_\theta+N_\alpha+N_\beta+N_\gamma)\times \\ (N_\theta+N_\alpha+N_\beta+N_\gamma)}}$	Identity matrix
$\boldsymbol{0}$		Matrix of zeroes of appropriate size
Operators, functions, and indicators		
$\mathcal{L}[\bullet]$		Functional form of known dynamics
$\widehat{\mathcal{L}}[\bullet]$		Unknown dynamics belonging to a set of candidate functional forms
$\{\widehat{\mathcal{L}}_1[\bullet], ..., \widehat{\mathcal{L}}_{N_m}[\bullet]\}$		Set of candidate functional forms
$\widetilde{\mathcal{L}}[\bullet]$		Functional form of completely unknown dynamics
$\{\mathcal{L}^i_1[\bullet], ..., \mathcal{L}^i_{N_v(i)}[\bullet]\}$		Right-hand-sides of i^{th} candidate model of different complexity
$\mathcal{N}(\bullet,\bullet)$		Gaussian distribution
$\Psi_k(\bullet)$		k^{th} Linear function
$\{S^{\phi^1}(\bullet), ..., S^{\phi^{N_\phi}}(\bullet)\}$		Biological reaction terms
$\widehat{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet)$		Unknown biological reaction terms belonging to a set of candidate functional forms
$\widetilde{\boldsymbol{S}}^{\boldsymbol{\phi}}(\bullet)$		Completely unknown biological reaction terms
$\nabla.(\bullet)$		Gradient operator
$\nabla^2(\bullet)$		Diffusion operator
$\langle\bullet,\bullet\rangle$		i^{th} Spatial inner-product
$\mathbb{E}[\bullet]$		Expectation
$C_{\bullet,\bullet}$		Cross-covariance
$\mathcal{B}[\bullet]$		Boundary condition operator
$\boldsymbol{V_S}$		DO Subspace
$(\bullet)^f$		Prior
$(\bullet)^a$		Posterior

References

[1] C. Lalli, T. R. Parsons, Biological Oceanography: An Introduction, Elsevier Butterworth-Heinemann, 1997.

[2] W. Fennel, T. Neumann, Introduction to the Modelling of Marine Ecosystems:(with MATLAB programs on accompanying CD-ROM), Vol. 72 of Oceanography, Elsevier, 2014.
[3] P. J. S. Franks, NPZ models of plankton dynamics: their construction, coupling to physics, and application, Journal of Oceanography 58 (2) (2002) 379–387.
[4] B. A. Ward, M. A. M. Friedrichs, T. R. Anderson, A. Oschlies, Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, Journal of Marine Systems 81 (1) (2010) 34–43.
[5] K. L. Denman, Modelling planktonic ecosystems: parameterizing complexity, Progress in Oceanography 57 (3-4) (2003) 429–452.
[6] P. J. S. Franks, J. S. Wroblewski, G. R. Flierl, Behavior of a simple plankton model with food-level acclimation by herbivores, Marine Biology 91 (1) (1986) 121–129.
[7] G. Flierl, D. J. McGillicuddy, Mesoscale and submesoscale physical-biological interactions, The sea 12 (2002) 113–185.
[8] C. S. Davis, J. H. Steele, Biological/physical modeling of upper ocean processes, Tech. rep., Woods Hole Oc. Inst. (1994).
[9] M. J. R. Fasham, H. W. Ducklow, S. M. McKelvie, A nitrogen-based model of plankton dynamics in the oceanic mixed layer, Journal of Marine Research 48 (3) (1990) 591–639.
[10] Ş. T. Beşiktepe, P. F. J. Lermusiaux, A. R. Robinson, Coupled physical and biogeochemical data-driven simulations of Massachusetts Bay in late summer: Real-time and post-cruise data assimilation, Journal of Marine Systems 40–41 (2003) 171–212. doi:10.1016/S0924-7963(03)00018-6.
