Gupta, A., P.J. Haley, D.N. Subramani, and P.F.J. Lermusiaux, 2019. Fish Modeling and Bayesian Learning for the Lakshadweep Islands.
In: OCEANS '19 MTS/IEEE Seattle, 27-31 October 2019, doi: 10.23919/OCEANS40490.2019.8962892

Fish Modeling and Bayesian Learning for the Lakshadweep Islands

Abhinav Gupta, Patrick J. Haley, Deepak N. Subramani, Pierre F. J. Lermusiaux†
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA-02139
†Corresponding Author: pierrel@mit.edu

***Abstract*—In fish modeling, a significant amount of uncertainty exists in the parameter values, parameterizations, functional form of model equations, and even the state variables themselves. This is due to the complexity and lack of understanding of the processes involved, as well as the measurement sparsity. These challenges motivate the present proof-of-concept study to simultaneously learn and estimate the state variables, parameters, and model equations from sparse observations. We employ a novel dynamics-based Bayesian learning framework for high-dimensional, coupled fish-biogeochemical-physical partial-differential equations (PDEs) models, allowing the simultaneous inference of the augmented state variables and parameters. After reviewing the status of ecosystem modeling in the coastal oceans, we first complete a series of PDE-based learning experiments that showcase capabilities for fish-biogeochemical-physical model equations and parameters, using nonhydrostatic Boussinesq flows past a seamount. We then showcase realistic ocean primitive-equation simulations and analyses, using fish catch data for the Lakshadweep islands in India. These modeling and learning efforts could improve fisheries management from a standpoint of sustainability and efficiency.**

***Index Terms*—Fish model, Biogeochemical model, Uncertainty quantification, Bayesian learning, SEAPODYM, Gaussian Mixture Models, Dynamically Orthogonal, Ocean Primitive Equations**

I. Introduction

Fisheries are a major industry in the coastal states of India, employing millions of people and contributing to 1.1% of GDP and 5.3% of agricultural GDP. Globally, the Indian fishing industry is the third largest in the world. The total marine fish production is around 3 billion metric tons. Indian waters contain about 2,500 species of finfishes and shellfishes. Among these, there are about 65 commercially important species or groups. In 2004, 52% of these commercially important groups were pelagic and midwater species. In 2006, over 600,000 metric tons of fish were exported, to some 90 countries, earning over $1.8 billion [10]. Increased demand for fish, coupled with unsustainable fishing practices lead to over-exploitation and fast depletion of fish stocks. Coastal fisheries and aquaculture stocks often thrive on very specific water conditions—building capabilities for coastal physical-ecosystem forecasting and monitoring will help ensuring and managing the survival and reproduction of healthy stock. Without sustainable fisheries management and conservation practices in place, there could be dire consequences for the many communities that rely on the ocean for their economic well-being.

Marine ecosystems are very complex, but in broad terms they can be seen as a flow of energy from nutrients, to phytoplankton, to zooplankton, to fish, ending with the nutrients being recycled back into the ecosystem. There does not exist a single generic model that encompasses all the detailed components of a marine food web due to the complexities and multiple scales of natural marine ecosystems. Hence, in general, current research seeks to model small, isolated parts of the food web. These isolated parts need to be properly parameterized in order to link them with other portions of the food web. The ocean ecosystem is commonly divided into two main parts, Lower Trophic Levels (LTL) and Higher Trophic Levels (HTL) including fish. Common ecosystem models are reviewed in Section II.

The LTL models have many sources of uncertainties [27, 41, 39, 22, 61]. Due to the semi-empirical methodology of developing HTL models, there is even more uncertainty associated with the parameters, functional forms, and the level of complexity of fish models. This is an inherent drawback to purely deterministic fish modeling approaches. A set of parameter values may only be valid in certain ocean regions, or may be subject to seasonal variability. Observations are integral to the formation of these models, but are in general used for data fitting in order to find appropriate parameter values or functional forms of these models in offline mode, or for interpolation/extrapolation of data using the models. Due to the lack of comprehensive and reliable fish data [29, 48], learning approaches such as the ANNs are also not very suitable. However, instead, it is possible to combine the existing models with the observational data in a Bayesian inference framework. A variety of data assimilation techniques are indeed used in biogeochemical modeling [54], but most employ standard parameter optimization techniques, where model parameters are calibrated by minimizing misfits between model output and independent observations [21, 44, 67]. Very few studies deal with the simultaneous estimation of state variables parameters and parameterizations [11, 31].

In this paper, we use a PDE-based Bayesian learning framework to showcase a series of learning experiments that simultaneously infer the augmented state variables, parameters, and parameterizations of the fish model SEAPODYM coupled with a LTL dynamical model and a nonhydrostastic varaible-density Boussinesq flow past a seamount. The principled Bayesian learning [46, 45, 23] combines the Dynamically Orthogonal (DO) methodology [55, 15] for reduced-dimension stochastic evolution with a Gaussian mixture model (GMM)-DO filtering scheme [59, 60]. Focusing on the Lakshadweep

978-0-578-57618-3

islands in India, where the main fishery is tuna, we then use the capabilities of our MSEAS primitive equation ocean modeling system [25, 40, 24] to capture the complex oceanic phenomenon in the region. We finally perform a qualitative analysis of these phenomena compared with some fish catch data available for this region [18]. Conclusions and future extensions are discussed in Section V.

II. Marine Ecosystem Modeling

A. Lower Trophic Level

There exist many well-studied models of varying complexities for LTL. A basic model is a simple, 3-component nutrient-phytoplankton-zooplankton (NPZ) model [19, 17, 5]; Franks, 2002 ([20]) provides a thorough review on development of such models. In a workshop on the status of upper layer coupled biological-physical modeling [9], researchers proposed couplings of various mixed layer physical models with NPZ and NPZD (NPZ-Detritus) biological models. Fasham et. al., 1990 ([14]) presented a 7-component model of the annual cycles of plankton dynamics and nitrogen in the ocean. One of the most complex lower-trophic level marine ecosystem model is the European Regional Seas Ecosystem Model (ERSEM, [3, 2, 6]), initially developed for the North Sea.

B. Higher Trophic Level

HTL models vary greatly in how they model fish; some model individual fishes in Lagrangian sense, some are empirical data-based models, and some treat them as aggregate (continuous) biomass and capture more realistic biological interactions and processes. The effective coupling of LTL models and HTL full-life cycle fish models is notoriously challenging, mainly due to the difficult practical and theoretical problems associated with resolving relevant temporal and spatial scales at all biologically meaningful trophic levels. Nonetheless, they can be coupled using different mathematical functions that model source and mortality terms, and close the ecosystem. One of the oldest models developed for fisheries management is the MultiSpecies Virtual Population Analysis (MSVPA) [47]. It solves a system of coupled nonlinear equations in terms of biomass of species and number of fishes belonging to each cohort averaged over large spatial and temporal scales. Parameterization requires stomach-content data of fish and estimates of the number of fish in a particular cohort; this requires lots of hard-to-obtain data, data only sparsely collected for a small number of fish species.

Larval Individual Based Models (IBMs) [26] attempt to model the larval stage, which is in-between LTL and HTL. They start with an ensemble of eggs seeded in the domain, and let them advect and interact with the underlying physical and biogeochemical fields, while mortality is also modeled as a stochastic event which determines whether individual eggs develop to the juvenile stage. A drawback to this approach is the fact that larvae cannot really represent fish population.

A prominent box model is NUMERO.FISH [50, 32], which simulates the daily predator-prey interactions and biogeochemical cycling of phytoplankton, zooplankton, nutrients, and detritus. The FISH model simulates the daily growth and mortality of herring in each of multiple age-classes and is coupled to NEMURO via zooplankton-dependent herring consumption, excretion, and egestion. The FISH model is based on an energy balance equation that equates energy consumed with energy expended and gained.

