ELSEVIER

Journal of Marine Systems 28 (2001) 19–44

JOURNAL OF MARINE SYSTEMS

www.elsevier.nl/locate/jmarsys

A weak constraint inverse for a zero-dimensional marine ecosystem model

L.-J. Natvik *, M. Eknes, G. Evensen

Nansen Environmental and Remote Sensing Center, Edv. Griegsvei 3A, 5059 Bergen, Norway

Received 29 September 1999; accepted 20 September 2000

Abstract

Data assimilation for nonlinear ocean models can be extremely complicated. A typical marine ecosystem model consists of equations that are both nonlinear and coupled to each other, and the data assimilation problem is therefore nontrivial.

While several papers describe data assimilation techniques for estimating the model parameters in marine ecosystem models, this paper discusses a different approach. We let the model parameters have fixed values and instead allow for an unknown error term in each of the model equations. Thus, the model is included as a weak constraint in a variational formulation. To be more specific, unknown error terms are assumed for the model, the initial conditions and the data, and a weak constraint cost function containing weighted squares of the unknown error terms is to be minimized. For this purpose, we use very simple iterative gradient descent schemes, where each iterate is based on the gradient of the cost function. An ordinary gradient steepest descent scheme is compared to a more elaborate conjugate gradient scheme denoted by the Fletcher and Reeves method. Several data assimilation experiments are performed in which the data are generated from an "exact" forward model integration (twin experiments). To make the experiments more realistic, normally distributed noise is added to the measurements. Properties like the sensitivity with respect to data density and length of the assimilation interval are investigated.

Also, a twin experiment with measurements of only the phytoplankton component is performed. Remotely sensed ocean colour data represent the major part of currently available observations and data sets of this type may be used to measure the phytoplankton compartment. Thus, showing that it is possible to obtain good results also for the other ecosystem components when only phytoplankton is measured seems to be of vital importance for creating a marine biochemical data assimilation system.

Keywords: Data assimilation; Weak constraint inverse; Zero-dimensional marine ecosystem model

1. Introduction

Data assimilation methods combine information contained in a dynamical model with a set of measurements, and a more realistic estimate of the true state can be found. A dynamical model will always be a simplification of the truth. For example, many processes and interactions are probably not included in the dynamical equations. Further, small scale processes may be included using inappropriate parameterization schemes or poorly known parameter values. Also, the initial and boundary conditions may not be well known. A set of observations contains

* Corresponding author. Tel.: +47-55-29-72-88; fax: +47-55-20-00-50.

E-mail address: larsj@nrsc.no (L.-J. Natvik).

0924-7963/01/$ - see front matter
PII: S0924-7963(00)00082-8

information about the true ocean state. However, a data set may be sparse in space and time, and may not resolve small scale processes. Also, different sources of error have probably influenced the measurements. By combining all information available in a data assimilation scheme, one can expect to find a better estimate of the true field than the different sources of information provide by themselves.

Data assimilation for linear ocean models is well understood, see e.g. Bennett (1992). However, the extension to nonlinear dynamics is nontrivial. As a consequence of the chaotic behavior that can be seen in many nonlinear models, the data assimilation problem can become extremely complicated (Miller et al., 1994; Evensen, 1997). In this paper, an inverse problem for a very simple ecosystem model is discussed. It will be illustrated that nonlinearities in the dynamical model equations complicate the inverse problem, but that one still can find realistic estimates in many cases. In our twin experiments, we use an "exact" forward model integration as a reference solution, from which we also generate measurements, see Sections 4 and 5.

It is the parameter estimation problem that has received most attention in connection with data assimilation for ecosystem models. In fact, all of the following papers discuss data assimilation as a tool for estimating model parameters. In Spitz et al. (1998), a variational adjoint method was used to assimilate both simulated data and data from the BATS (Bermuda Atlantic Time-series Study) site. An adjoint technique was also used by Lawson et al. (1995, 1996) in identical twin experiments. In Prunet et al. (1996a,b), a variational integral (cost function) was minimized directly using an iterative Gauss–Newton scheme, and in Fasham and Evans (1995), a corresponding variational problem was solved using a gradient free conjugate direction method. Simulated annealing is another method used in parameter estimation experiments by Matear (1995) and Hurtt and Armstrong (1996). Finally, a Markov chain Monte Carlo method was investigated by Harmon and Challenor (1996), where a Metropolis–Hastings algorithm was implemented to sample the error statistics of the model parameters.

