INSTITUT FÜR INFORMATIK

Optimization of Parameters and Initial Values in a Marine NPZD-Type Ecosystem Model

Johannes Rückelt, Andreas Oschlies
Thomas Slawig

Bericht Nr. 1013
November 2010

CHRISTIAN-ALBRECHTS-UNIVERSITÄT

KIEL

Institut für Informatik der
Christian-Albrechts-Universität zu Kiel
Olshausenstr. 40
D – 24098 Kiel

Optimization of Parameters and Initial Values in a Marine NPZD-Type Ecosystem Model

Johannes Rückelt, Andreas Oschlies
Thomas Slawig

Bericht Nr. 1013
November 2010

e-mail: joru@informatik.uni-kiel.de, ts@informatik.uni-kiel.de

Optimization of Parameters and Initial Values in a Marine NPZD-Type Ecosystem Model

J. Rückelt[*], A. Oschlies[†], T. Slawig[*]

Abstract

Parameters and initial values of a one-dimensional marine ecosystem model are optimized using a gradient-based optimization algorithm taking into account parameter bounds. Sensitivities of the optimized parameters w.r.t. errors in observations and initial values are studied numerically and found to yield parameter ranges narrow relative to the a priori parameter uncertainty reflected in upper and lower bounds on the permitted parameter range. This means, that optimal parameters can be determined accurately. We find, that optimizing for the initial values along with the parameters can greatly improve the model's fit to the observations.

1 Introduction

A spatially one-dimensional marine biogeochemical model that simulates the interaction of dissolved inorganic nitrogen N, phytoplankton P, zooplankton Z and detritus D was developed by Oschlies and Garcon [9], with the aim of simulating the nitrogen and carbon cycles in the North Atlantic [7], [11]. Oschlies and Schartau [12] showed that local calibration of the ecosystem model also resulted in an improved performance when this model was embedded into the basin-scale circulation model. In the one-dimensional configuration, the model simulates one water column at a given horizontal position, which is motivated by the fact that there have been special observational time series studies at fixed locations, one of which was used here. Hourly profiles of turbulent diffusivities and temperatures are taken from a global three-dimensional circulation model assumed to provide a perfect representation of the real ocean state. Such a local off-line approach has several restrictions, e.g. it neglects horizontal transport of biogeochemical tracers and possible feedbacks of the ocean biology on the absorption of solar radiation and thus on the vertical heating profile.

The model was already used several times for the optimization of model parameters: In [17], a so-called micro-genetic algorithm (μGA) was used, and the cost function combined three observational data sets from different locations. Also, noise was added to the data. The authors observed the well-known behavior of stochastic optimization methods such as genetic algorithms (GAs) to require a huge number of function evaluations to get to terminate. A number of

[1]Institut für Informatik, Christian-Albrechts-Universitt Kiel, Germany. Email: joru,ts@informatik.uni-kiel.de, research supported by DFG Cluster of Excellence *The Future Ocean*.

[2]IFM-GEOMAR, Kiel, Germany. Email: aoschlies@ifm-geomar.de

the parameters turned out to be hard to identify or estimate by the used cost function.

In [18] (compare also [17],[19]), the authors used the same model (and another more complex one) to optimize the parameters again. In addition to a GA they also employed a variational technique using a gradient-based method for optimization. Since no parameter bounds were applied for the latter method, some parameters went out of bounds, and the number of optimized parameters had to be restricted for this method.

In [16], the same NPZD model was studied for only the single dataset of the Bermuda Atlantic Time-series Station (BATS). Here a variant of a GA as well as a different gradient-based optimization method were applied. The latter was now chosen to take into account the parameter bounds, and turned out to be superior to other comparable implementations since it includes a special line search procedure, compare [13]. It could be shown that it was on the one hand superior to the GA with respect to computing time and that, one the other hand, optimal parameters could be identified also with incorporated data uncertainties.

Based on these results and methods, in this paper the model is studied concerning its dependency on the model parameters *and* additionally with respect to the initial values, which had been kept fixed for the investigations mentioned above. Since initial values are difficult to find for biogeochemical models, it is important to know how their choice effects the model output and the parameters obtained by any identification or estimation method. The applied method of optimization is general and can be used for any initial or initial-boundary value problem at hand, since its components are readily available and no special software was designed to achieve the presented results.