Interacting Particle Model for Migration of Pelagic Fish [1] models individuals rather than keeping track of the density of a population. Particles look to their neighbors to determine their directional heading at each time step, averaging the neighbors' directional headings to determine their own. This allows the particles to move together as a group. Size spectra models [49] are based on the biomass spectrum theory, which assumes that size governs biological rates and predatory interactions. In size-spectra studies, the whole ecosystem or community is represented by a continuum of biomass and organisms are represented only in terms of their body size. The bio-ecological processes taken into account to model consumers are predation, mortality, assimilation and use of energy for maintenance, growth and reproduction.

Ordinary differential equations based models include the Ecopath with Ecosim (EwE) Ecosystem Modeling Suite [7, 34]. EwE facilitates the construction of a static ecosystem model (Ecopath) that can then be used to run time-based dynamic (Ecosim) and spatial (Ecospace) simulations. Modelling in EwE begins with creating a mass-balance model using Ecopath to obtain a static snapshot of the ecosystem under study. The underlying principle behind the mass balance approach is to balance the energy flow among different trophically linked functional groups by solving a set of simultaneous linear equations (one equation for each functional group).

SEAPODYM (Spatial Ecosystem And Populations Dynamics Model) [38, 37] is an Advection-Diffusion-Reaction (ADR) equation-based model that couples a physical-biological interaction model at basin scales, combining a forage (prey) production model with an age-structured population model of targeted (tuna predator) species. An adult habitat index combines the spatial distribution of tuna forage biomass with a temperature function defined for each species. Young and adult tuna movements are constrained by this adult habitat index while a spawning habitat index is used to constrain the recruitment to environmental conditions. Related ADR models were used for regional fisheries management [53].

Lastly, recent research involves machine learning approaches such as training Artificial-Neural-Nets (ANNs). When using ANNs, typically the output is in the form of catch-per-unit-effort (CPUE) and input includes Sea-Surface-Temperature (SST), Sea-Surface-Height (SSH), gradient of SST, chlorophyll-a, latitude, longitude, time, and other relevant quantities [66, 65, 35, 63].

III. Learning and Modeling Methodology

A. PDE-based Bayesian Learning Machines

A Bayesian learning setting involves choosing a prior probability distribution for the state variables ($\mathbf{X} \in \mathbb{R}^{N_X}$) of interest,

taking into account all sources of uncertainties, $p_{\mathbf{X}}(\mathbf{x})$. Observations ($\mathbf{Y} \in \mathbb{R}^{N_Y}$), with likelihood ($p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$) are used to estimate the posterior probability of the states ($p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$) [4, 57]. In the present problem, the state variable consists of physical, lower-trophic-level and higher-trophic level biological variables, governed by a coupled physical-biological-fish model, along with initial conditions, parameters, and parameterization uncertainties. For the observation likelihood, we assume a linear model given by, $\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V}$; where $\mathbf{H} \in \mathbb{R}^{N_Y \times N_X}$ is the sparse observation matrix; and $\mathbf{V} \in \mathbb{R}^{N_Y}$ is a zero-mean, uncorrelated Gaussian noise with covariance matrix $\mathbf{R} \in \mathbb{R}^{N_Y \times N_Y}$.

1) Physical Model: The physical model is described by the stochastic nonhydrostatic Navier-Stokes equations with a variable-density Boussinesq approximation,

$$\begin{aligned} \nabla.\boldsymbol{u} &= 0 \\ \frac{\partial \boldsymbol{u}}{\partial t} &= -\nabla.(\boldsymbol{u}\boldsymbol{u}) - \nabla P + \Lambda_{Re}(\omega)\nabla^2\boldsymbol{u} + \left[\frac{\rho'}{\rho_o}\right]\mathbf{g} \\ \frac{\partial \rho'}{\partial t} &= -\nabla.(\rho'\boldsymbol{u}) + \kappa\nabla^2\rho' \end{aligned} \tag{1}$$

where $\boldsymbol{u} \equiv \boldsymbol{u}(\boldsymbol{x}, t; \omega)$ is the two-dimensional stochastic velocity field; $P \equiv P(\boldsymbol{x}, t; \omega)$, the stochastic pressure field that contains contributions from the hydrostatic pressure due to the variable density as well as the nonhydrostatic pressure; ρ_o, the mean density; $\rho' \equiv \rho'(\boldsymbol{x}, t; \omega) = \rho(\boldsymbol{x}, t; \omega) - \rho_o$, the density perturbation from the mean; $\boldsymbol{g} = -g\boldsymbol{e_z}$; κ, the constant of kinematic diffusivity; and $\Lambda_{Re}(\omega)$ is here an uncertain parameter equivalent to the inverse of eddy-viscosity (ν_E) Reynolds number ($Re = \frac{UL}{\nu_E}$). This stochastic system belongs to a domain $\boldsymbol{x} : \{x, z\} \in \mathcal{D}$, and ω is a realization index belonging to a measurable sample space Ω. We also consider the density perturbation to be solely a function of temperature $T(\boldsymbol{x}, t; \omega)$, given by the relation, $\rho' = \alpha(T - T_o)$, where α is the coefficient of expansion and T_o is a reference temperature. We specify uncertain initial velocity $\boldsymbol{u}(\boldsymbol{x}, t_{init}; \omega) = \boldsymbol{u}_{init}(\boldsymbol{x}; \omega)$ and temperature $T(\boldsymbol{x}, t_{init}; \omega) = T_{init}(\boldsymbol{x}; \omega)$ fields. The velocity uncertainty is initialized by adding sinusoidal perturbations to a divergence-free domain-confirming potential flow, while for density, different stable stratified profiles are considered for each realization.

2) LTL-Biological Model: The lower-trophic-level biogeochemical model used in the present study is adapted from Newberger et. al., 2003 ([52]). We employ the three-component NPZ model (nutrients ($N(\boldsymbol{x}, t; \omega)$), phytoplankton ($P(\boldsymbol{x}, t; \omega)$), and zooplankton ($Z(\boldsymbol{x}, t; \omega)$)). The NPZ model is given by,

$$\begin{aligned} S^N &= -G\frac{PN}{N + K_u} + \Xi P + \Gamma_1 Z + a(\omega)\Gamma_2 Z^2 \\ &\quad + R_m\gamma Z(1 - \exp^{-\Lambda(\omega)P}) \\ S^P &= G\frac{PN}{N + K_u} - \Xi P - R_m Z(1 - \exp^{-\Lambda P}) \\ S^Z &= R_m(1 - \gamma)Z(1 - \exp^{-\Lambda P}) - \Gamma_1 Z - a(\omega)\Gamma_2 Z^2 \end{aligned} \tag{2}$$

where: G represents the optical model given by, $G = V_m \frac{\alpha I_l}{(V_m^2 + \alpha^2 I_l^2)^{1/2}}$ and $I_l(\boldsymbol{x}) = I_l^o \exp^{k_w z}$. Along with the uncertain state variables, we assume a uncertain Ivlev grazing parameter ($\Lambda(\omega)$) and a special binary stochastic parameter ($a(\omega) \in \{0, 1\}$).