[11] J. W. Baretta, W. Ebenhöh, P. Ruardij, The European Regional Seas Ecosystem Model, a complex marine ecosystem model, Netherlands Journal of Sea Research 33 (3) (1995) 233–246.
[12] J. W. Baretta, Preface to the European Regional Seas Ecosystem Model II, Journal of Sea Research 38 (3) (1997) 169–171.
[13] J. C. Blackford, J. I. Allen, F. J. Gilbert, Ecosystem dynamics at six contrasting sites: a generic modelling study, Journal of Marine Systems 52 (1) (2004) 191–215.
[14] 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, Vol. 3038 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2004, pp. 685–692. doi:10.1007/978-3-540-24688-6_89.
[15] P. F. J. Lermusiaux, Adaptive modeling, adaptive data assimilation and adaptive sampling, Physica D: Nonlinear Phenomena 230 (1) (2007) 172–196. doi:10.1016/j.physd.2007.02.014.
[16] 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, Vol. 12 of The Sea, John Wiley and Sons, New York, 2002, chapter 12, pp. 475–536.
[17] M. Dowd, E. Jones, J. Parslow, A statistical overview and perspectives on data assimilation for marine biogeochemical models, Environmetrics 25 (4) (2014) 203–213.
[18] M. A. M. Friedrichs, et al., Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups, Journal of Geophysical Research: Oceans 112 (C8) (2007).
[19] S. N. Losa, G. A. Kivman, V. A. Ryabchenko, Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data?, Journal of Marine Systems 45 (1) (2004) 1–20.
[20] T. Toyoda, et al., Improved state estimations of lower trophic ecosystems in the global ocean based on a green's function approach, Progress in Oceanography 119 (2013) 90–107.
[21] J. P. Mattern, M. Dowd, K. Fennel, Sequential data assimilation applied to a physical–biological model for the bermuda atlantic time series station, Journal of Marine Systems 79 (1) (2010) 144–156.
[22] J. I. Allen, M. Eknes, G. Evensen, An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea, in: Annales Geophysicae, Vol. 21, 2003, pp. 399–411.
[23] L. J. Natvik, G. Evensen, Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments, Journal of Marine Systems 40 (2003) 127–153.
[24] M. Doron, P. Brasseur, J.-M. Brankart, Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments, Journal of Marine Systems 87 (3) (2011) 194–207.
[25] E. Jones, J. Parslow, L. Murray, A bayesian approach to state and parameter estimation in a phytoplankton-zooplankton model, Australian Meteorological and Oceanographic Journal 59 (SP) (2010) 7–16.
[26] E. Giricheva, Aggregation in ecosystem models and model stability, Progress in Oceanography 134 (2015) 190–196.
[27] B. A. Ward, et al., When is a biogeochemical model too complex? objective model reduction and selection for north atlantic time-series sites, Progress in Oceanography 116 (2013) 49–65.
[28] C. Evangelinos, R. Chang, P. F. J. Lermusiaux, N. M. Patrikalakis, Rapid real-time interdisciplinary ocean forecasting using adaptive sampling and adaptive modeling and legacy codes: Component encapsulation using XML, in: Computational Science–ICCS 2003, Springer, 2003, pp. 375–384. doi:10.1007/3-540-44864-0_39.
[29] 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, Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA (May 2004).
[30] 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, Special Issue on the Philippine Straits Dynamics Experiment. doi:10.5670/oceanog.2011.05.
[31] S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS 113 (15) (2016) 3932–3937.
[32] S. Rudy, A. Alla, S. L. Brunton, J. N. Kutz, Data-driven identification of parametric partial differential equations, SIAM Journal on Applied Dynamical Systems 18 (2) (2019) 643–660.

[33] D. A. Messenger, D. M. Bortz, Weak sindy for partial differential equations, Journal of Computational Physics (2021) 110525.
[34] 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., Vol. 12312 of Lecture Notes in Computer Science, Springer, Cham, 2020, pp. 208–216. doi:10.1007/978-3-030-61725-7_25.