The motivation for developing parameter estimation techniques when assimilating data into ecosystem models is obvious since these models often include a large number of parameters of more or less well known values. However, in this paper, the model parameters are kept fixed with predefined values. Thus, we do not investigate a parameter estimation problem. Instead, we let each of the model equations contain an unknown error term, and these error terms are then minimized in a least squares sense. That is, the model is regarded as not being perfect, and it is included in a variational formulation as a so-called weak constraint. This will be more fully explained in Section 2. A popular approach in physical oceanography has been to use a so-called strong constraint approximation. However, using the strongly nonlinear and coupled Lorenz model, Miller et al. (1994) reported an extreme sensitivity with respect to the initial conditions as a consequence of including the dynamics as a strong constraint (see also Section 5.4).

Data assimilation methods may be divided into two major groups, namely filtering and smoothing methods, see Gelb (1979) or Bennett (1992). In a true filter, information only from "present time" observations is used to get an improved estimate. Thus, a sequential algorithm can be formulated, where the dynamics are integrated forward in time and improved estimates are found at measurement times. In a smoothing method, an estimate is found based on information from both past and future observations.

In this paper a weak constraint formulation will be used to find a smoothing estimate. This approach has not yet been used in data assimilation experiments for ecosystem models. The representer method is an optimal smoothing method for linear dynamics, as one obtains an analyzed field as a forecast plus a linear combination of M_{obs} representers or influence functions, one for each measurement. Thus, the solution is a member of an M_{obs}-dimensional space. The representer technique is extensively explained in Bennett (1992), and also in Eknes and Evensen (1997). For nonlinear dynamics, an iterative scheme can be used, where a linear inverse problem can be solved using the representer method in each iteration (Bennett and Thorburn, 1992; Evensen et al., 1998). However, it is believed that this approach will not handle too strong nonlinearities (Evensen et al., 1998), and other methods should be used in such cases.

We choose to minimize the weak constraint formulation directly using gradient descent methods, where the model solution as a function of time (and space) is treated as the control variable (see Section 3 for details). Two variants will be used, namely gradient steepest descent (GSD) and a conjugate gradient procedure denoted by the Fletcher and Reeves method (FR). Both methods are iterative, and the gradient with respect to the full discretized state has to be found in each iteration (The state vector $\boldsymbol{\psi}$ contains the discrete model variables on a finite time interval, see Section 3 for more information). Since our model (described in Appendix A) is very simple containing three ecosystem variables and being time dependent only, the evaluation of the gradient will be relatively simple (the gradient is given in Appendix B). Also, the numerical storage requirements will be moderate for a zero-dimensional approach (see Section 3). The gradient descent methods are relatively easy to implement. As will be seen Section 3, no model integrations are performed, instead a search direction calculated from the gradient is substituted at every iteration. A gradient descent procedure has previously been investigated by Evensen (1997) and Evensen and Fario (1997). They showed that the method worked well with the strongly nonlinear Lorenz model. A weakness is the possibility of convergence to a stationary point not being the global minimum. The main reasons for choosing gradient descent methods here are simplicity and the successful performance with the Lorenz model. For more information on optimization, see e.g. Shanno (1978), Gill et al. (1982), Minoux (1986) or Kincaid and Cheney (1991).

A model by Evans and Parslow (1985) will be used in this paper. It consists of one physical variable, the mixed layer depth M, as input, and three ecosystem variables, nutrients N, phytoplankton P and herbivores H, and is dependent of time only (zero dimensional). (Actually, light intensity is also used as a physical input as described in Appendix A). For our purpose, it is of interest to use a simple model. However, it should include the type of nonlinearities which are typical for more general ecosystem models. Thus, one should start with a simple and somewhat general model, and several assimilation experiments should be carried out before extensions to more elaborate ones are made. Note also that at this time there are not enough data to constrain all the variables in a multi-compartment model. The Evans and Parslow (1985) model describes a simple annual cycle of nutrients, phytoplankton and herbivores, a full description is given in Appendix A. It is convenient to define $\boldsymbol{\Psi}^T(t) = [N(t),\ P(t),\ H(t)]$. The coupled set of model Eqs. (23)–(25) may then be written as

$$\frac{\mathrm{d}\boldsymbol{\Psi}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{\Psi}, t), \tag{1}$$

where $\boldsymbol{f}$ contains the right hand sides of the equations.

In Section 2, the weak constraint formulation (cost function) is discussed, and a hypothesis about the prior statistics is given. The gradient descent methods used to minimize the cost function are described in Section 3. In Section 4, additional information about the prior statistics is given. Also, the setup of the twin experiments is explained in this section. Section 5 contains the description of the different experiments, where the main goal is to investigate how the nonlinearities in the dynamics complicate the inversion. To be more specific, it is interesting to see how close the inverse estimate is to the reference solution in the twin experiments. Properties like the sensitivity with respect to data density (Case 2) and length of the assimilation interval (Case 4) are investigated. Also, an experiment with measurements of phytoplankton only is discussed (Case 3). There are also three appendices. The model is summarized in Appendix A and the gradient is described in Appendix B. Finally, a list describing the nomenclature is given in Appendix C.

