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 ($\mu$GA) 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 $K_\rho$ 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 $\alpha, \kappa$ 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,

| index | symbol | value/ range | unit (d=86400 s) | parameter |
|---|---|---|---|---|
| | $C_{\text{ref}}$ | 1.066 | 1 | growth coefficient |
| | $c$ | 1 | $^\circ\text{C}^{-1}$ | growth coefficient |
| | $R$ | 6.625 | 1 | molar carbon to nitrogen ratio |
| | $k_w$ | 25 | $\text{m}^{-1}$ | PAR extinction length |
| | $f_{\text{PAR}}$ | 0.43 | 1 | short-wave PAR fraction |
| 1 | $\beta$ | $[0, 1]$ | 1 | assimilation efficiency of zooplankton |
| 2 | $\mu_m$ | $\mathbb{R}_0^+$ | $\text{d}^{-1}$ | phytoplankton growth rate parameter |
| 3 | $\alpha$ | $\mathbb{R}_0^+$ | $\text{m}^2\text{W}^{-1}\text{d}^{-1}$ | slope of photosynthesis vs light intensity |
| 4 | $\Phi_m^z$ | $\mathbb{R}_0^+$ | $\text{d}^{-1}$ | zooplankton loss rate |
| 5 | $\kappa$ | $\mathbb{R}_0^+$ | $\text{m}^2(\text{mmol N})^{-1}$ | light attenuation by phytoplankton |
| 6 | $\epsilon$ | $\mathbb{R}_0^+$ | $\text{m}^6(\text{mmol N})^{-2}\text{d}^{-1}$ | grazing encounter rate |
| 7 | $g$ | $\mathbb{R}_0^+$ | $\text{d}^{-1}$ | maximum grazing rate |
| 8 | $\Phi_m^p$ | $\mathbb{R}_0^+$ | $\text{d}^{-1}$ | phytoplankton linear mortality |
| 9 | $\Phi_z^*$ | $\mathbb{R}_0^+$ | $\text{m}^3(\text{mmol N})^{-1}\text{d}^{-1}$ | zooplankton quadratic mortality |
| 10 | $\gamma_m$ | $\mathbb{R}_0^+$ | $\text{d}^{-1}$ | detritus remineralization rate |
| 11 | $k_N$ | $\mathbb{R}_0^+$ | $\text{mmol Nm}^{-3}$ | half saturation for $NO_3$ uptake |
| 12 | $w_s$ | $\mathbb{R}_0^+$ | $\text{m d}^{-1}$ | detritus sinking velocity |

**Table 2:** Model parameters. Parameters with a index in the first column were optimized. The ranges given are theoretical, the ranges actually used are given in table 5.

- chlorophyll a (Chl a) (in mg (Chl a) $m^{-3}$), corresponding to variable $P$, here using a constant conversion factor of 1.59 mg (Chl a)/(mmol N),
- zooplankton biomass (ZOO) (in mmol N $m^{-3}$), corresponding to $Z$,
- particulate organic nitrogen (PON) (in mmol N $m^{-3}$), corresponding to $P + Z + D$,
- and carbon fixation or primary production as carbon uptake (here abbreviated as PP, in mmol C $m^{-3}$ $d^{-1}$), corresponding to the term $J(\mu, u)PR$, with $R$ being the Redfield ratio, see [14] and table 2.

Except for zooplankton biomass, the tracer variable value is its concentration at the corresponding depth. For zooplankton biomass, the tracer value is the vertically averaged concentration in the water column from the given depth (approximately 200 meters) to the ocean surface. In order to attempt to estimate total zooplankton biomass from measured mesozooplankton biomass, the observed zooplankton ZOO is transformed to $Z_{obs} = 1.23 \cdot ZOO + 0.097$, see also [19]. Except for zooplankton, only data in the euphotic zone, translating into the upper 20 model layers, is considered. To compare modeled primary production $PP$ with observations from 24-hour incubation measurements, the 24-hour mean of the model output $J(\mu, u)PR$ is taken.

### 2.2 Initial Values and Forcings

An initial vertical concentration profile for $N$ is calculated as the mean depth profile derived from the available DIN observations, the components of the other profiles $P, Z, D$ are set to small values following [10], see figure 1. The $NPZD$ model is forced by output from the OCCAM global circulation model, see [6], namely the hourly vertical profiles of temperature $T$ (in °C) and vertical diffusivity $K_\rho$ (in $m^2 s^{-1}$), respectively. The time resolution $\Delta t$ of the forcing data is one hour, and the vertical grid consists of 66 layers with thickness increasing with depth. The PDE is integrated by operator splitting. The source minus sink equations (2) are integrated using four explicit Euler steps with $\frac{1}{4}\Delta t$ after which the advection equation in (1) is also solved by an explicit Euler step and then the diffusion equation in (1) is solved by an implicit step (Thomas method).

## 3 The Optimization Problem

The optimization problem consists of finding a set of constant parameters giving a minimal misfit to the BATS data for the years 1991-1995 as defined by a cost function. The problem is of least-squares type with box constraints and can be written as

$$\min_{x \in \mathbb{R}^n} F(x) := \|f(y(x)) - y_d\|^2 \quad s.t. \quad l \leq x \leq u, \quad F : \mathbb{R}^n \to \mathbb{R}, \tag{3}$$

where the parameters are summarized in the vector $x \in \mathbb{R}^n$, see again table 2.

Here $y$ shall denote the discrete values of the four variables $N, P, Z, D$ on the space-time grid used by the model. The vector $y_d$ denotes the observational data, or in the case of twin experiments, data precomputed by the model itself. The function $f$ stands for transformations on the variables as in section 2.1.