[35] R. K. Niven, A. Mohammad-Djafari, L. Cordier, M. Abel, M. Quade, Bayesian identification of dynamical systems, in: Multidisciplinary Digital Publishing Institute Proceedings, Vol. 33, 2020, p. 33.
[36] A. Gupta, P. F. J. Lermusiaux, Neural closure models for dynamical systems, Proceedings of The Royal Society A 477 (2252) (2021) 1–29. doi:10.1098/rspa.2020.1004.
[37] M. Maslyaev, A. Hvatov, A. Kalyuzhnaya, Data-driven pde discovery with evolutionary approach.(2019), arXiv preprint arXiv:1903.08011.
[38] M. Bassenne, A. Lozano-Durán, Computational model discovery with reinforcement learning, arXiv preprint arXiv:2001.00008 (2019).
[39] G. Novati, H. L. de Laroussilhe, P. Koumoutsakos, Automating turbulence modelling by multi-agent reinforcement learning, Nature Machine Intelligence 3 (1) (2021) 87–96.
[40] Y. Wang, Z. Shen, Z. Long, B. Dong, Learning to discretize: solving 1d scalar conservation laws via deep reinforcement learning, arXiv preprint arXiv:1905.11079 (2019).
[41] M. Raissi, G. E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, Journal of Computational Physics 357 (2018) 125–141.
[42] 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).
[43] P. Lu, P. F. J. Lermusiaux, Bayesian learning of stochastic dynamical models, Physica D: Nonlinear Phenomena 427 (2021) 133003. doi:10.1016/j.physd.2021.133003.
[44] T. Sondergaard, P. F. J. Lermusiaux, Data assimilation with Gaussian Mixture Models using the Dynamically Orthogonal field equations. Part I: Theory and scheme., Monthly Weather Review 141 (6) (2013) 1737–1760. doi:10.1175/MWR-D-11-00295.1.
[45] T. Sondergaard, P. F. J. Lermusiaux, Data assimilation with Gaussian Mixture Models using the Dynamically Orthogonal field equations. Part II: Applications., Monthly Weather Review 141 (6) (2013) 1761–1785. doi:10.1175/MWR-D-11-00296.1.
[46] M. Bayes, M. Price, 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., Philosophical Transactions (1683-1775) 53 (1763) 370–418. doi:10.1098/rstl.1763.0053.
[47] T. P. Sapsis, P. F. J. Lermusiaux, Dynamically orthogonal field equations for continuous stochastic dynamical systems, Physica D: Nonlinear Phenomena 238 (23–24) (2009) 2347–2360. doi:10.1016/j.physd.2009.09.017.
[48] T. P. Sapsis, P. F. J. Lermusiaux, Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty, Physica D: Nonlinear Phenomena 241 (1) (2012) 60–76. doi:10.1016/j.physd.2011.10.001.
[49] F. Feppon, P. F. J. Lermusiaux, A geometric approach to dynamical model-order reduction, SIAM Journal on Matrix Analysis and Applications 39 (1) (2018) 510–538. doi:10.1137/16M1095202.
[50] A. Gelb, Applied optimal estimation, MIT Press, 1974.
[51] P. E. Hart, D. G. Stork, R. O. Duda, Pattern classification, Wiley Hoboken, 2000.
[52] P. A. Newberger, J. S. Allen, Y. H. Spitz, Analysis and comparison of three ecosystem models, Journal of Geophysical Research: Oceans (1978–2012) 108 (C3) (2003).
[53] P. J. Haley, Jr., P. F. J. Lermusiaux, Multiscale two-way embedding schemes for free-surface primitive equations in the "Multidisciplinary Simulation, Estimation and Assimilation System", Ocean Dynamics 60 (6) (2010) 1497–1537. doi:10.1007/s10236-010-0349-4.
[54] P. J. Haley, Jr., A. Agarwal, P. F. J. Lermusiaux, Optimizing velocities and transports for complex coastal regions and archipelagos, Ocean Modeling 89 (2015) 1–28. doi:10.1016/j.ocemod.2015.02.005.
