Contents lists available at ScienceDirect

Nonlinear Analysis: Real World Applications

journal homepage: www.elsevier.com/locate/nonrwa

Parameter optimization and uncertainty analysis in a model of oceanic CO_2 uptake using a hybrid algorithm and algorithmic differentiation

J. Rückelt[a,1], V. Sauerland[a,1], T. Slawig[a,*,1], A. Srivastav[a,1], B. Ward[b], C. Patvardhan[c]

[a] Institut für Informatik, Christian-Albrechts-Universität Kiel, Germany
[b] National Oceanographic Centre, Southampton, UK
[c] Deemed University, Faculty of Engineering, Dayalbagh, Agra, India

ARTICLE INFO

Article history:
Received 22 February 2010
Accepted 10 March 2010

Keywords:
Parameter optimization
Biogeochemical modeling
Gradient-based optimization
Genetic algorithm
Algorithmic differentiation

ABSTRACT

Methods and results for parameter optimization and uncertainty analysis for a one-dimensional marine biogeochemical model of NPZD type are presented. The model, developed by Schartau and Oschlies, simulates the distribution of nitrogen, phytoplankton, zooplankton and detritus in a water column and is driven by ocean circulation data. Our aim is to identify parameters and fit the model output to given observational data. For this model, it has been shown that a satisfactory fit could not be obtained, and that parameters with comparable fits can vary significantly. Since these results were obtained by evolutionary algorithms (EA), we used a wider range of optimization methods: A special type of EA (called quantum-EA) with coordinate line search and a quasi-Newton SQP method, where exact gradients were generated by Automatic/Algorithmic Differentiation. Both methods are parallelized and can be viewed as instances of a hybrid, mixed evolutionary and deterministic optimization algorithm that we present in detail. This algorithm provides a flexible and robust tool for parameter identification and model validation. We show how the obtained parameters depend on data sparsity and given data error. We present an uncertainty analysis of the optimized parameters w.r.t. Gaussian perturbed data. We show that the model is well suited for parameter identification if the data are attainable. On the other hand, the result that it cannot be fitted to the real observational data without extension or modification, is confirmed.

1. Introduction

In the current discussion on climate change, carbon dioxide (CO_2) as one of the major greenhouse gases plays a central role. On the one hand, the greenhouse effect in the Earth's atmosphere is the reason that the temperature on the surface is in the range that we are used to and need to survive. On the other hand, the increasing concentration of greenhouse gases (such as carbon dioxide, methane and others) also causes the atmosphere's temperature to increase more and more, with a number of mostly undesired consequences.

One part of the global carbon cycle is the uptake of CO_2 by the ocean. Via the sea surface, CO_2 molecules are taken up by the ocean water from the atmosphere. It is estimated that about 30%–50% of the CO_2 that is currently emitted is taken up by the ocean. To some extend, this carbon dioxide is transformed into organic material in phytoplankton (algae) via photosynthesis. The phytoplankton again serves as a nutrient for zooplankton, and the dead material (detritus) of both is

* Corresponding author. Tel.: +49 431 8804490; fax: +49 431 8807618.
E-mail address: ts@informatik.uni-kiel.de (T. Slawig).

1 DFG Cluster of Excellence *The Future Ocean*.

1468-1218/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.nonrwa.2010.03.006

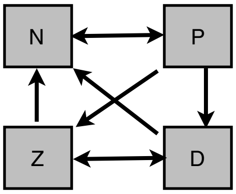

Fig. 1. Schematic description of the coupling between the variables in the NPZD model.

sinking down, but also may serve again as nutrient. Thus there is a complex coupled cycle of carbon, depending on the amount of phyto-, zooplankton, and detritus in the water.

The amount of CO_2 that stays in the ocean for longer time scales is important, since then this effect mitigates the rising CO_2 concentration in the atmosphere. The challenge is to understand, simulate and predict the underlying processes and their dependency on the climate change itself (e.g. rising sea water temperature).

As noted above, the most important process is photosynthesis which combines inorganic carbon (C), nitrogen (N), and phosphorus (P) with water to form organic matter and release oxygen. Photosynthesis is performed in phytoplankton. The *Redfield ratio* [1] states that the relation in marine phytoplankton is (see also [2, Section 4.2])

$$\text{C:N:P} = 106{:}16{:}1. \tag{1}$$

It it thus possible to calculate the amount of carbon contained in the phytoplankton by knowing for example the amount of nitrogen N. Thus N can be used as a model variable from which the potential uptake of CO_2 can be estimated, assuming that there is no limit on phosphorus P and carbon C (i.e. on carbon dioxide CO_2) in the water. The model that is used in this paper is of NPZD type. Here N is the model variable for inorganic nitrogen and P the model variable for phytoplankton (note: not for phosphorus!). Additionally two model variables Z and D are used to represent the amount of zooplankton and detritus (particulate dead organic matter). Phytoplankton P serves as nutrient for Z, whereas both P, Z contribute to D, which itself contributes to the amount of N. The dependencies between the four model variables are schematically depicted in Fig. 1, a more detailed mathematical description is given in the next section.

There are lots of marine biogeochemical models, see e.g. again [2, Section 4.3]. The NPZD model with four variables (or *species*) used here was presented and studied e.g. in [3,4]. It lies in the middle range of complexity; there are models just taking into account N and P, at the other extremum there are models that distinguish between different types of phyto- and/or zooplankton. It is not clear whether a finer model (in the sense that it incorporates more species), is better in the sense that it can better reproduce given measurement data.

This is one reason why parameter estimates and optimization is important: It can help to tune and improve models, and to show their limitations, thus giving rise to extend or vary them. The other reason of course is the necessary validation of all model simulations and predictions that shall be used in political advice and decision making. For the model under consideration, parameter optimization studies have been performed in [5–7]. Therein, mostly stochastic optimization methods as Evolutionary/Genetic Algorithms (EA/GAs) have been used. The use of gradient-based optimization methods was limited by the applied algorithms that had problems with the necessary box constraints for the parameters and the state variables. It can be summarized that the optimization to match real measurement data showed that

- no optimal fit of the data could be achieved,
- and a wide spread of suboptimal parameters that give a similar, significant deviation from the data was obtained.

A natural consequence would be to ask if the model is sufficient to fit the realistic data at all, or if it has to be extended or changed substantially. To provide an even more profound answer to this questions, our aims in this paper are the following:

- We extend the optimization methods to
  - a new variant of an EA called Quantum-Genetic Algorithm,
  - superlinear convergent local search methods that take into account the parameter bounds and make use of algorithmically generated exact derivative information.
- We show how both methods can be combined in a hybrid algorithm. We discuss which variants are applicable for the model.
- With this tool at hand, we perform
  - parameter optimization for model-generated, i.e. attainable *twin data*, and for the real measurements.
  - uncertainty analysis to study the above-mentioned parameter spread, to show how Gaussian disturbed data effect the optimization results.

The structure of the paper is as follows: In the next section we describe the model, the used data and the cost function to be optimized. In the third section we present the hybrid optimization algorithm and its main parts, namely the quantum evolutionary algorithm and the gradient-based optimizer. Afterwards we present our results with respect to optimization and uncertainty analysis and end the paper with some conclusions.