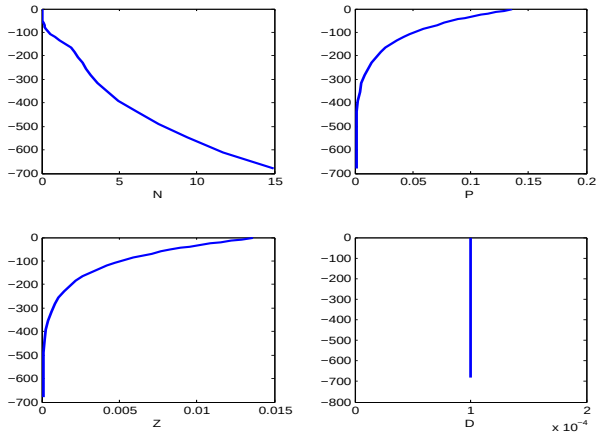


**Figure 1:** Original initial values.

The norm in $F$ is an Euclidean norm weighted by assumed standard deviations of the measurements $\sigma = (\sigma_j)_{j=1,\dots,5} = (0.1\,\mathrm{mmolNm}^{-3}, 0.01\,\mathrm{mg(Chla)m}^{-3}, 0.01\,\mathrm{mmolNm}^{-3}, 0.0357\,\mathrm{mmolNm}^{-3}, 0.025\,\mathrm{mmolCm}^{-3}\mathrm{d}^{-1})$ for the five respective data types introduced above, the number of years for which data exists for each data type, and the respective numbers of measurements. For instance, zooplankton data are available only from 1994 onwards. For details of the cost function see [16].
The bound constraints on the parameters are to be understood componentwise for the vectors $l, u \in \mathbb{R}^n$ with the actual numbers given in table 5. A main restriction is that the state variables $N, P, Z, D$ have to be non-negative. This additional constraint is not treated in the optimization, only the optimized product is checked for this condition.

## 3.1 Global Optimization

We are interested in a possibly unique global optimum, i.e. a minimal misfit $F$ within the given bounds $l, u$ on the parameters $x$. It is known that for methods using local information (function values, derivatives) only, to converge for any function $f$ to its global minimizer it is necessary and sufficient to evaluate a sequence of points that is dense in the parameter space. This obviously

amounts to evaluating all (up to some given tolerance) points and can only be improved upon by using additional global information. For this density theorem and additional information see [5]. With no such global information available, the simplest incomplete (i.e. no guarantees on global convergence) algorithm is the *multiple random start method* which performs a number of local optimizations from random starting points for the parameter vector. In the absence of information or heuristics on how to choose starting points (global information), these are chosen independent and identically distributed (i.i.d.) within the bounds $l, u$.
The *multiple random start method* is easily parallelized in an obvious way.

## 3.2 Sequential quadratic local Optimization Method

The desired properties of the local optimization method include:

- superlinear convergence rate without using second derivatives,
- treatment of bound or box constraints to avoid irrelevant results,
- off-the-shelf availability as source or library without additional coding needed,
- flexibility to make use of exact derivative information provided by tools of Algorithmic/Automatic Differentiation (AD), i.e. no restriction to built-in finite differences.

We used the software CFSQP which stands for *Feasible Sequential Quadratic Programming in C*, see [13], which is well suited for the problem on hand. It can moreover treat general nonlinear constraints. The SQP method treats the constraints via Lagrange multipliers and solves the optimality system by a Quasi-Newton method, thus providing superlinear convergence in the ideal case while requiring only first derivatives.

## 3.3 Algorithmic Differentiation

If the coded function is smooth enough, efficient code for exact derivatives can be generated automatically by the technology of *Algorithmic or Automatic Differentiation* (AD), see e.g. [3]. This was the case here, with only two non differentiable terms, one being the minimum evaluation in $J(\mu, u)$, see table 1, the other being a safeguard in the integration of the reaction equations assuring non-negativity. Since the occurring *min* and *max* functions are directionally differentiable, the AD software has no problem to generate a correct directional derivative that is sufficient for a local optimizer. Moreover, the use of AD-generated gradients naturally avoids approximation errors and instabilities which are well-known for numerically approximated derivatives. It is also superior w.r.t. the convergence speed in an optimization, since the implemented gradient is exactly the one of the implemented discretized function to be optimized, a fact that is assumed in all convergence results that are analytically proven.

We used the proprietary AD software TAF (Transformations of Algorithms in Fortran), see [2], that generates new source code from the original one, thus providing a code that evaluates $F$ and its gradient simultaneously. We used TAF

| # parameters ($n$) | | relative cpu time for evaluation of $F$ | $F + \nabla F$ (AD forward) | $F + \nabla F$ (finite diff.) |
|---|---|---|---|---|
| 12 | mean | 1 | 10.5 | 13 |
| | min | 1 | 7.8 | 13 |
| 140 | mean | 1 | 48.3 | 141 |
| | min | 1 | 56.5 | 141 |

**Table 3:** Relations of cpu times for function ($F$) and gradient ($\nabla F$) computations, the latter generated algorithmically with AD software and approximately by finite-differences. The "min" in the second column refers to the time for evaluation of $F$.

in forward mode, i.e. propagating derivative information in the same order as is used for evaluating the function itself.

The forward mode's effort to evaluate $F$ and $\nabla F$ is $\mathcal{O}(n)$ times the effort of one function evaluation itself, where $n$ is the number of independent parameters to be optimized. Here we give some numbers based on actual optimization runs. In our case with $n = 12$ (on an AMD Shanghai) one function evaluation costs on average 2.14s with a minimum of 1.62s (736 evaluations). The mean runtime for the model plus the gradient was 22.5s with a minimum of 12.5s (370 evaluations). The mean ratio was 10.5 and the ratio of the minima was 7.8.

In our case with initial conditions used as control parameters as well, yielding $n = 140$, (on an AMD Barcelona) the respective numbers for the model were 1.40s, 1.08s (520 evaluations) and for model plus gradient 67.6s, 61.3s (301 evaluations) with a mean ratio of 48.3 and that of the minima of 56.5. These numbers may vary wildly due to circumstances (computational environment) not under the user's control.

A finite difference approximation to the gradient would cost at least $n + 1$ function evaluations. Especially in the case of $n = 140$, AD in the forward mode clearly did beat this, compare table 3 which summarizes these results.

It is well known that for computing the derivative of a function $F : \mathbb{R}^n \to \mathbb{R}$, as is needed in all single-objective optimization, the reverse mode (also known as adjoint mode) generally is superior w.r.t. performance if the number $n$ of optimization parameters gets large. This is because in this case the number of elementary operations (+,-,*,...) to evaluate $F$ and $\nabla F$ is $\mathcal{O}(1)$ times that for the function $F$ alone. This comes at the price of having to store (or recompute) every intermediate variable of the whole calculation. Naive implementation can slow the calculation practically to a halt. This usually can be overcome and so using the reverse mode is highly recommended when cpu time grows too much.

# 4 Preliminary tests with synthetic data

To assess the capabilities of the chosen optimization algorithm, we tested its ability to reconstruct i.i.d. target parameter vectors $x^i \in \mathbb{R}^n, i = 1, \ldots, 64$ from their simulated data $y_d^i := y(x^i)$ taken at the points of the BATS observations for the years 1991-1995.