The biogeochemical models are coupled with the physical model using stochastic Advection-Diffusion-Reaction (ADR) equations. Let $\phi^i(\boldsymbol{x}, t; \omega)$, $i = \{1, 2, 3\}$ represent the three stochastic biological tracers, the ADR equations are then,

$$\frac{\partial \phi^i}{\partial t} + \nabla.(\boldsymbol{u}\phi^i) - \kappa\nabla^2\phi^i = S^{\phi^i}(\phi^1, \phi^2, \phi^3, \boldsymbol{x}, t; \omega), \quad i = \{1, 2, 3\}\ , \tag{3}$$

where $\boldsymbol{u}(\boldsymbol{x}, t; \omega)$ is the stochastic velocity field which is derived from the physical model (Eq. 1), Pe is the Peclet number, and $S^{\phi^i}(\phi^1, \phi^2, \phi^3, \boldsymbol{x}, t)$ are the reaction equations for various biological variables which are given by the NPZ biogeochemical model (Eq. 2). The initial conditions for this model are here generated by solving for stable equilibrium solution ($S^{\phi^i} = 0,\ i = \{1, 2, 3\}$) for each of the parameter realizations.

3) Fish Model: We use the spatial ecosystem and population dynamics model (SEAPODYM) based on an ADR formulation that focuses on spatial tuna population dynamics [36]. It couples low-trophic-level (LTL) and high-trophic-level (HTL) biological models. The physical and LTL biological models, given in sections III-A1 and III-A2, respectively, provide estimates of stochastic physical state variables such as velocity ($\boldsymbol{u}(\boldsymbol{x}, t; \omega)$), temperature ($T(\boldsymbol{x}, t; \omega)$), and primary production ($P(\boldsymbol{x}, t; \omega)$). The primary production acts as a source for the forage ($F(\boldsymbol{x}, t; \omega)$), after taking into account the recruitment time and mortality, given by source $S(\boldsymbol{x}, t; \omega) = \frac{1}{\lambda}P(\boldsymbol{x}, t; \omega)\exp^{-m_r T_r(\omega)}$, where λ is the mortality, m_r is a loss coefficient, and $T_r(\omega)$ is the uncertain recruitment time. Thus, the forage field is governed by,

$$\frac{\partial F}{\partial t} + \nabla.(\boldsymbol{u}F) - \kappa\nabla^2 F = -\lambda F + S\ . \tag{4}$$

Tuna tend to favor certain temperature ranges and high food concentrations; the habitat index, given by $I(\boldsymbol{x}, t; \omega) = g(F(\boldsymbol{x}, t; \omega))\phi(T(\boldsymbol{x}, t; \omega) - T_o)$, acts as a spatial field which defines the favorability of location for the fish, here tune. We take $g(F(\boldsymbol{x}, t; \omega)) = F(\boldsymbol{x}, t; \omega)$ and

$$\phi(T(\boldsymbol{x}, t; \omega) - T_o) = 1/(1 + \exp^{-(T(\boldsymbol{x}, t; \omega) - T_o)})\ .$$

Gradients of the habitat index then affect the movement of the fishes. This is captured by defining effective advection velocities, $A_x(\boldsymbol{x}, t; \omega) = u(\boldsymbol{x}, t; \omega) + \chi\frac{\partial I(\boldsymbol{x}, t; \omega)}{\partial x}$ and $A_y(\boldsymbol{x}, t; \omega) = v(\boldsymbol{x}, t; \omega) + \chi\frac{\partial I(\boldsymbol{x}, t; \omega)}{\partial y}$. Hence, the population density ($P_{den}(\boldsymbol{x}, t; \omega)$) of tuna is then governed by an ADR equation with the effective advection,

$$\begin{aligned} \frac{\partial P_{den}}{\partial t} + \nabla.(\mathbf{A}P_{den}) &= \nabla.(D\nabla P_{den})F \\ &\quad - Z(I)P_{den} + R\ , \end{aligned} \tag{5}$$

where D is the diffusion coefficient; $Z(I)$ is a habitat index dependent mortality coefficient given by $\lambda_z \exp^{-\lambda_I I}$; and R

is growth rate. Again the initial uncertainty estimates for $F(\boldsymbol{x}, t; \omega)$ and $P_{den}(\boldsymbol{x}, t; \omega)$ are here found by solving for the stable equilibrium solutions of Eqs. 4 and 5 for each realizations.

4) GMM-DO Bayesian Learning: The above coupled physical-biological-fish PDE model is a high-dimensional stochastic dynamical system encompassing multidisciplinary state variables, rendering uncertainty quantification and Bayesian inference a challenging task. Uncertainty propagation is performed using an efficient reduced-dimension uncertainty quantification method, the Dynamically Orthogonal (DO) equations [55, 56, 16]. To perform inference of the augmented state variables and parameters, we make use of a PDE-based machine learning framework developed by combining the DO method with a Gaussian mixture model (GMM) filtering algorithm [46, 45, 23], and implemented in a finite volume framework [64].

5) Simulated Experiments and Dynamics: The experimental setup for the Bayesian learning consists of a 2-dimensional domain with a seamount representing an idealized sill or strait (Figure 1) that can create an upwelling of the nutrients and thus phytoplankton blooms. The seamount also forces the advection of cold water upward, that leads to a competing effect on the habitat index, thus limiting the tuna to very specific depths. With the nonhydrostatic dynamics, internal waves, recirculations, and other instabilities can also be created downstream of the seamount, leading additional biogeochemical-fish responses. This domain is inspired by the Stellwagen Bank off of Massachusetts. Here, flow occurs from left to right in the positive x-direction over the seamount. For velocity, we apply a Dirichlet boundary condition for the inlet, no-slip for the bump, free-slip at top and bottom, and open boundary at the outlet. For the tracer fields, we use zero-Neumann on all the boundaries. The parameter values associated with the domain are provided in table I.

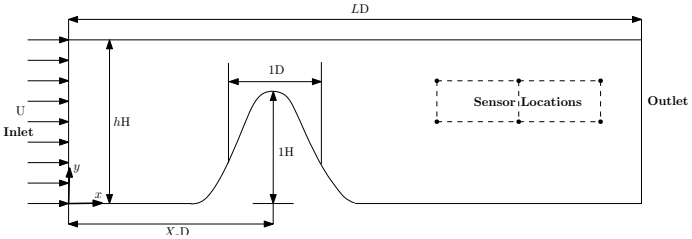


Fig. 1: Two-dimensional spatial domain of the nonhydrostatic flow past a seamount. All lengths are given in terms of length scales (D & H) of the seamount, described by height $H \exp^{(x-X_c)^2}$. Observation locations are marked by dots, downstream of the seamount.

TABLE I: Values of the parameters used in the coupled nonhydrostatic physical-biological-fish model. For the non-dimensionalization, the scalings used are: $N_T = 30\ mmol\, N\, m^{-3}$, $H = 50\ m$, $D = 1\ km$, and time-scale of $12.5\ d$.