[55] J. C. McWilliams, The nature and consequences of oceanic eddies, Ocean modeling in an eddying regime 177 (2008) 5–15.
[56] M. W. Hecht, H. Hasumi, Ocean modeling in an eddying regime, Vol. 177, John Wiley & Sons, 2013.
[57] J. H. Ferziger, M. Perić, R. L. Street, Computational methods for fluid dynamics, Vol. 3, Springer, 2002.
[58] D. N. Subramani, P. F. J. Lermusiaux, Energy-optimal path planning by stochastic dynamically orthogonal level-set optimization, Ocean Modeling 100 (2016) 57–77. doi:10.1016/j.ocemod.2016.01.006.
[59] D. N. Subramani, Q. J. Wei, P. F. J. Lermusiaux, Stochastic time-optimal path-planning in uncertain, strong, and dynamic flows, Computer Methods in Applied Mechanics and Engineering 333 (2018) 218–237. doi:10.1016/j.cma.2018.01.004.
[60] M. Branicki, A. J. Majda, Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabilities, Communications in mathematical sciences 11 (1) (2013) 55–103.
[61] D. McGillicuddy, D. Lynch, A. Moore, W. Gentleman, C. Davis, C. Meise, An adjoint data assimilation approach to diagnosis of physical and biological controls on pseudocalanus spp. in the gulf of maine–georges bank region, Fisheries Oceanography 7 (3-4) (1998) 205–218.
[62] P. F. J. Lermusiaux, Evolving the subspace of the three-dimensional multiscale ocean variability: Massachusetts Bay, Journal of Marine Systems 29 (1) (2001) 385–422. doi:10.1016/S0924-7963(01)00025-2.
[63] J. Pineda, V. Starczak, J. C. da Silva, K. Helfrich, M. Thompson, D. Wiley, Whales and waves: Humpback whale foraging response and the shoaling of internal waves at stellwagen bank, J. of Geophysical Res.: Oceans 120 (4) (2015) 2555–2570.
[64] R. Tian, C. Chen, J. Qi, R. Ji, R. C. Beardsley, C. Davis, Model study of nutrient and phytoplankton dynamics in the gulf

of maine: patterns and drivers for seasonal and interannual variability, ICES J. of Marine Science 72 (2) (2015) 388–402.
[65] A. Pershing, et al., Temperature and circulation conditions in the gulf of maine in 2050 and their expected impacts, Scientific scenario paper, Gulf of Maine 2050 International Symposium, 2019.
[66] T. L. Silva, State of the science report: An addendum to the stellwagen bank national marine sanctuary 2020 condition report (2021).
[67] 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
[68] B. Van Leer, Towards the ultimate conservative difference scheme. iv. a new approach to numerical convection, Journal of computational physics 23 (3) (1977) 276–299.
[69] A. Gupta, Scientific machine learning for dynamical systems: Theory and applications to fluid flow and ocean ecosystem modeling, Ph.D. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts (Sep. 2022).
[70] M. P. Ueckermann, P. F. J. Lermusiaux, T. P. Sapsis, Numerical schemes for dynamically orthogonal equations of stochastic fluid and ocean flows, Journal of Computational Physics 233 (2013) 272–294. doi:10.1016/j.jcp.2012.08.041.
[71] F. Feppon, P. F. J. Lermusiaux, Dynamically orthogonal numerical schemes for efficient stochastic advection and Lagrangian transport, SIAM Review 60 (3) (2018) 595–625. doi:10.1137/16M1109394.
[72] P. F. J. Lermusiaux, On the mapping of multivariate geophysical fields: Sensitivities to size, scales, and dynamics, Journal of Atmospheric and Oceanic Technology 19 (10) (2002) 1602–1637. doi:10.1175/1520-0426(2002)019<1602:OTMOMG>2.0.CO;2.
[73] L. Bengtsson, M. Ghil, E. Källén, Dynamic meteorology: data assimilation methods, Springer, 1981.