In general, i.e. without restriction to target parameter vectors that lead to feasible states $y \geq 0$ component-wise for all depths and times, these reconstruction problems could not be solved, i.e. the parameter vectors were not unique;

e.g. a vector yielding a cost $F < 10^{-14}$ could differ in ten components from 20 to 270 percent from its target vector. A heuristic restriction on the feasibility of the state variables showed better results: We restricted the set of otherwise i.i.d. parameter vectors $x^i$ (by elimination after evaluation) to those which yielded a state component $y_2 = P > 10^{-10}$ for all points of their trajectories. All those vectors had a parameter $x_8 = \Phi_m^p < 0.015$, thus this seems to be a reasonable and easy way to achieve feasibility. In our test experiments, no restrictions based on other components of the state vector (N, Z, D) turned out to be necessary.

For the reconstruction of these $x^i$, 30 trials each were undertaken, starting from i.i.d. starting points $x^{i_k}, k = 1, \ldots, 30$. In 63 (out of a total 64) cases, the $x^i$ were reconstructed with a precision of four significant digits, mostly multiple times. In the remaining case the component-wise error was up to 5 percent.

This is a good sign in three respects: Firstly, an optimization method that could not reconstruct such parameters would definitely be ruled out. Secondly, a mathematical model not yielding to inverse methods would be suspect, or at least would ask for more inquiry. Thirdly, the temporal and spatial observation schedule available at BATS appears sufficient to make the NPZD model under investigation here fully observable.

On the contrary, we take this as an indication that the random restart method using the local SQP method with appropriately set stopping criteria is a promising candidate for solving the parameter optimization problem in our case, at least as long as the optimal trajectories do not include parts with too low concentrations of $P$.

# 5 Spatial Model Reduction

We find that for the purpose at hand and with used observational data restricted to the upper 194 m (i.e., the upper 20 levels) the model can be reduced to 32 instead of the original 66 depth levels. This corresponds to a deepest spatial gridpoint of 679 m instead of 6366 m. Optimization for parameters gives practically identical results w.r.t. parameters and misfit $F$ for both models thus defined. Thus we argue that the reduced model can represent the larger one.
This is only justified for short time periods as the five year interval considered here. Tests with synthetic (5-year cyclic) forcing for longer periods on the order of hundreds of years show that for given parameters (those optimized later on) much of the total mass sinks below the euphotic zone when using 66 levels. Overall mass conservation holds, though, as is easily verified numerically. This effect is much less pronounced for 32 levels. In other words, the trajectories of the respective models tend to diverge after some time. For that reason alone, different optimal parameters can be expected.
Further reducing the depth extent, e.g. to 24 levels leads to considerably different optimal parameters (biologies), albeit with a comparable misfit. Such a model was considered as unable to represent the original one.

# 6 Sensitivity of Parameters w.r.t. Data Error

In this section we solve our optimization problem for the years 1991-1995 with an additional spin-up year 1990, i.e. we start the model at 1.1.1990 but no

| $x_i$ | $l_i$ | $u_i$ | min$x_i$ | max$x_i$ | $s_i$ | $\bar{x}_i$ | $x_i^*$ |
|---|---|---|---|---|---|---|---|
| $F$ | 0.0000 | 0.0000 | 69.4820 | 71.9196 | 0.3305 | 70.6951 | 69.9368 |
| $\beta$ | 0.3000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 1.0000 |
| $\mu_m$ | 0.2000 | 1.4600 | 1.1555 | 1.2056 | 0.0069 | 1.1798 | 1.1831 |
| $\alpha$ | 0.0010 | 0.2530 | 0.1028 | 0.1211 | 0.0014 | 0.1070 | 0.1068 |
| $\Phi_m^z$ | 0.0000 | 0.6300 | 0.0471 | 0.0519 | 0.0006 | 0.0502 | 0.0498 |
| $\kappa$ | 0.0100 | 0.7300 | 0.0730 | 0.0730 | 0.0000 | 0.0730 | 0.0730 |
| $\epsilon$ | 0.0250 | 4.0000 | 4.0000 | 4.0000 | 0.0000 | 4.0000 | 4.0000 |
| $g$ | 0.0400 | 4.0000 | 4.0000 | 4.0000 | 0.0000 | 4.0000 | 4.0000 |
| $\Phi_m^p$ | 0.0000 | 0.6300 | 0.0028 | 0.0037 | 0.0001 | 0.0034 | 0.0034 |
| $\Phi_z^*$ | 0.0100 | 1.0000 | 0.0310 | 0.0456 | 0.0020 | 0.0360 | 0.0373 |
| $\gamma_m$ | 0.0100 | 0.1500 | 0.0100 | 0.1017 | 0.0122 | 0.0391 | 0.0447 |
| $k_N$ | 0.1000 | 1.0000 | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 1.0000 |
| $w_s$ | 2.0000 | 128.0000 | 128.0000 | 128.0000 | 0.0000 | 128.0000 | 128.0000 |

**Table 4:** The $x_i^*$ are the optimal parameters of section 6. $\bar{x}_i$ , $s_i$ , min$x$ and max$x$ are the results of the uncertainty analysis of section 6.1.

observations for this first year enter the cost function. Using a single spin-up year for upper-ocean one-dimensional ecosystem models is common practice, and we will return to this in section 7.

Out of 320 random starts we get 180 minima with identical misfit (up to six significant digits) and a set of optimal parameters with less than 1% spread. The mean of these optimal parameters is denoted as $x^*$ in table 4, where also the misfit $F$ is given. It is also the lowest $F$ found.

This, of course, is no proof that the global minimum is found, as discussed in section 3.1. The fact that the local optimization method is so heavily attracted by one (and so far the best) local minimum together with the overall goodnaturedness of the problem as witnessed in section 4 convinces us that it is unlikely to find a better one (with this method) and to tentatively accept this one as global.

## 6.1 Uncertainty of parameters due to observational error

To assess the uncertainty (or variance) of the optimal parameters w.r.t. observational error we use the direct way of repeatedly adding noise to the data and optimizing again for the parameters. In this way no theoretical work is needed and quick results are available when we start the optimization from $x^*$. This is justified by tests using the restart method as before with 30 i.i.d. starting points. Furthermore, any error distribution can be assessed this way, although we confine ourselves to normal error with standard deviations $\sigma$ as in section 3.

2560 such optimizations result in mean $\bar{x}_i$, empirical standard deviation $s_i$, and spread [min $x_i$, max $x_i$] as given in table 5. A histogram of the parameters showing variance is seen in figure 2. It shows a roughly bell shaped distribution for these parameters.