Parameters	Values
Domain	
Horiz. length scale, D (km)	1
Vert. length scale, H (m)	50
Domain length, L (non-dim.)	20
Domain height, h (non-dim.)	2
Seamount center, X_c (non-dim.)	7.5
Physical Model	
Inlet velocity, U (cm/s)	1
Eddy viscosity, ν_E (m^2/s)	10
Inverse of Eddy viscosity based Reynolds number (Λ_{Re})	unif(0.01, 1)
Diffusion constants in horiz. and vert., $\mathcal{K}_x$ & $\mathcal{K}_z$ $(m^2/s$; same for all tracers, except fish density)	0.01 & 0.001
Reference temperature, T_o $(^\circ C)$	15
Coeff. of expansion, α $(kg/m^3/^\circ C)$	1.5×10^{-7}
LTL-Biological Model	
Light attenuation coeff. k_w (m^{-1})	0.067
Slope of the P-I curve α $((W\ m^{-2}\ d)^{-1})$	0.025
Surface available radiation I_l^o $(W\ m^{-2})$	158.075
Phyto. maximum uptake rate V_m (d^{-1})	1.5
Half-sat. for phyto. uptake of nutrients, K_u $(mmol\ N\ m^{-3})$	1
Phyto. specific mortality rate Ξ (d^{-1})	0.1
Linear zoo. mortality rate Γ_1 (d^{-1})	0.145
Presence or absence of quad. zoo. mortality term a	unif{0, 1}
Quad. zoo.mortality rate Γ_2 (d^{-1})	0.2
Zoo. max grazing rate R_m (d^{-1})	0.52
Ivlev constant Λ $((mmol\ N\ m^{-3})^{-1})$	unif(0.1, 0.2)
Fraction of zoo. grazing egested γ	0.3
Fish Model	
Forage mortality, λ (yr^{-1})	4.6
Forage loss coeff. m_r (day^{-1})	0.025
Recruitment time, T_r (day)	unif(75, 100)
Fish mortality coeff. parameter, λ_z (day^1)	0.8
Fish mortality coeff. parameter, λ_I (non-dim)	100
Taxis coeff., χ (m/day)	400
Fish recruitment rate, R (kg/m^2/day)	4
Fish diffusion coefficient, D_x & D_y (m^2/s)	0.1 & 0.01
Others	
Number of realizations, N_{MC}	10,000
State being observed	Z
Observation error std. dev. $(\sqrt{\boldsymbol{R}})$	0.05
Number of obs. locations, N_Y	6
Observation start time (non-dim)	3
Time interval between obs. (non-dim)	2
Observation end time (non-dim)	11

B. MSEAS Primitive Equation Ocean Model

The MIT Multidisciplinary Simulation Estimation and Assimilation System (MSEAS) primitive equation (PE) modeling system [25, 42, 24] was used to produce hindcasts for the Lakshadweep region. The Lakshadweep domain (Figure 7) off the southwest coast of India has a 1.5 km horizontal resolution and 70 vertical levels with optimized level depths (i.e., higher resolution near the surface or large vertical derivatives). This resolution was needed to develop and maintain complex layered features (not shown). The bathymetry was obtained from the 15 arc-second SRMT15 data [58]. The sub-tidal initial and boundary conditions were downscaled from 1/12 degree Hybrid Coordinate Ocean Model (HYCOM) analyses [8] via optimization for higher-resolution coastlines and bathymetry [24]. Tidal forcing was computed from the high-resolution TPXO8-Atlas from OSU [12, 13], by reprocessing for higher

resolution bathymetry/coastline and quadratic bottom drag (a nonlinear extension of Logutov and Lermusiaux [43]). The atmospheric fields consisted of the wind stresses, net heat flux and surface fresh water flux from the 1/2 degree NAVGEM 3-hourly analyses [28].

IV. Results

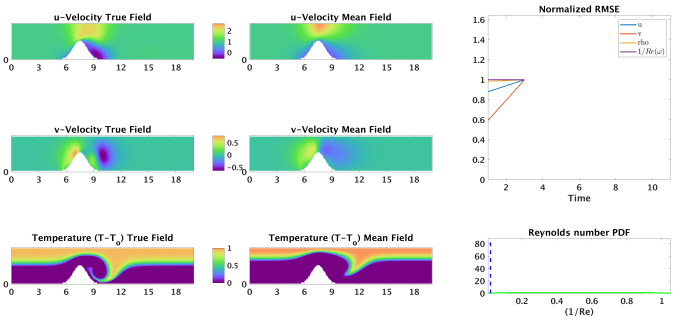
A. Bayesian Learning Experiments

To demonstrate the capabilities of the Bayesian learning framework, we perform simultaneous estimations of state variables, parameters, and parameterizations in the coupled physical-biological-fish model using only very sparse observations. We employ so-called "identical twin experiments" in which observations made from a simulated truth are generated using a deterministic run with a particular set of parameter values which lie within the uncertain realization space.

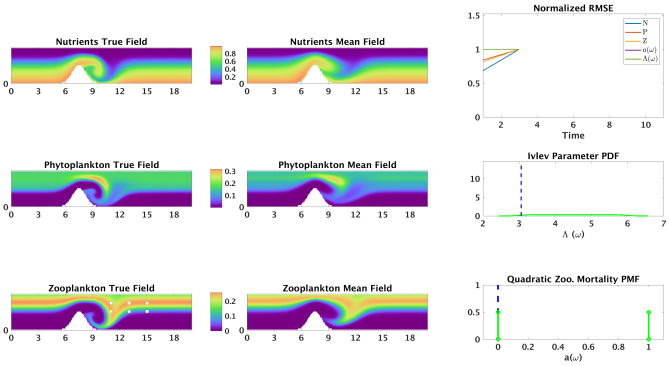
1) Experiment - 1: In the first main experiment, for the physical model, uncertainty is in the initial conditions for the state variables and the eddy viscosity parameter (Λ_{Re}). We consider the Boussinesq coupling between momentum and density equation to be absent, with the density as a passive tracer. In the lower-trophic-level biological model, uncertainty is introduced by the ambiguity in the presence or absence of quadratic zooplankton mortality functional ($a \in \{0, 1\}$), along with the Ivlev grazing parameter (Λ). In the fish model, the uncertainty comes from the presence of uncertain primary production and the recruitment time (T_r) in the source term of the forage equation (Eq. 4) and from uncertain physical variables in the effective advection velocities (Eq. 5). The goal is to learn all the state variables fields, along with the uncertain parameters and parameterizations, through a few observations of the Zooplankton field. These data are sparse in both space and time, with the observations only available at six locations every two non-dimensional times, starting at time $T = 3$ and ending at $T = 11$. The parameter values used in this experiment, adapted from the literature, are given in Table I.

Figure 2 shows the prior of the system at $T = 3$, i.e., just before the first set of observations are available. There are many differences between the mean and true fields of all the state variables. The blue dotted line in the probability plots of the parameters marks the true non-dimensional values. The prior probabilities of these parameters are considered to be uniform within a certain range. A phytoplankton bloom develops in top-right of the seamount due to the upwelling of nutrients from the bottom, which causes an increase in the forage concentration. In-turn, the fish population increases. A vortex also starts to develop in the wake of the seamount. We provide the corresponding standard deviation fields in Figure 3. There exists a large amount of uncertainty in the exact location and size of both the bloom and the vortex.

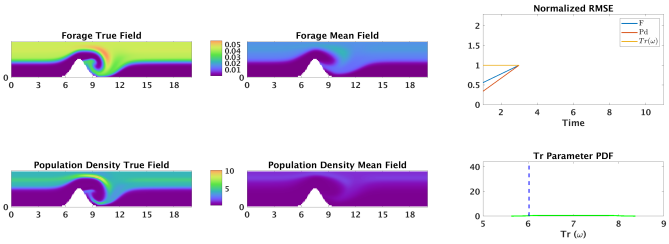
In Figure 4, we provide the posterior of the system after two observational episodes, i.e. at $T = 5$. By observing the zooplankton field, we are not only able to correct the biological model tracers and its parameters, but also the dynamics of the flow, as seen by the clustering of the Λ_{Re} distribution around its true value. Though we do not see a large correction in the



(a) Physics model



(b) LTL-Biological model



(c) Fish model

Fig. 2: The prior state of the stochastic dynamical system used in the experiment-1, at $T = 3$ (i.e. just before the 1st observational episode). (a), (b), (c): The first two columns consist of the true (left) and mean (right) field of the state variables of the corresponding models. In the third column, the first two plots show the variation of RMSE with time for various stochastic state variables and parameters. The remaining plots contain the probability distribution of the uncertain parameters $\Lambda_{Re}(\omega)$, $\Lambda(\omega)$, $a(\omega)$ (to learn the presence or absence of quadratic zoo. mortality), and recruitment time $T_r(\omega)$ parameters, respectively. The white circles on the zooplankton true field mark the observation locations.