The structure of the paper is as follows: We start by briefly describing the model structure and the parameters that are optimized or subject to estimation in the second section. In section 3, we present the parameter optimization problem and describe the relevant parts of the used optimization algorithm. In sections 4, we summarize the results for parameter estimation with synthetic data. Afterward, we make a short note about opportunities of spatial model reduction and how they affect the optimization in section 5. Section 6 presents the uncertainty analysis w.r.t. observational errors. In the main part in section 7, we show results of optimization and uncertainty analysis w.r.t. initial values and parameters simultaneously. We end the paper with conclusions in section 8.

2 Mathematical Model Equations

Biogeochemical models are coupled PDE systems consisting of time-dependent advection-diffusion-reaction equations with nonlinear coupling terms. The turbulent diffusivity, temperature and sometimes also salinity fields are either computed simultaneously or in advance by a physical ocean model. Clearly, the second variant (where the physical ocean model output is used as a kind of forcing for the ecosystem model) that is used in this paper is computationally cheaper but neglects the biology's feedback effects via impacts on the absorption of solar radiation, generally assumed to be small relative to uncertainties in the boundary conditions such as surface heat fluxes, see [8].

symbol	equation	meaning
V_P	$= \mu_m \cdot (C_{ref})^{cT}$	maximum growth rate of phytoplankton
u	$= \dfrac{N}{k_N + N}$	factor for nutrient limited growth rate of phytoplankton
$J(\mu, u)$	$= \min(\bar{\mu}(z), V_p u)$	growth rate of phytoplankton after Liebig's Law of the Minimum
$\bar{\mu}(z)$		light limited growth rate of phytoplankton, according to Evans and Parslow [1]
$G(\epsilon, g)$	$= \dfrac{g\epsilon P^2}{g + \epsilon P^2}$	zooplankton grazing function

Table 1: Auxiliary variables in the $NPZD$ model.

In the model, the concentrations (in mmol N m^{-3}) of dissolved inorganic nitrogen, phytoplankton, zooplankton, and detritus, denoted by $y = (y_j)_{j=n,p,z,d} = (N, P, Z, D)$ are described by the following PDE system:

$$\frac{\partial y_j}{\partial t} = -w_s \frac{\partial y_j}{\partial z} + \frac{\partial}{\partial z}\left(K_\rho \frac{\partial y_j}{\partial z}\right) + q_j(y), \quad j = n, p, z, d. \tag{1}$$

Here z denotes the vertical spatial coordinate, i.e. the depth in the water column. The output taken from the physical ocean model are hourly profiles of the turbulent mixing coefficients K_ρ and temperature, the latter needed in the biological process parameterizations below. The vertical sinking velocity w_s is a parameter of the biological model that is nonzero only for D.

The biogeochemical coupling (or source-minus-sink) terms for the four species are given by (see [9]):

$$\begin{array}{lrcl} \text{for } N: & q_n(y) & = & -J(\mu,u)P + \gamma_m D + \Phi_m^z Z, \\ \text{for } P: & q_p(y) & = & J(\mu,u)P - \Phi_m^p P - G(\epsilon,g)Z, \\ \text{for } Z: & q_z(y) & = & \beta G(\epsilon,g)Z - \Phi_m^z Z - \Phi_z^* Z^2, \\ \text{for } D: & q_d(y) & = & (1-\beta)G(\epsilon,g)Z + \Phi_z^* Z^2 + \Phi_m^p P - \gamma_m D. \end{array} \tag{2}$$

Tables 1 and 2 give a further description of the parameters and functions.

The source minus sink equations of the $NPZD$ model are affected by the light-limited growth rate $\mu(z,t)$ of phytoplankton, which varies with depth z and time t. Average light-limited phytoplankton growth rates $\overline{\mu(k,t)}$ are calculated for each depth layer k using a simplified version of an approximative formula by Evans and Parslow [1]. This is where the parameters α, κ enter the equations. For more details see [16].

2.1 Measurement Data and corresponding Model Output

Observational data is taken from the *Bermuda Atlantic Time-series Study* (called BATS, located at 31N 64W). The used data and their corresponding model variables are

- dissolved inorganic nitrogen (DIN) (in mmol N m^{-3}), corresponding to state variable N in the model,