



A Bayesian parameter estimation method applied to a marine ecosystem model for the coastal Gulf of Alaska



J. Fiechter^{a,*}, R. Herbei^b, W. Leeds^c, J. Brown^d, R. Milliff^e, C. Wikle^f, A. Moore^g, T. Powell^h

^a Institute of Marine Sciences, University of California, Santa Cruz, USA

^b Department of Statistics, Ohio State University, Columbus, USA

^c Department of Statistics, University of Chicago, Chicago, USA

^d Principal Scientific Group, Fishers, USA

^e Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, USA

^f Department of Statistics, University of Missouri, Columbia, USA

^g Ocean Sciences Department, University of California, Santa Cruz, USA

^h Integrative Biology Department, University of California, Berkeley, USA

ARTICLE INFO

Article history:

Received 4 December 2012

Received in revised form 1 March 2013

Accepted 7 March 2013

Available online 12 April 2013

Keywords:

Marine ecosystem model

Bayesian sampling

Adaptive Metropolis–Hasting

Parameter estimation

Posterior distributions

Gulf of Alaska

ABSTRACT

The present study describes a state-of-the-art methodology based on an adaptive Metropolis–Hastings algorithm to facilitate efficient Bayesian sampling for realistic lower trophic level (LTL) marine ecosystem models. The main objective is to explore the ability to differentiate between biological parameters that can learn from observations and those that cannot. The Bayesian approach is applied to the northwestern coastal Gulf of Alaska region and uses both synthetic and actual (*in situ* and remotely sensed) observations. LTL ecosystem dynamics in the Bayesian framework are described by a process model consisting of a 1-dimensional Nutrient–Phytoplankton–Zooplankton–Detritus formulation with iron limitation (NPZDFe) and vertical mixing. The results illustrate the ability to determine parameter posterior distributions for fundamental biological rates, such as maximum phytoplankton growth or zooplankton grazing. By using various observational platforms as data stage inputs, the results also demonstrate the impact of spatial and temporal sampling on parameter posterior distributions, as well as the benefits of having concurrent measurements for two or more state variables of the process model (e.g., chlorophyll and nitrate concentrations). Extending the method to multiple parameters is non-trivial, as posterior distributions become impacted by correlated and/or disproportionate contributions for certain model parameters. Controlled experiments with “near perfect data” were useful to characterize parameter identifiability based on information content in the BHM data stage inputs, as well as to separate uncertainties due to sampling issues vs. uncertain ecosystem process interpretation.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Lower trophic level (LTL) marine ecosystem models have become increasingly important for understanding how marine ecological systems respond to and affect climate forcing. In particular, when coupled to physical models, LTL marine ecosystem models are one of the main tools for evaluating primary productivity and its impact on the global carbon cycle (e.g., Friedrichs et al., 2009). Since available *in situ* observations are often sparse and remotely sensed observations are generally limited to surface or upper-ocean vertically averaged values for a single variable (i.e., chlorophyll), models are also important for exploring the dynamical linkages between multiple components of the marine food web. LTL marine ecosystem models are most often represented

as a system of multi-component nonlinear differential equations with various physical and biological parameterizations (e.g., Miller, 2004). It is well-known that there is dependence among model parameters and that there is typically not enough data to adequately calibrate these parameters (i.e., the “under-determination problem” in marine ecosystem modeling; e.g., Ward et al., 2010, and references therein). In addition, previous studies have determined that a particular parameter set that optimizes a model in one location may not be the same for another location (e.g., Friedrichs et al., 2007, and references therein).

LTL marine ecosystem models have traditionally been considered from a deterministic perspective and parameter estimation (or calibration) has been conducted with standard optimization algorithms (e.g., Ward et al., 2010). However, in the case of an underdetermined system, such methods require that poorly constrained parameters be fixed and/or limited to realistic ranges to give realistic results relative to observations. As described by Ward et al. (2010), arbitrarily fixing parameters may lead to too little

* Corresponding author. Tel.: +1 831 459 1306; fax: +1 831 459 4882.

E-mail addresses: fiechter@ucsc.edu, jfiechter@yahoo.com (J. Fiechter).

model variability, while prescribing parameter ranges may yield too much model variability. Ward et al. (2010) recommend the adoption of a more sophisticated use of prior information and suggest that a Bayesian approach which assimilates data in the context of informative prior knowledge might be favored. Recent work by Weir et al. (2012) also provides useful insight on ocean ecosystem models and parameter estimation in the contexts of Maximum Likelihoods and Bayes Theorem. More specifically, they investigate state and parameter estimation for a predator–prey system (i.e. the Lotka–Volterra equations) employing solution procedures adapted from deterministic data assimilation methodologies (e.g., cost function minimization, particle filters) and analyses based in non-linear systems theory (e.g., bifurcations, limit cycles, and fixed points). Their approach offers interesting connections and potential synergies with the Markov Chain Monte Carlo (MCMC) parameter estimation methods implemented in the present study.