fish model state variables, the probability distribution for the recruitment time (T_r) begins to approach the true value. We use the variation of Root Mean Square Error (RMSE) over

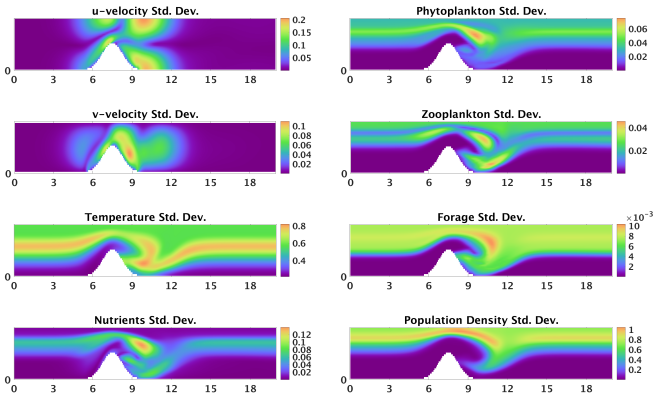


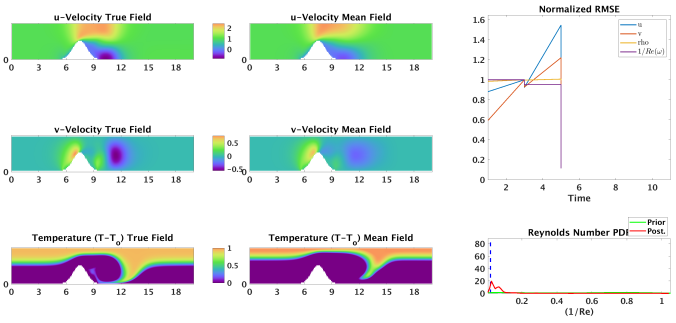
Fig. 3: The prior standard deviation of the stochastic dynamical system used in the experiment-1, at $T = 3$ (i.e. just before the 1st observational episode).

time to judge performance. RMSE is the L_2 distance between the mean of the random variables in the stochastic run and the simulated truth. The RMSE value for each of the variables at every time is normalized by the corresponding RMSE value just before the first assimilation step. Hence, our findings are corroborated by the decrease in RMSE for the parameters and state variables (except the temperature field), and the fact that assimilating the first observation at $T = 3$ was not effective.

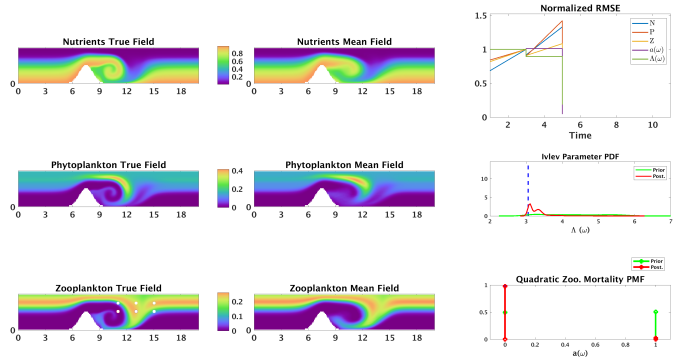
Finally, in Figure 5, we present the posterior after 5 observational episodes at $T = 11$. We unambiguously learn all the parameter values from the data, even detecting the absence of quadratic mortality term from our NPZ model. We observe agreement between the mean and true fields for the velocities, NPZ tracers, and the forage. It is interesting to note that we make no correction to the temperature field. This is perfectly as expected, because it does not affect the biological tracers: since in the present simulation, temperature is a passive tracer, the zooplankton data contains no information about the temperature field: it is thus not identifiable from the given data (also called the problem of identifiability [51]). The fish population density is directly affected by the temperature field, about which we have no information, but is indirectly related to zooplankton through the primary production and forage; hence, we are able to learn the fish population from zooplankton data through the somewhat weak link of primary production and forage.

2) Experiment-2: In this experiment, to show how the overall learnability of the fish model from zooplankton observations can be improved, we consider the physical model to be known, i.e. deterministic. However, we turn on the full temperature-momentum Boussinesq coupling, hence leading to more complex nonhydrostatic dynamics including internal waves.

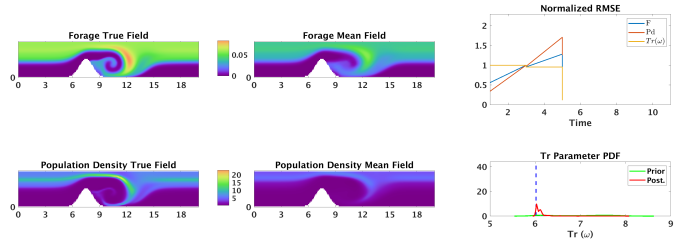
In Figure 6, we provide the posterior state of the system directly after 10 observational episodes at $T = 21$, and as expected, there is a better match between the GMM-DO mean fields and the true fields for the fish model tracers. The probability distribution for the T_r parameter has also



(a) Physics model



(b) LTL-Biological model



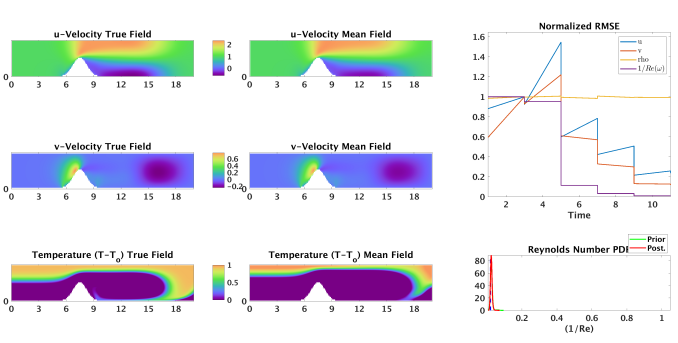
(c) Fish model

Fig. 4: Posterior state of the stochastic dynamical system used in the experiment-1, at $T = 5$ (i.e. just after the 2nd observational episode).

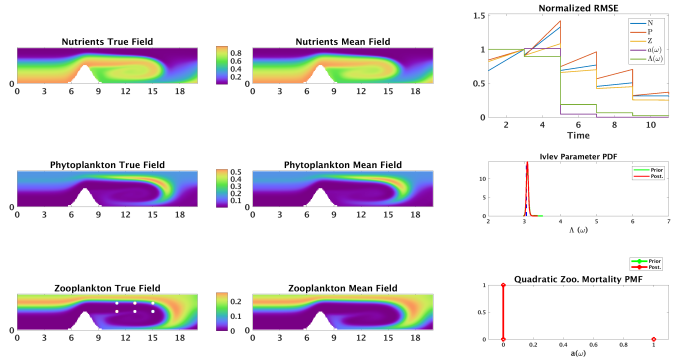
become concentrated around its true value. The effects of the known internal lee waves are clearly visible on all coupled physics-LTL-fish fields. As a result, the forage field is more challenging to learn than before. Even though the physics is known, due to the complicated nature of this flow dynamics, a larger number of observational episodes were indeed needed to achieve the learning objectives.