## 6.2 Results

At this point, we note the following two results:

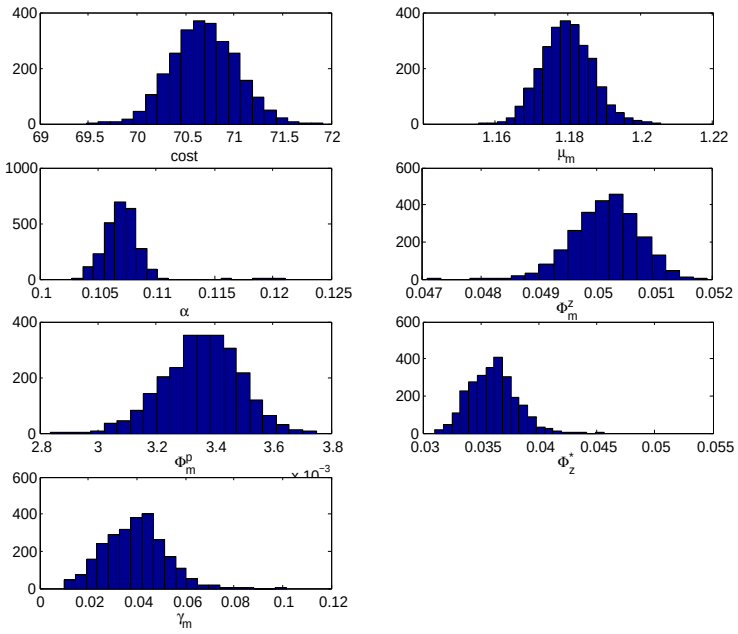


**Figure 2:** Histogram of 2560 parameter optimizations for normal noise added to data, section 6.1. Parameter bounds as in table 4.

- Using the random restart method with an SQP solver and AD-provided exact gradients, we find an unique optimal parameter set when initial values, data and boundary conditions (forcing) are held fixed.
- Based on Gaussian noise added to the data,with everything else held fixed, for the given bounds we get narrow estimates for 9 of the 12 parameters. Six parameters are at the upper bounds, for the assimilation efficiency $\beta = 1$ this also implies that the terms $(1 - \beta)$ turn out to be irrelevant

in the reaction equations. For the parameters $\Phi_m^p, \Phi_z^*, \gamma_m$ we find standard deviations of $3\%, 6\%, 30\%$ of their respective mean value. Such we believe that only limited uncertainty in the optimal parameters is due to observational error, provided the given $\sigma$ are somewhat realistic.

# 7 Sensitivities w.r.t. initial Values

## 7.1 Simultaneous Optimization of Parameters and initial Values using a spinup year

As mentioned in section 6, so far the optimizations were done using a single spin-up year. The underlying assumption is that at times much longer than the internal time scales of the ecological processes resolved by the model (of the order of days, as evident from the range of parameter values in Table 4), initial values are largely irrelevant and the trajectories are forced in the 'right direction' by the model parameters and the boundary conditions (i.e. the forcing). We understand this assumption to mean that at least for all globally optimal parameter sets $x^{opt}$ (we haven't established uniqueness) the trajectories (and therefore the misfits) are independent of the initial values. To check this, we optimize the parameters and the initial values together with bounds as in section 7.2. Our best result out of 40 starts is seen in table 5 as $\hat{x}_i^*$ with the misfit $\hat{F} = 67.78$. Therefore $x^*$, which gives $F = 69.94$, cannot be the globally optimal set with the claimed property. It also cannot be the set $\hat{x}_i^*$ which gives a misfit of $F = 136.27$ when using the original initial values. If there is such a global optimizer, we have not found it.

## 7.2 Simultaneous Optimization of Parameters and initial Values without using a spinup year

With the assumed loss of information about the inititial conditions during the spin-up year being much in doubt, we turn to the simultaneous optimization of parameters and initial values without using a spinup year, i.e. we start the optimization for 1991-1995 at 1.1.91. This does not necessarily mean a lifting of restrictions, since we do not know what kind of mapping the spin-up year represents with regard to the state variables. However, it turns out that still lower values of the cost function $F$ can be reached that way.

The bounds $l_i^j, u_i^j$ (with $j = n, p, z, d$ for the four state variables $N, P, Z, D$, and $i = 1, \ldots, 32$) on the initial values are based on the initial values $v_i^j$ used in the previous sections and shown in figure 1. For the lower bound we take

$$l_i^j = \max\{0, \min\{0.1 v_i^j, v_i^j - 2 s_j\}\}$$

and for the upper bound

$$u_i^j = \max\{2 v_i^j, v_i^j + 2 s_j\},$$

with values $s_j = 0.1, 0.01, 0.01, \text{and} 0.02\,\text{mmol N m}^{-3}$ for $j = 1, 2, 3, 4$, roughly corresponding to the weights $\sigma_j$ given in section 3.

We use 120 random starts for the optimization. The distribution of the minima obtained is displayed in figure 3.

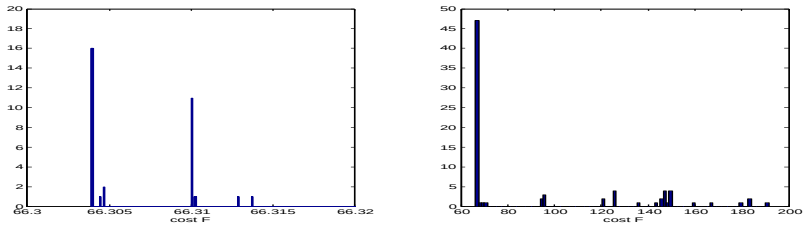


**Figure 3:** Histogram of cost function values obtained for simultaneous optimization of parameters and initial values (section 7.2). Left a zoom.

We get 16 identical (w.r.t. parameters) results $x^{**}$ with $F = 66.30$. The initial values also show little variance for these 16 hits, see figure 4. Again we tentatively accept this as an unique optimum (albeit only w.r.t. the parameters) within the given bounds.