2. Inverse formulation

In this section, the inverse formulation is introduced, mainly using the notation of Bennett (1992). The model equations are certainly not perfect, several assumptions were made when deriving them (for example, the model is zero-dimensional, it does not include more than three ecosystem compartments, the physical forcing is too simple to be realistic, etc.) Therefore, it is natural to allow for an unknown model error term or dynamical misfit in each of the

model equations. The error terms are denoted by q^N, q^P and q^H, and Eq. (1) now becomes

$$\frac{\mathrm{d}\boldsymbol{\Psi}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{\Psi}, t) + \mathbf{q}(t), \tag{2}$$

where $\mathbf{q}^T = [q^N,\ q^P,\ q^H]$.

Unknown error terms $\mathbf{a}^T = [a^N,\ a^P,\ a^H]$ for the initial conditions are also assumed

$$\boldsymbol{\Psi}(0) = \boldsymbol{\Psi}_0 + \mathbf{a}, \tag{3}$$

where $\boldsymbol{\Psi}_0^T = [N_0,\ P_0,\ H_0]$ are first guess values for the initial conditions. The system of differential equations together with the initial conditions would have a unique solution if all the error terms were zero (or for any realization of the error terms). This is valid provided that the function $\boldsymbol{f}(\boldsymbol{\Psi}, t)$ is bounded and continuous in t and Lipschitz continuous in the variables contained in $\boldsymbol{\Psi}$, see e.g. Theorem 7.19 in Apostol (1969), or McOwen (1996).

Now, we assume that in addition to the dynamical model, a set of measurements $\mathbf{d}$ of the true solution is given. These are assumed to be linearly related to the model variables contained in $\boldsymbol{\Psi}$ by the measurement equation

$$\mathbf{d} = \mathscr{L}[\boldsymbol{\Psi}] + \boldsymbol{\epsilon}, \tag{4}$$

where $\mathscr{L}$ is a vector of linear measurement functionals and $\boldsymbol{\epsilon}$ is a vector of unknown measurement errors.

If all the error terms were zero, the system (Eqs. (2)–(4)) would be over-determined. This is easily seen, since the initial value problem would have a unique solution that would not generally coincide with the data. The idea here is to allow for the errors to get the system under-determined, and then try to minimize them with a least squares procedure. This should provide a better estimate of the state variables than the dynamical model or the data can give when used separately, see e.g. Bennett (1992).

Additional information about the prior error statistics is introduced through a statistical hypothesis. It is assumed that the mean of the errors are zero, i.e.

$$\begin{aligned} \overline{\mathbf{q}(t)} &= \mathbf{0}, \\ \bar{\mathbf{a}} &= \mathbf{0}, \\ \bar{\boldsymbol{\epsilon}} &= \mathbf{0}, \end{aligned} \tag{5}$$

and that the covariances are known as

$$\begin{aligned} \overline{\mathbf{q}(t_1)\mathbf{q}^T(t_2)} &= \mathbf{Q}_{\mathbf{qq}}(t_1, t_2), \\ \overline{\mathbf{a}\mathbf{a}^T} &= \mathbf{Q}_{\mathbf{aa}}, \\ \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T} &= \mathbf{w}^{-1}, \end{aligned} \tag{6}$$

where the overbar denotes ensemble averaging. Further, any cross-correlations between the different error terms $\mathbf{q}$, $\mathbf{a}$ and $\boldsymbol{\epsilon}$ are assumed to be zero. Allowing such cross-correlations would clearly make the cost function (defined below) much more complicated and could make the inversion more difficult. Also, it is certainly difficult enough to get good estimates of our prior error covariance matrices without dealing with cross-correlations. Setting the cross-correlations to zero is believed to be a realistic assumption in most cases. As long as the initial conditions are chosen independently of our data and model, $\overline{\mathbf{a}\boldsymbol{\epsilon}}$ and $\overline{\mathbf{q}\mathbf{a}}$ will probably be small. If the model errors $\mathbf{q}$ are essentially white in time, the initial conditions can even be chosen from a model spin-up using the same model without violating $\overline{\mathbf{q}\mathbf{a}} = \mathbf{0}$ too much. Further, as long as our data set has not been used to calibrate the model, $\overline{\mathbf{q}\boldsymbol{\epsilon}} = \mathbf{0}$ should be valid (when saying that the cross-covariances are small, we mean relative to the covariances (6)). In our twin experiments, we sample the errors of the measurements and initial conditions by adding random Gaussian noise to the "reference solution" (see Section 4). This is consistent with assuming that $\overline{\mathbf{q}\mathbf{a}} = \mathbf{0}$, $\overline{\mathbf{q}\boldsymbol{\epsilon}} = \mathbf{0}$ and $\overline{\mathbf{a}\boldsymbol{\epsilon}} = \mathbf{0}$.