B. Realistic Simulations for the Lakshadweep Islands

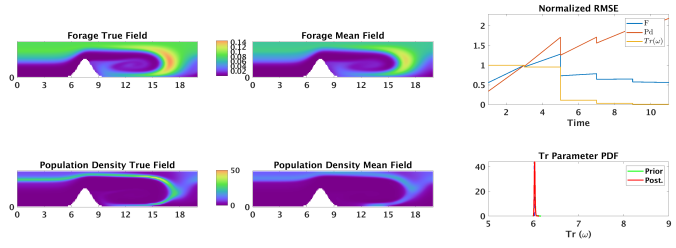
The main fishery in the Lakshadweep islands is the live-bait-pole and line tuna fishery, targeting Skipjack and Yellowfin. We obtained fish catch data for four islands, Agatti, Kadmat, Kavaratti, and Minicoy, in the region. The data was collected



(a) Physics model



(b) LTL-Biological model

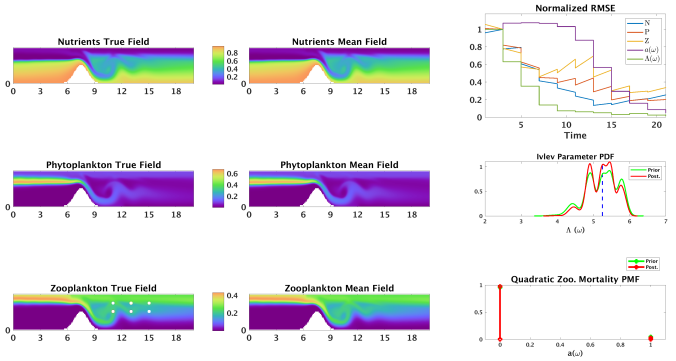


(c) Fish model

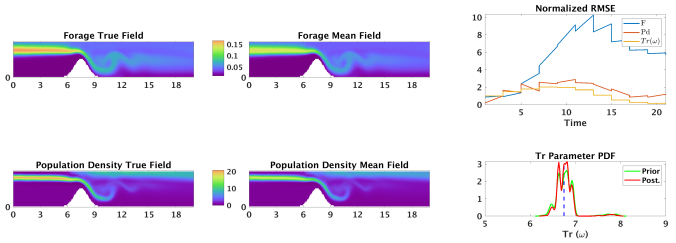
Fig. 5: The posterior state of the coupled physical-biological-fish model based stochastic dynamical system used in the experiment-1, at $T = 11$ (i.e. just after the 5th observational episode).

over a span of four years between January 2014 to 2018 under the community-based fisheries monitoring program of the Dakshin Foundation in India [18]. Fish catch data includes features such as date, total time and fuel, fish catch, etc. [30], that can be used to compute a normalized catch-per-unit-effort (CPUE) categorized into four categories, ranging from poor (1) to very good (4) fish availability [66]. As we will show, using our realistic ocean simulations for the region, we can qualitatively explain trends in fish catch data with physical ocean processes.

We performed a case study on March 2015, and related



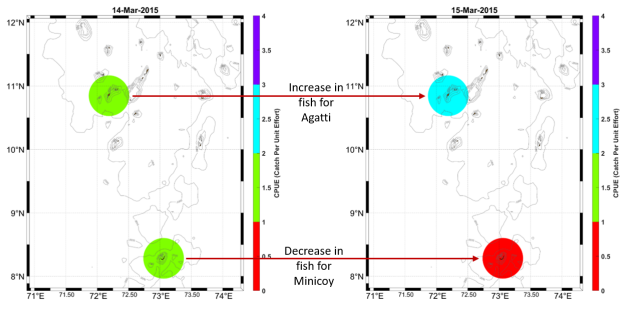
(a) LTL-Biological model



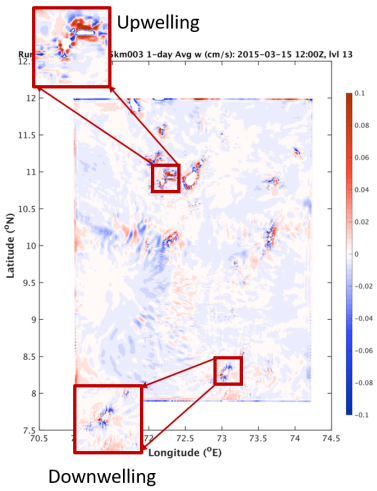
(b) Fish model

Fig. 6: The posterior state of the coupled physical-biological-fish model based stochastic dynamical system used in the experiment-2, at $T = 21$ (i.e. just after the 10th observational episode).

the changes in fish availability with physical processes. For example, we show in Figure 7a the change in fish availability for the islands of Agatti and Minicoy on March 15$^{\text{th}}$, 2015. There is an increase in fish availability for Agatti and a decrease for Minicoy. This change in CPUE may be related to our realistic MSEAS hindcast simulations. Indeed, we find upwelling and downwelling processes near these islands—there is a positive 1-day averaged vertical velocity at a depth of $52m$ around Agatti, while mostly negative vertical velocities for Minicoy. In Figure 8, we also compute a 2-day backward z flow map at $52\ m$ depth, as computed by the PDE-based method of flow map composition [33]: the result provides the depths at which the water masses originated 48-hours ago. Blue represents depths greater than $52\ m$ indicating Agatti experiences upwelling, while Minicoy is surrounded by mostly red, signifying water originating from lower than $52\ m$ depths. Upwelling occurs at Agatti because the impinging flow gets pushed above the surrounding bathymetry, while for Minicoy flows go around the bathymetry, generating negative vorticity and downwelling. As deeper water comes laden with nutrients, it provides a fertile zone, resulting in increase in fish availability at Agatti.



(a) Measured change in CPUE value



(b) 1-Day averaged simulated vertical velocity at $52m$

Fig. 7: (a) Change in CPUE value for the islands of Agatti and Minicoy on 15th March, 2015. (b) 1-day averaged vertical velocity at a depth of $52m$ on 15th March, 2015, as hindcast by the MSEAS PE model simulation.

V. Conclusions and Extensions

We provided a comprehensive overview of the current state-of-the-art in fish modeling. Taking into account different sources of uncertainty including initial conditions, parameters, and parameterizations, we numerically integrate a stochastic coupled nonhydrostatic ocean physical-biological-fish dynamical model in an idealized domain, using the Dynamically Orthogonal (DO) methodology, an adaptive reduced-dimension stochastic modeling technique. We then use a rigorous PDE-based Bayesian learning framework to perform a nonlinear inference of high-dimensional states containing multidisciplinary unknown states, parameters, and parameterizations. The learning results are promising for use with realistic simulations. We then investigate the capability of realistic physical simulations at predicting the advection of biogeochemical quantities relevant for fish populations. Specifically, we analyze the changes in tuna fish concentration from collected fishing data for the Lakshadweep islands in India, by connecting these changes to upwelling and downwelling

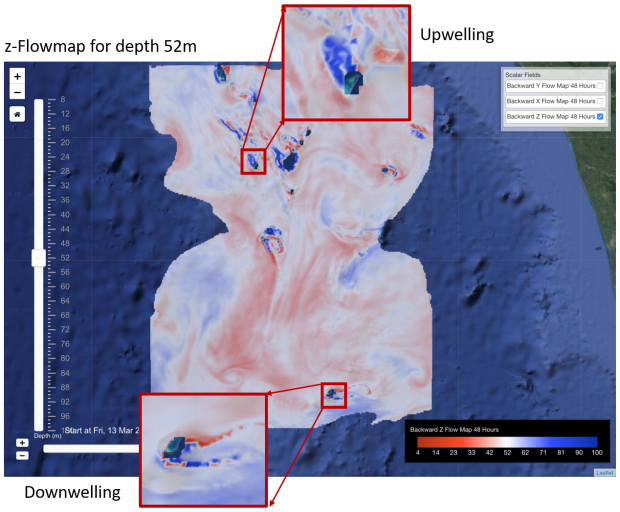


Fig. 8: 48-hour backward z flow map at a depth of $52m$ on 15th March, 2015, as hindcast by the MSEAS PE model simulation.

processes predicted to occur by the MSEAS primitive equation ocean model. The result indicate that the realistic MSEAS simulations can forecast the physics relevant to the regional Lakshadweep ecosystem.