We see in table 5 that seven parameters lie outside the spread found in the sensitivity test. This is especially prominent for the parameters $\Phi^z_m, \Phi^p_m, \Phi^*_z$ and $w_s$. To put it differently, for these parameters the optimal values are not found without optimizing for the initial values, too.

| $x_i$ | min $x_i$ | max $x_i$ | $x_i^*$ | $\hat{x}_i^*$ | $x_i^{**}$ | $x_i^{***}$ |
|---|---|---|---|---|---|---|
| $F$ | 69.4820 | 71.9196 | 69.9368 | 67.7760 | 66.3039 | 65.8409 |
| $\beta$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $\mu_m$ | 1.1555 | 1.2056 | 1.1831 | 1.3742 | 1.2530 | 1.3109 |
| $\alpha$ | 0.1028 | 0.1211 | 0.1068 | 0.0765 | 0.1298 | 0.1312 |
| $\Phi^z_m$ | 0.0471 | 0.0519 | 0.0498 | 0.0301 | 0.0166 | 0.0175 |
| $\kappa$ | 0.0730 | 0.0730 | 0.0730 | 0.0399 | 0.0730 | 0.0730 |
| $\epsilon$ | 4.0000 | 4.0000 | 4.0000 | 4.0000 | 4.0000 | 4.0000 |
| $g$ | 4.0000 | 4.0000 | 4.0000 | 4.0000 | 4.0000 | 4.0000 |
| $\Phi^p_m$ | 0.0028 | 0.0037 | 0.0034 | 0.0053 | 0.0015 | 0.0017 |
| $\Phi^*_z$ | 0.0310 | 0.0456 | 0.0373 | 0.1543 | 0.2225 | 0.2142 |
| $\gamma_m$ | 0.0100 | 0.1017 | 0.0447 | 0.1500 | 0.1500 | 0.1500 |
| $k_N$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $w_s$ | 128.0000 | 128.0000 | 128.0000 | 9.2566 | 9.5248 | 9.0023 |

**Table 5:** min $x_i$, max $x_i$ and $x_i^*$ as in table 4. $\hat{x}_i^*$ is the result of section 7.1. $x_i^{**}$ and $x_i^{***}$ are the results of section 7.2 and section 7.2.1, respectively.

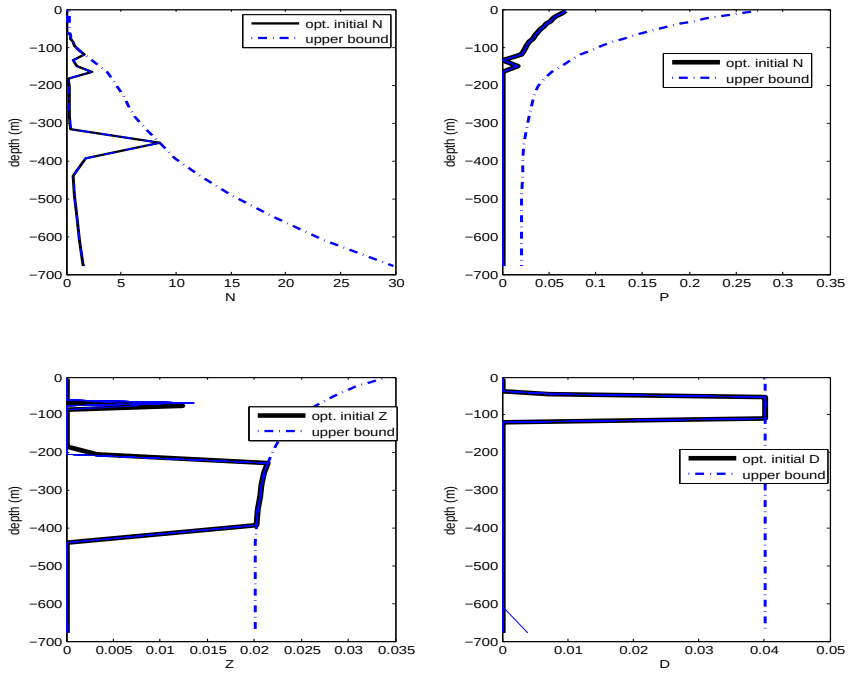


**Figure 4:** Optimized initial values of section 7.2 and upper bound used. Blurred lines are the only visible differences in 16 hits.

### 7.2.1 Broader Bounds

Now we broaden the lower bounds on the initial values substantially to $l_i^j = 0$, for $j = n, p, z, d$ and the upper bounds to

$$u_i^n = \max\{2, 3v_i^n\}, u_i^p = 2, u_i^z = 1, u_i^d = 1$$

and do 200 random starts.

Here we are interested in the best misfit and do not claim to be able to identify the initial state with any precision. We see no conceptual difference between a (possible) particularly bad fit in the (unknown) initial values and one at any other (temporal) point. Given the magnitude of the misfits obtained so far, we feel justified to allow for relatively large bounds.

For the broader bounds, we get only one distinct optimizer $x^{***}$ with a misfit of $F = 65.84$, see table 5, but a cluster of minima nearby as seen in figure 6. The representation of the solution space by the local optimization method indicates that a substantially higher number of optimizations would be needed to build trust in a minimum based on multiple identical hits. We conjecture, that since

$x^{***}$ (and its most direct neighbors in terms of misfit $F$) differ only marginally from $x^{**}$, that we are very near the optimizer within the bounds prescribed on the parameters $x_i$.

The optimized initial values are well within their prescribed bounds as depicted in figure 5.

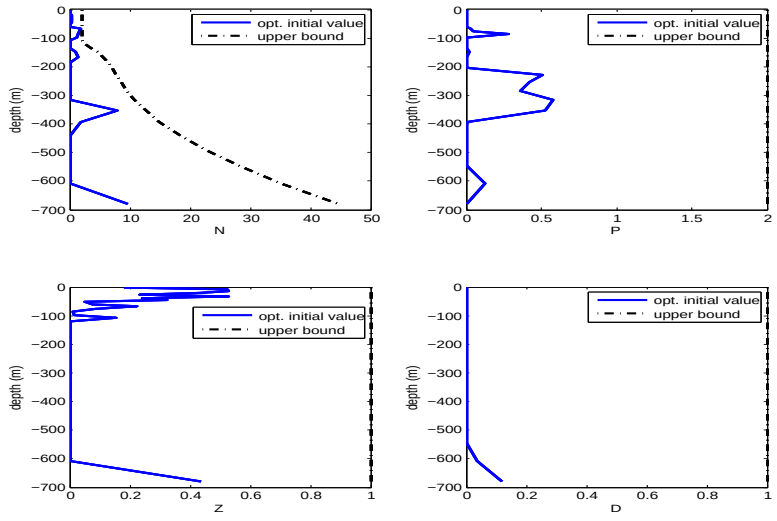


**Figure 5:** Optimized initial values, model of section 7.2.1.

For this final result of our optimization efforts in terms of misfit, see figures 7, 8. The trajectories of the model's state variables compared with those of the original optimization of section 6 are depicted in figures 9, 10, 11, 12.