1.1. Benefits of Bayesian approach

In the context of LTL marine ecosystem models, the Bayesian framework has been considered for both state estimation (e.g., Harmon and Challenor, 1997; Evensen, 2003; Dowd, 2006, 2007; Jones et al., 2010) and parameter estimation (Harmon and Challenor, 1997; Malve et al., 2007; Jones et al., 2010; Dowd, 2011) for marine biogeochemical models. For the state estimation problem, the preferred approach has been based on derivative of sequential importance sampling (e.g., Doucet et al., 2001), primarily ensemble Kalman filtering (e.g., Evensen, 2003) and particle filtering (Dowd, 2006), although MCMC methods (e.g., Robert and Casella, 2004) have been considered as well (Dowd, 2007). For parameter estimation, Malve et al. (2007) have implemented a modern MCMC algorithm that relies on an adaptive Metropolis–Hastings (M–H) algorithm. The strength of this methodology is that it accounts for dependence in the parameters and adapts the associated proposal distribution to increase sampling efficiency. The simultaneous inference on both parameters and states has been approached from both the hybrid particle filtering/M–H perspective (Jones et al., 2010; Dowd, 2011) and the MCMC sampling perspective (e.g., Harmon and Challenor, 1997; Jones et al., 2010). However, these particular implementations considered quite low-dimensional parameter and state-spaces.

Jones et al. (2010) suggest that the MCMC–Gibbs approach is more efficient than the particle filter/M–H approach, but it may be more limited in practical problems due to the difficulty of obtaining efficient proposals for high-dimensional state processes. In addition, the MCMC approach of Jones et al. (2010) used a quasi-linear approximation of biogeochemical model dynamics to facilitate Gibbs updates. Such an approximation may not be reasonable in models formulations more complex than the two-component Lotka–Volterra predator–prey model that motivated their example. Given the importance of LTL marine ecosystem models, there is a need for a Bayesian inference methodology that can accommodate high dimensional, underdetermined parameter spaces, as well as high-dimensional state-spaces and non-Gaussian data and process models.

The present study describes a state-of-the-art methodology based on adaptive M–H methodology to facilitate efficient Bayesian sampling for realistic LTL marine ecosystem models, with one of the main objectives being to explore the ability to differentiate between parameters that can learn from the data and those that cannot. More specifically, the Bayesian hierarchical model (BHM) is applied to the northwestern coastal Gulf of Alaska (CGOA) region (Supplementary Fig. S1) and uses both actual observations (*in situ* and remotely sensed) and “pseudo-observations” generated from a 3-dimensional (3-D) coupled physical–biological model. LTL ecosystem dynamics in the Bayesian framework are

described by a process model consisting of a 1-dimensional (1-D) NPZD (Nutrient–Phytoplankton–Zooplankton–Detritus) formulation with iron limitation (NPZDFe; Fiechter et al., 2009) and vertical mixing.

See Supplementary Fig. S1 as supplementary file. Supplementary material related to this article found, in the online version, at <http://dx.doi.org/10.1016/j.ecolmodel.2013.03.003>.

By using NPZDFe explicitly as the process model in the Bayesian framework, the results presented here complement those from previous studies which have attempted to characterize parameter influence and posterior distribution based on: (1) ensemble calculations with the 3-D deterministic physical–biological model (Fiechter, 2012), and (2) first-order emulator models constructed on the basis of the deterministic ensemble simulations (Hooten et al., 2011; Leeds et al., 2012b). Since the NPZDFe formulation is directly embedded in the Bayesian hierarchy, the associated computational requirements are reduced as the solution procedure no longer requires multiple runs of the 3-D coupled physical–biological model to produce meaningful ensemble statistics.

Straight-forward implementation of the parameter estimation BHM for multiple parameters of the NPZDFe model was not readily tractable, as Bayesian estimation was not converging after lengthy and expensive iterative solution procedures. The use of deterministic ensemble calculations with the NPZDFe model coupled to a regional ocean circulation model (Fiechter, 2012) helped re-scope parameter estimation experiments to focus on identifying two key biological parameters (i.e., phytoplankton growth rate and zooplankton grazing rate). Furthermore, “near perfect data” experiments with subsets of the deterministic model output were useful to characterize parameter identifiability based on information content in the datasets used as BHM data stage inputs. These controlled experiments also helped separate uncertainties due to sampling issues vs. uncertainties associated with ecosystem process interpretation.

1.2. CGOA physical–biological environment

The CGOA exhibits a highly productive shelf supported by spring bloom dynamics and a high nutrient-low chlorophyll (HNLC) region offshore where, despite elevated nitrate and silicate concentrations, primary production remains low because of severe iron limitation on phytoplankton growth (Martin and Fitzwater, 1988). Physical mechanisms contributing to cross-shelf exchange, such as seasonal anticyclonic eddies, are critically important for alleviating phytoplankton growth limitation in the HNLC region via offshore transport of iron-rich shelf waters (Brown and Fiechter, 2012; Crawford et al., 2007; Fiechter and Moore, 2012; Ladd et al., 2005). Consequently, primary and secondary productivity in the CGOA exhibit significant spatial and temporal variability in response to physical and biological cross-shelf gradients (Strom et al., 2006, 2007). For example, intrinsic phytoplankton growth rates along the Seward Line (a cross-shelf transect off Seward, Alaska, routinely sampled during the GLOBEC program) were observed to vary by an order of magnitude (from ca. 0.1 to 1.0 day^{−1}) within a given month and across different shelf regions (Strom et al., 2006). Because of its intrinsic physical and biological variability, the CGOA is a well-suited test bed to estimate parameter distribution and uncertainty with respect to existing and future LTL ecosystem models for that region.

2. BHM methodology and implementation

Let X be the random variable denoting the process of interest, let Y be observations and let θ_p and θ_d be parameters associated

Download English Version:

<https://daneshyari.com/en/article/4376110>

Download Persian Version:

<https://daneshyari.com/article/4376110>

[Daneshyari.com](https://daneshyari.com)