Possible future extensions involve the derivation and implementation of the LTL-biological and fish stochastic model within the recently developed probabilistic Dynamically Orthogonal primitive equation (DO-PE) regional ocean modeling system [62]. The MSEAS DO-PE is combination of the MSEAS PE model with the DO methodology to perform uncertainty propagation in realistic ocean models. Finally, employing Bayesian learning in conjunction with real fish catch data using the PDE-based GMM-DO filter [46, 45, 23] constitutes the subject of ongoing research.

Acknowledgment

We thank the members of the MSEAS group for useful discussions. We are grateful to the MIT Tata Center for Design and Technology for providing fellowships to AG and DNS. We thank Drs. Robert Stoner and Angeliki Diane Rigos for their valuable feedback and support. We thank the Dakshin foundation for providing the Lakshadweep fish catch data. PFJL and PJH are grateful to the Office of Naval Research (ONR) for support under grants N00014-14-1-0725 (Bays-DA) and N00014-19-1-2693 (IN-BDA) to the Massachusetts Institute of Technology (MIT).

References

[1] Alethea Barbaro. Modelling and simulations of the migration of pelagic fish. *ICES Journal of Marine Science*, 66(5):826–838, 2009.

[2] J. W. Baretta. Preface to the European Regional Seas Ecosystem Model II. *Journal of Sea Research*, 38(3):169–171, 1997.

[3] J. W. Baretta, W. Ebenhöh, and P. Ruardij. The European Regional Seas Ecosystem Model, a complex marine ecosystem model. *Netherlands Journal of Sea Research*, 33(3):233–246, 1995.

[4] Mr. Bayes and Mr. 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)*, 53:370–418, 1763.

[5] Ş. T. Beşiktepe, P. F. J. Lermusiaux, and 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*, 40–41:171–212, 2003.

[6] J. C. Blackford, J. I. Allen, and F. J. Gilbert. Ecosystem dynamics at six contrasting sites: a generic modelling study. *Journal of Marine Systems*, 52(1):191–215, 2004.

[7] Villy Christensen, Marta Coll, Jeroen Steenbeek, Joe Buszowski, Dave Chagaris, and Carl J Walters. Representing variable habitat quality in a spatial food web model. *Ecosystems*, 17(8):1397–1412, 2014.

[8] James A Cummings and Ole Martin Smedstad. Variational data assimilation for the global ocean. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, pages 303–343. Springer, 2013.

[9] Cabell S. Davis and John H. Steele. Biological/physical modeling of upper ocean processes. Technical report, Woods Hole Oceanographic Institution, 1994.

[10] Department of Animal Husbandy, Dairing & Fisheries. Handbook on fisheries statistics, 2014.

[11] Maéva Doron, Pierre Brasseur, and Jean-Michel Brankart. Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical–biogeochemical model: Twin experiments. *Journal of Marine Systems*, 87(3):194–207, 2011.

[12] Gary D Egbert and Svetlana Y Erofeeva. Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology*, 19(2):183–204, 2002.

[13] Gary D Egbert and Svetlana Y Erofeeva. OSU tidal inversion. http://volkov.oce.orst.edu/tides/tpxo8_atlas.html, 2013.

[14] M. J. R. Fasham, H. W. Ducklow, and S. M. McKelvie. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, 48(3):591–639, 1990.

[15] Florian Feppon and Pierre F. J. Lermusiaux. Dynamically orthogonal numerical schemes for efficient stochastic advection and Lagrangian transport. *SIAM Review*, 60(3):595–625, 2018.

[16] Florian Feppon and Pierre F. J. Lermusiaux. The extrinsic geometry of dynamical systems tracking nonlinear matrix projections. *SIAM Journal on Matrix Analysis and Applications*, 40(2):814–844, 2019.

[17] Glenn Flierl and Dennis J. McGillicuddy. Mesoscale and submesoscale physical-biological interactions. *The sea*, 12:113–185, 2002.

[18] Dakshin Foundation. Community-based fisheries monitoring program. https://www.dakshin.org/, Started in January 2014.

[19] P. J. S. Franks, J. S. Wroblewski, and G. R. Flierl. Behavior of a simple plankton model with food-level acclimation by herbivores. *Marine Biology*, 91(1):121–129, 1986.

[20] Peter J. S. Franks. NPZ models of plankton dynamics: their construction, coupling to physics, and application. *Journal of Oceanography*, 58(2):379–387, 2002.

[21] Marjorie A. M. Friedrichs, Jeffrey A. Dusenberry, Laurence A. Anderson, Robert A. Armstrong, Fei Chai, James R. Christian, Scott C. Doney, John Dunne, Masahiko Fujii, Raleigh Hood, 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.

[22] Marjorie AM Friedrichs, Mary-Elena Carr, Richard T Barber, Michele Scardi, David Antoine, Robert A Armstrong, Ichio Asanuma, Michael J Behrenfeld, Erik T Buitenhuis, Fei Chai, et al. Assessing the uncertainties of model estimates of primary productivity in the tropical pacific ocean. *Journal of Marine Systems*, 76(1-2):113–133, 2009.

[23] Abhinav Gupta. Bayesian Inference of Obstacle Systems and Coupled Biogeochemical-Physical Models. Master's thesis, IIT Kanpur, Kanpur India, 2016.

[24] P. J. Haley, Jr., A. Agarwal, and P. F. J. Lermusiaux. Optimizing velocities and transports for complex coastal regions and archipelagos. *Ocean Modeling*, 89:1–28, 2015.

[25] Patrick J. Haley, Jr. and Pierre 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):1497–1537, December 2010.

[26] Sarah Hinckley. *Biophysical mechanisms underlying the recruitment process in walleye pollock (Theragra chalcogramma)*. PhD thesis, 1999.

[27] EE Hofmann and Marjorie AM Friedrichs. Predictive modeling for marine ecosystems. *The Sea*, 12:537–565, 2002.

[28] Timothy F Hogan, Ming Liu, James A Ridout, Melinda S Peng, Timothy R Whitcomb, Benjamin C Ruston, Carolyn A Reynolds, Stephen D Eckermann, Jon R Moskaitis, Nancy L Baker, et al. The navy global environmental model. *Oceanography*, 27(3):116–125, 2014.

[29] Jeremy BC Jackson, Michael X Kirby, Wolfgang H Berger, Karen A Bjorndal, Louis W Botsford, Bruce J Bourque, Roger H Bradbury, Richard Cooke, Jon Erlandson, James A Estes, et al. Historical overfishing and the recent collapse of coastal ecosystems. *science*, 293(5530):629–637, 2001.

[30] Mahima Jaini, Shwetha Nair, Ishaan Khot, Meera Oommen, Naveen Namboothri, and Kartik Shanker. Linking conservation and livelihoods in lakshadweeps fisheries: Long-term monitoring of the live-bait pole and line tuna fishery, 2016.

[31] Emlyn Jones, John Parslow, and Lawrence Murray. A bayesian approach to state and parameter estimation in a phytoplankton-zooplankton model. *Australian Meteorological and Oceanographic Journal*, 59(SP):7–16, 2010.

[32] Michio J Kishi, Shin-ichi Ito, Bernard A Megrey, Kenneth A Rose, and Francisco E Werner. A review of the nemuro and nemuro. fish models and their application to marine ecosystem investigations. *Journal of oceanography*, 67(1):3–16, 2011.

[33] Chinmay S. Kulkarni and Pierre F. J. Lermusiaux. Advection without compounding errors through flow map composition. *Journal of Computational Physics*, 398:108859, December 2019.

[34] Rajeev Kumar, Szymon Surma, Tony J Pitcher, Divya Varkey, Mimi E Lam, Cameron H Ainsworth, and Evgeny Pakhomov. An ecosystem model of the ocean around haida gwaii, northern british columbia: Ecopath, ecosim and ecospace. 2016.