### 7.2.2 Dependency on Error in initial Values

Here we step back to the problem in section 6 to assess how much uncertainty in initial values may affect optimal parameters. This is interesting in the case that some initial values are insisted upon, e.g. when using observations. To investigate this further, we add Gaussian noise to the 'original' initial values of figure 1, i.e. we treat these values as measurements subject to error and optimize only for the parameters. We use 640 initial distributions and start the optimization from the 'optimal' parameters $x^*$ of section 6.

For the assumed standard deviations of the error in the initial conditions, we find for parameter $\gamma_m$ a standard deviation of 10% of the mean and less

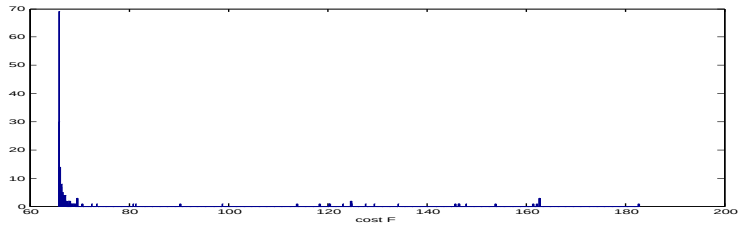


**Figure 6:** Histogram of optimization of parameters and initial values in section 7.2.1.

than 0.5% for the others, see $x^\sigma, s^\sigma$ together with $x^*$ in table 6. Only when we allow a normal distribution with the $\sigma_{n,p,z,d}$ multiplied by five, we get for $\gamma_m$ a standard deviation of 50% of the mean and less than 3% for the others.

We conclude, that for initial values known up to measurement error, this error adds only little uncertainty to optimized parameters.

Often, measurements will be available for N only, whereas P, Z, D have to be estimated. To assess the respective contributions of the state variables we do two tests with the results seen in table 6:

- First we add Gaussian noise with $\sigma_{n,p,z,d}$ multiplied by 20 to get mean $x_i^{20\sigma}$ and standard deviation $s_i^{20\sigma}$.
- Secondly we repeat this test with the original $\sigma_n$ and only $\sigma_{p,z,d}$ multiplied by 20 to get $x_i^N, s_i^N$.

The results of both tests are also displayed in table 6.

As a second test we optimize (using the spinup) firstly for the parameters and the initial N-profile and secondly for the parameters and the initial P,Z,D-profiles simultaneously, in each case holding the other profiles fixed at the 'original' values of figure 1. Again the bounds were as in section 7.2. In the first case we get 4 identical results (in $F$ with $F = 67.94$) after 80 starts. For the parameters $\mu_m, \alpha, \phi_m^z, \kappa, \phi_m^p, \phi_z^*, w_s$ the relative errors (i.e. $|\hat{x}_i^* - x_i|/\hat{x}_i^*$) are $1.6, 3.8, 6.4, 12.2, 6.7, 7.0$ and 16% respectively.

In the second case no definite result is obtained. The optimizing algorithm does only find a minimum $x$ at $F = 69.78$ if started at $x^*$ and the original initial values. Hundreds of random starts yield only suboptimal results. For this optimum we get relative errors $|x_i^* - x_i|/x_i^*$ of $0.4, 1.7, 0.6, 3.8, 6.7$ and 17.4% for the parameters $\mu_m, \alpha, \phi_m^z, \phi_m^p, \phi_z^*, w_s$.

We see that although much of the variance in the optimized parameters can be attributed to the uncertainty in the initial $N$-profile, $P, Z$ and $D$ can not

be chosen completely arbitrary. To achieve variances on the same order as in section 6.1 all profiles will have to be determined with some accuracy.

| $x_i$ | $x_i^*$ | $x_i^\sigma$ | $s_i^\sigma$ | $x_i^N$ | $s_i^N$ | $x_i^{20\sigma}$ | $s_i^{20\sigma}$ |
|---|---|---|---|---|---|---|---|
| $F$ | 69.9368 | 69.9495 | 0.0373 | 70.1853 | 0.0651 | 70.7074 | 0.8207 |
| $\beta$ | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |
| $\mu_m$ | 1.1831 | 1.1825 | 0.0040 | 1.1614 | 0.0107 | 1.1034 | 0.0987 |
| $\alpha$ | 0.1068 | 0.1071 | 0.0004 | 0.1149 | 0.0050 | 0.1216 | 0.0117 |
| $\Phi_m^z$ | 0.0498 | 0.0498 | 0.0001 | 0.0495 | 0.0004 | 0.0469 | 0.0034 |
| $\kappa$ | 0.0730 | 0.0730 | 0.0000 | 0.0730 | 0.0000 | 0.0730 | 0.0000 |
| $\epsilon$ | 4.0000 | 4.0000 | 0.0000 | 4.0000 | 0.0000 | 4.0000 | 0.0000 |
| $g$ | 4.0000 | 4.0000 | 0.0000 | 4.0000 | 0.0000 | 4.0000 | 0.0000 |
| $\Phi_m^p$ | 0.0034 | 0.0034 | 0.0000 | 0.0032 | 0.0001 | 0.0025 | 0.0011 |
| $\Phi_z^*$ | 0.0373 | 0.0370 | 0.0001 | 0.0349 | 0.0007 | 0.0421 | 0.0081 |
| $\gamma_m$ | 0.0447 | 0.0419 | 0.0042 | 0.0107 | 0.0023 | 0.0281 | 0.0347 |
| $k_N$ | 1.0000 | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9439 | 0.0984 |
| $w_s$ | 128.0000 | 128.0000 | 0.0000 | 128.0000 | 0.0000 | 127.3863 | 5.5200 |

**Table 6:** Results of section 7.2.2.

# 8 Conclusions

We summarize our results as follows.

- Based on our experience, the random start method using a SQP local solver with AD generated derivatives gives reliable and quick results for parameter optimization problems encountered here. This is true even for relatively high number of free parameters as when initial values are optimized, too. The performance though, depends on the parameter space.
- With fixed initial values, only little variance in the optimal parameters due to observational uncertainty is found with the notable exception of $\gamma_m$.
- Small disturbances as normal error in the initial values will also have only small influence on the optimal parameters, again with the exception of the detritus remineralization rate $\gamma_m$.
- Using optimized initial values and parameters gives results that lie way outside the uncertainty inflicted by measurement error alone. When there is no cogent reason to use particular initial values, e.g. observations, to get the best fit and the matching parameters, simultaneous optimization for initial values and parameters is required.
- The use of a spin-up without optimization of the initial conditions does not change the situation. It still gives suboptimal results.