[74] K. Ide, M. Ghil, Extended kalman filtering for vortex systems. part 1: Methodology and point vortices, Dynamics of Atmospheres and Oceans 27 (1-4) (1998) 301–332.
[75] K. Ide, M. Ghil, Extended kalman filtering for vortex systems. part ii: Rankine vortices and observing-system design, Dynamics of Atmospheres and Oceans 27 (1-4) (1998) 333–350.
[76] P. F. J. Lermusiaux, Data assimilation via Error Subspace Statistical Estimation, part II: Mid-Atlantic Bight shelfbreak front simulations, and ESSE validation, Monthly Weather Review 127 (7) (1999) 1408–1432. doi:10.1175/1520-0493(1999)127<1408:DAVESS>2.0.CO;2.
[77] T. Lolla, P. J. Haley, Jr., P. F. J. Lermusiaux, Time-optimal path planning in dynamic flows using level set equations: Realistic applications, Ocean Dynamics 64 (10) (2014) 1399–1417. doi:10.1007/s10236-014-0760-3.
[78] 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. doi:10.23919/OCEANS40490.2019.8962892.
[79] J. A. Bilmes, et al., A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, International Computer Science Institute 4 (510) (1998) 126.
[80] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, Signal Processing Magazine, IEEE 21 (4) (2004) 36–47.
[81] G. Wornell, Inference and information. lecture notes for mit course 6.437 in spring 2013 (May 2016).
[82] J. Lin, Bayesian learning for high-dimensional nonlinear systems: Methodologies, numerics and applications to fluid flows, Ph.D. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts (Sep. 2020).
[83] A. Gupta, Bayesian inference of obstacle systems and coupled biogeochemical-physical models, Master's thesis, Indian Institute of Technology Kanpur, Kanpur, India (2016).
[84] P. F. J. Lermusiaux, et al., A future for intelligent autonomous ocean observing systems, Journal of Marine Research 75 (6) (2017) 765–813, the Sea. Volume 17, The Science of Ocean Prediction, Part 2. doi:10.1357/002224017823524035.
[85] P. F. J. Lermusiaux, et al., Optimal planning and sampling predictions for autonomous and Lagrangian platforms and sensors in the northern Arabian Sea, Oceanography 30 (2) (2017) 172–185, special issue on Autonomous and Lagrangian Platforms and Sensors (ALPS). doi:10.5670/oceanog.2017.242.
[86] A. Gupta, W. H. Ali, P. F. J. Lermusiaux, Boundary conditions for stochastic DO equations, MSEAS Report, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA (2016).
[87] M. J. Humara, Stochastic acoustic ray tracing with dynamically orthogonal equations, Master's thesis, Massachusetts Institute of Technology, Joint Program in Applied Ocean Science and Engineering, Cambridge, Massachusetts (May 2020).
[88] M. J. Humara, W. H. Ali, A. Charous, M. Bhabra, P. F. J. Lermusiaux, Stochastic acoustic ray tracing with dynamically orthogonal differential equations, in: OCEANS 2022 IEEE/MTS, IEEE, 2022, sub-judice.
[89] A. Charous, M. J. Humara, W. H. Ali, M. S. Bhabra, A. Gupta, P. F. Lermusiaux, Dynamically orthogonal ray equations with adaptive reclustering, The Journal of the Acoustical Society of America 150 (4) (2021) A209–A209. doi:10.1121/10.0008139.
[90] P. F. J. Lermusiaux, D. G. M. Anderson, C. J. Lozano, On the mapping of multivariate geophysical fields: Error and variability subspace estimates, Quarterly Journal of the Royal Meteorological Society 126 (565) (2000) 1387–1429. doi:10.1256/smsqj.56509.
[91] D. N. Subramani, Probabilistic regional ocean predictions: Stochastic fields and optimal planning, Ph.D. thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts (Feb. 2018).
[92] D. Subramani, P. F. J. Lermusiaux, Probabilistic ocean predictions with dynamically-orthogonal primitive equationsIn preparation (2021).