[35] Kuo-Wei Lan, Teruhisa Shimada, Ming-An Lee, Nan-Jay Su, and Yi Chang. Using remote-sensing environmental and fishery data to map potential yellowfin tuna habitats in the tropical pacific ocean. *Remote Sensing*, 9(5):444, 2017.

[36] P. Lehodey, I. Senina, and R. Murtugudde. A spatial ecosystem and populations dynamics model (SEAPODYM)–modeling of tuna and tuna-like populations. *Progress in Oceanography*, 78(4):304–318, 2008.

[37] Patrick Lehodey, JEAN-MICHEL ANDRE, Michel Bertignac, John Hampton, Anne Stoens, Christophe Menkès, Laurent Mémery, and Nicolas Grima. Predicting skipjack tuna forage distributions in the equatorial pacific using a coupled dynamical bio-geochemical model. *Fisheries Oceanography*, 7(3-4):317–325, 1998.

[38] Patrick Lehodey, Fei Chai, and John Hampton. Modelling climate-related variability of tuna populations from a coupled ocean–biogeochemical-populations dynamics model. *Fisheries Oceanography*, 12(4-5):483–494, 2003.

[39] P. F. J. Lermusiaux. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *Journal of Computational Physics*, 217(1):176–199, 2006.

[40] P. F. J. Lermusiaux, P. J. Haley, W. G. Leslie, A. Agarwal, O. Logutov, and L. J. Burton. Multiscale physical and biological dynamics in the Philippine Archipelago: Predictions and processes. *Oceanography*, 24(1):70–89, 2011. Special Issue on the Philippine Straits Dynamics Experiment.

[41] Pierre F. J. Lermusiaux, P. Malanotte-Rizzoli, D. Stammer, J. Carton, J. Cummings, and A. M. Moore. Progress and prospects of U.S. data assimilation in ocean research. *Oceanography*, 19(1):172–183, 2006.

[42] W. G. Leslie, P. J. Haley, Jr., P. F. J. Lermusiaux, M. P. Ueckermann, O. Logutov, and J. Xu. MSEAS Manual. MSEAS Report 06, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 2010.

[43] O. G. Logutov and P. F. J. Lermusiaux. Inverse barotropic tidal estimation for regional ocean applications. *Ocean Modelling*, 25(1–2):17–34, 2008.

[44] Svetlana N. Losa, Gennady A. Kivman, and Vladimir 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):1–20, 2004.

[45] P. G. Y. Lu and Pierre 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.

[46] Peter Guang Yi Lu. Bayesian inference of stochastic dynamical models. Master's thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts, February 2013.

[47] Kjartan G Magnusson. An overview of the multispecies vpa theory and applications. *Reviews in Fish Biology and Fisheries*, 5(2):195–212, 1995.

[48] Nicholas C Makris, Purnima Ratilal, Deanelle T Symonds, Srinivasan Jagannathan, Sunwoong Lee, and Redwood W Nero. Fish population and behavior revealed by instantaneous continental shelf-scale imaging. *Science*, 311(5761):660–663, 2006.

[49] Olivier Maury, Blaise Faugeras, Yunne-Jai Shin, Jean-Christophe Poggiale, Tamara Ben Ari, and Francis Marsac. Modeling environmental effects on the size-structured energy flow through marine ecosystems. part 1: the model. *Progress in Oceanography*, 74(4):479–499, 2007.

[50] Bernard A Megrey, Kenneth A Rose, Robert A Klumb, Douglas E Hay, Francisco E Werner, David L Eslinger, and S Lan Smith. A bioenergetics-based population dynamics model of pacific herring (clupea harengus pallasi) coupled to a lower trophic level nutrient–phytoplankton–zooplankton model: description, calibration, and sensitivity analysis. *Ecological Modelling*, 202(1):144–164, 2007.

[51] Ralph F Milliff, Jerome Fiechter, William B Leeds, Radu Herbei, Christopher K Wikle, Mevin B Hooten, Andrew M Moore, Thomas M Powell, and Jeremiah Brown. Uncertainty management in coupled physical-biological lower trophic level ocean ecosystem models. *Oceanography*, 26(4):98–115, 2013.

[52] Priscilla A. Newberger, John S. Allen, and Y. H. Spitz. Analysis and comparison of three ecosystem models. *Journal of Geophysical Research: Oceans (1978–2012)*, 108(C3), 2003.

[53] A. R. Robinson, B. J. Rothschild, W. G. Leslie, J. J. Bisagni, M. F. Borges, W. S. Brown, D. Cai, P. Fortier, A. Gangopadhyay, P. J. Haley, Jr., H. S. Kim, L. Lanerolle, P. F. J. Lermusiaux, C. J. Lozano, M. G. Miller, G. Strout, and M. A. Sundermeyer. The development and demonstration of an advanced fisheries management information system. In *Proc. of the 17th Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, pages 186–190, Albuquerque, New Mexico, 2002. American Meteorological Society.

[54] Allan R. Robinson and Pierre F. J. Lermusiaux. Data assimilation for modeling and predicting coupled physical–biological interactions in the sea. In Allan R. Robinson, James J. McCarthy, and Brian J. Rothschild, editors, *Biological-Physical Interactions in the Sea*, volume 12 of *The Sea*, chapter 12, pages 475–536. John Wiley and Sons, New York, 2002.

[55] Themistoklis P. Sapsis and Pierre F. J. Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D: Nonlinear Phenomena*, 238(23–24):2347–2360, December 2009.

[56] Themistoklis P. Sapsis and Pierre F. J. Lermusiaux. Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty. *Physica D: Nonlinear Phenomena*, 241(1):60–76, 2012.

[57] Simo Särkkä. *Bayesian filtering and smoothing*, volume 3. Cambridge University Press, 2013.

[58] Walter HF Smith and David T Sandwell. Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, 277(5334):1956–1962, 1997.

[59] T. Sondergaard and 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):1737–1760, 2013.

[60] T. Sondergaard and P. F. J. Lermusiaux. Data assimilation with Gaussian Mixture Models using the Dynamically Orthogonal field equations. Part II: Applications. *Monthly Weather Review*, 141(6):1761–1785, 2013.

[61] Craig A Stow, Jason Jolliff, Dennis J McGillicuddy Jr, Scott C Doney, J Icarus Allen, Marjorie AM Friedrichs, Kenneth A Rose, and Philip Wallhead. Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, 76(1-2):4–15, 2009.

[62] Deepak Narayanan Subramani. *Probabilistic Regional Ocean Predictions: Stochastic Fields and Optimal Planning*. PhD thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, Massachusetts, February 2018.

[63] Iragavarapu Suryanarayana, Antonio Braibanti, Rupenaguntla Sambasiva Rao, Veluri Anantha Ramam, Duvvuri Sudarsan, and Gollapalli Nageswara Rao. Neural networks in fisheries research. *Fisheries Research*, 92(2-3):115–139, 2008.

[64] M. P. Ueckermann and 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.

[65] Jintao Wang, Xinjun Chen, Kevin W Staples, and Yong Chen. The skipjack tuna fishery in the west-central pacific ocean: applying neural networks to detect habitat preferences. *Fisheries Science*, 84(2):309–321, 2018.

[66] Jintao Wang, Wei Yu, Xinjun Chen, Lin Lei, and Yong Chen. Detection of potential fishing zones for neon flying squid based on remote-sensing data in the northwest pacific ocean using an artificial neural network. *International journal of remote sensing*, 36(13):3317–3330, 2015.

[67] Ben A. Ward, Marjorie A. M. Friedrichs, Thomas R. Anderson, and Andreas Oschlies. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems*, 81(1):34–43, 2010.