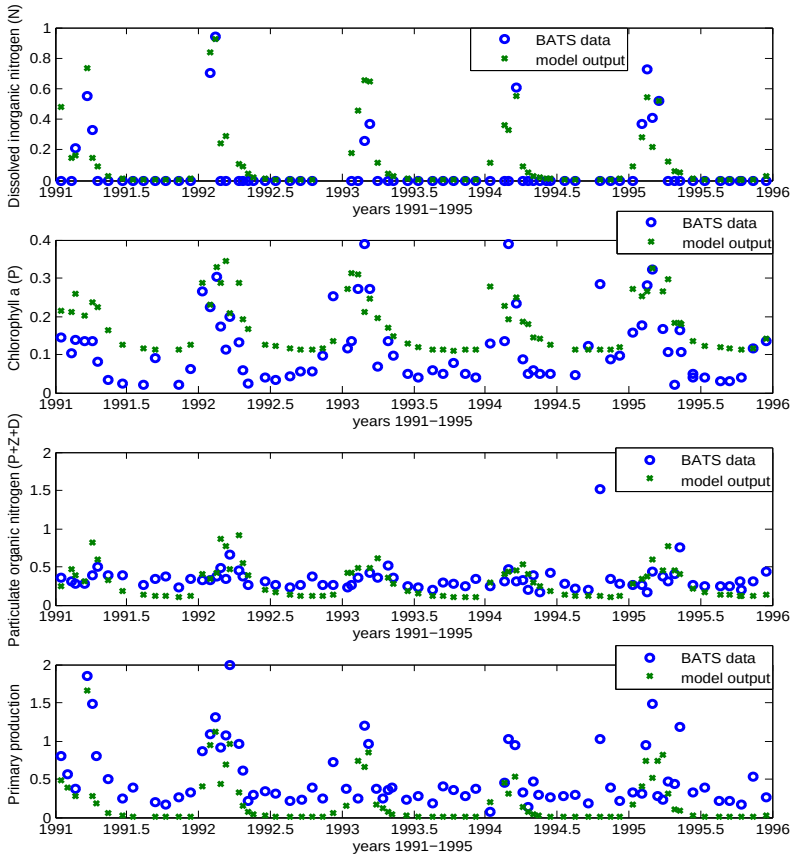


**Figure 7:** Fit to data in depth 0-5 meters for optimized model of section 7.2.1

# References


[1] G. T. Evans and J. S. Parslow, *A model of annual plankton cycles*, Biological Oceanography **3** (1985), 328–347.

[2] R. Giering and T. Kaminski, *Recipes for Adjoint Code Construction*, ACM Trans. Math. Software **24** (1998), no. 4, 437–474.

[3] Andreas Griewank, *Evaluating derivatives principles and techniques of algorithmic differentiation*, Frontiers in Appl. Math., no. 19, SIAM, Philadelphia, PA, 2000.

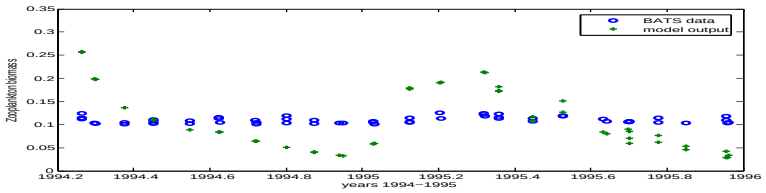


**Figure 8:** Fit to zooplankton data for optimized model of section 7.2.1

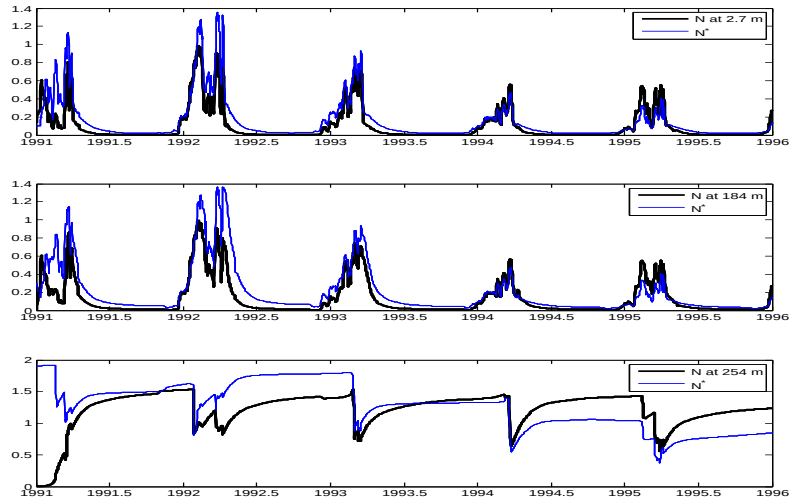


**Figure 9:** Comparison between the best N-trajectory ($x^{***}$ of section 7.2.1) and the result of optimization of parameters only $N^*$ of section 6.

[4] K. H. Han and J. H. Kim, *Quantum-inspired evolutionary algorithm for a class of combinatorial optimization*, IEEE Transactions on Evolutionary Computation **6(6)** (2002), 580–593.

[5] A. Neumaier, *Complete Search in Continuous Global Optimization and*

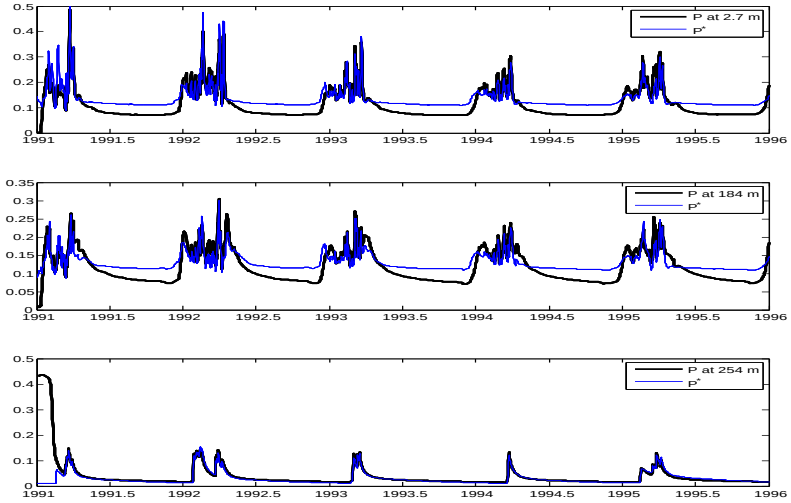


**Figure 10:** Comparison between the best P-trajectory ($x^{***}$ of section 7.2.1) and the result of optimization of parameters only $P^*$ of section 6.

