Gupta, A. and P.F.J. Lermusiaux, 2022. Bayesian Learning of Coupled Biogeochemical-Physical Models. Progress in Oceanography, sub-judice.

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 the sparse measurements and limited understanding of the myriad of ocean processes, there is however significant uncertainty. There is model uncertainty in the parameter values, functional forms with diverse parameterizations, and level of complexity needed, and thus in the state variable 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 variable 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. It seamlessly and rigorously discriminates among existing models, but also 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 the learning skill, and evaluate convergence and the sensitivity to hyper-parameters. Our PDE framework successfully discriminates among functional forms and model complexities, and learns in the absence of prior knowledge by searching in dense function spaces. The probabilities of biogeochemical-physical fields and parameters, and of known, uncertain, and unknown model formulations are updated jointly using Bayes' law. Non-Gaussian statistics, ambiguity, and biases are captured. The parameter values and the model formulations that best explain the data are identified. When observations are sufficiently informative, model complexity and model functions are discovered.

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

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