*Constraint Satisfaction*, Acta Numerica 2004, 271–369.

[6] NOCS, *Occam ocean model, http://www.noc.soton.ac.uk/jrd/occam.*

[7] A. Oschlies, *Nutrient supply to the surface waters of the North Atlantic – a model study*, Journal of Geophysical Research **107** (2002), doi:10.1029/2000JC000275.

[8] A. Oschlies, *Feedbacks of biotically induced radiative heating on upper-ocean heat budget, circulation, and biological production in a coupled ecosystem-circulation model*, Journal of Geophysical Research **110** (2004), doi:10.1029/2004JC002430.

[9] A. Oschlies and V. Garcon, *An eddy-permitting coupled physical-biological model of the north Atlantic. 1. sensitivity to advection numerics and mixed layer physics*, Global Biogeochemical Cycles **13** (1999), 135–160.

[10] A. Oschlies, W. Koeve, and V. Garcon, *An eddy-permitting coupled physical-biological model of the north Atlantic. 2. ecosystem dynamics and comparison with satellite and jgofs local studies data*, Global Biogeochemical Cycles **14** (2000), 499–523.

[11] A. Oschlies, and P. Kähler, *Biotic contribution to air-sea fluxes of $CO_2$ and $O_2$ and its relation to new production, export production, and net community production*, Global Biogeochemical Cycles **18** (2004), GB1015, doi:10.1029/2003GB002094.

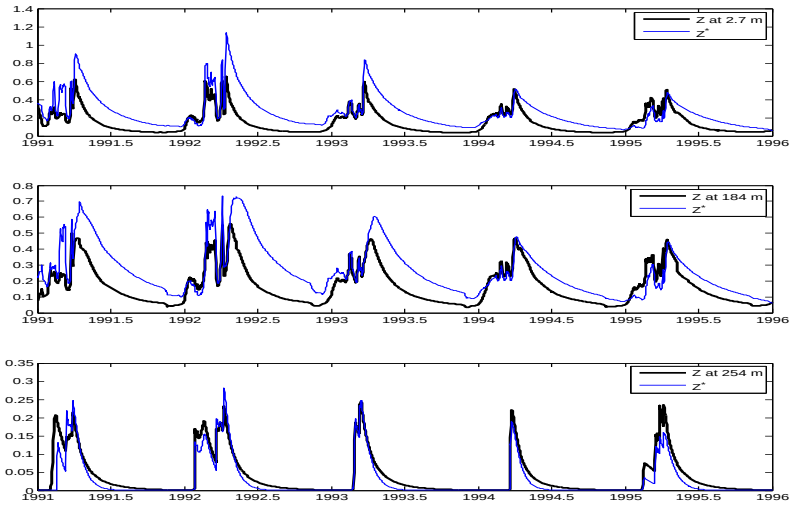


**Figure 11:** Comparison between the best Z-trajectory ($x^{***}$ of section 7.2.1) and the result of optimization of parameters only $Z^*$ of section 6.

[12] A. Oschlies, and M. Schartau, *Basin-scale performance of a locally optimised marine ecosystem model*, Journal of Marine Research **63** (2005), 335–358.

[13] E. Panier and A. L. Tits, *On combining feasibility, descent and superlinear convergence in inequality constrained optimization*, Mathematical Programming **59** (1993), 261–276.

[14] A.C. Redfield, B.H. Ketchum, and F.A. Richard, *The influence of organisms on the composition of sea water*, The Sea (M.N. Hill, ed.), Wiley, New York, 1963, pp. 26–77.

[15] M.R. Roman, D.A. Caron, and P. Kremer, *Spatial and temporal changes in the partitioning of organic carbon in the plankton community of the sargasso sea of Bermuda*, Deep-Sea Research I **42(6)** (1995), 922–973.

[16] J. Rückelt, V. Sauerland, T. Slawig, A. Srivastav, B. Ward, C. Patvardhan, *Parameter Optimization and Uncertainty Analysis in a Model of Oceanic CO2-Uptake using a Hybrid Algorithm and Algorithmic Differentiation*, Nonlinear Analysis B Real World Applications (2010), doi:10.1016/j.nonrwa.2010.03.006 (c) Elsevier Ltd.

[17] M. Schartau and A. Oschlies, *Simultaneous data-based optimization of a 1d-ecosystem model at three locations in the north Atlantic: Part i - method and parameter estimates*, Journal of Marine Research **61** (2003), 765–793.

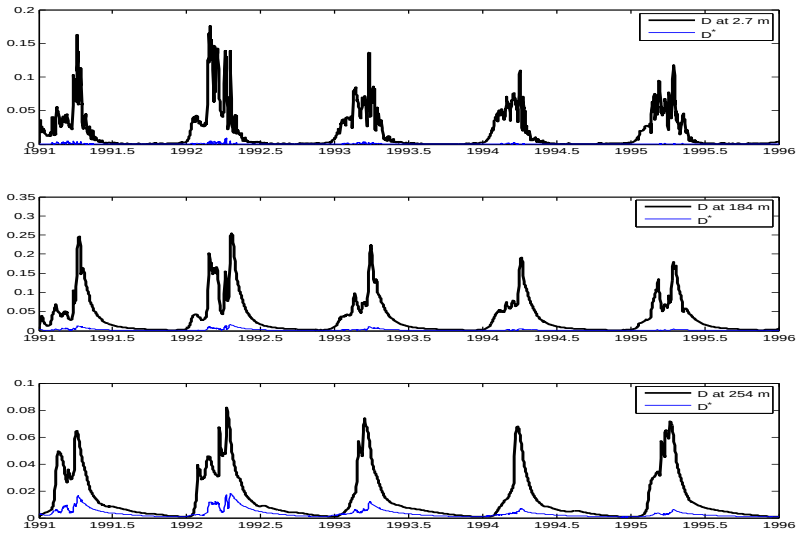


**Figure 12:** Comparison between the best D-trajectory ($x^{***}$ of section 7.2.1) and the result of optimization of parameters only $D^*$ of section 6.

[18] Ward, B.A., Friedrichs, M.A.M., Anderson, T.R., Oschlies, A., 2010. *Parameter optimization techniques and the problem of underdetermination in marine biogeochemical models* Journal of Marine Systems **81**: 1-2, 34-43.

[19] B. Ward, *Marine ecosystem model analysis using data assimilation*, Ph.D. thesis, http://web.mit.edu/benw/www/Thesis.pdf, 2009.