Now, we wish to minimize the unknown error terms $\mathbf{q}(t)$, $\mathbf{a}$ and $\boldsymbol{\epsilon}$ in the interval $[0, T']$. To be more specific, this may be done by a weak constraint formulation, where the variational integral,

$$\begin{aligned} \mathscr{J}[\boldsymbol{\Psi}] = &\int_0^{T'}\int_0^{T'} \mathbf{q}^T(t_1)\mathbf{W}_{\mathbf{qq}}(t_1, t_2)\mathbf{q}(t_2)\,\mathrm{d}t_1\mathrm{d}t_2 \\ &+ \mathbf{a}^T\mathbf{W}_{\mathbf{aa}}\mathbf{a} + \boldsymbol{\epsilon}^T\mathbf{w}\boldsymbol{\epsilon}, \end{aligned} \tag{7}$$

where $\mathscr{J}[\boldsymbol{\Psi}]$ is the cost or penalty function, is to be minimized. In Eq. (7), $\mathbf{W}_{\mathbf{qq}}$ and $\mathbf{W}_{\mathbf{aa}}$ are the dynamical and the initial weight matrices, respectively. These are 3×3 matrices where 3 corresponds to the number of variables. The measurement weight matrix, denoted by $\mathbf{w}$, is $M_{\mathrm{obs}} \times M_{\mathrm{obs}}$ where M_{obs} is the

number of measurements. The weights are symmetric and positive definite, see Bennett (1992) and Daley (1991). The dynamical weight $\mathbf{W_{qq}}$ is a functional inverse of the dynamical error covariance $\mathbf{Q_{qq}}$. This means that

$$\int_0^{T'} \mathbf{Q_{qq}}(t_1,t_2)\mathbf{W_{qq}}(t_2,t_3)\mathrm{d}t_2 = \boldsymbol{\delta}(t_1 - t_3), \qquad (8)$$

where $\boldsymbol{\delta}(t_1 - t_3)$ is a 3×3 identity matrix whenever $t_1 = t_3$ and a 3×3 zero matrix whenever $t_1 \neq t_3$. Further, $\mathbf{W_{aa}}$ is the inverse of the initial error covariance matrix $\mathbf{Q_{aa}}$ and $\mathbf{w}$ is the inverse of the measurement error covariance matrix $\mathbf{w}^{-1}$. The penalty function defined above is a minimum error variance estimator (the weight matrices are defined as functional inverses of the corresponding error covariance matrices).

Note that other estimators than least squares could have been defined. The least squares formulation is, however, an attractive formulation for several reasons. For example, if the unknown error terms, **q**, **a** and $\boldsymbol{\epsilon}$, are Gaussian, i.e. completely explained by the first two statistical moments mean and covariance, minimizing the penalty function (7) is equivalent to maximizing the multi-variate probability function defined as

$$p(\mathscr{J}[\boldsymbol{\Psi}]) = A_{\mathscr{J}} \mathrm{e}^{-\mathscr{J}[\boldsymbol{\Psi}]}, \qquad (9)$$

where $A_{\mathscr{J}}$ is a normalizing constant. Note that whenever the errors fail to be normally distributed, $\mathscr{J}$ is still the minimum error variance, but no longer the maximum likelihood estimate. Further, some solution methods require that the Euler–Lagrange equations are derived. This is possible from the above cost function $\mathscr{J}$ and the use of the fundamental lemma of calculus of variations (Courant and Hilbert, 1953). In this paper, gradient descent methods, which minimize the cost function directly, are used. By a proper choice of weights, the smoothness of $\mathscr{J}$ is ensured and the gradient exists everywhere. For more information on inverse methods and data assimilation, see e.g. Gelb (1979), Jazwinski (1970) or Bennett (1992).

3. Numerical solution methods

As said before, the weak constraint formulation will be minimized directly using gradient descent methods, where the model solution as a function of time is treated as the control variable. We start by discretizing the model variables, i.e. let $\boldsymbol{\psi}$ be a vector containing the variables N, P and H at every time step of a finite time interval.

As mentioned in Section 1, two algorithms for minimizing $\mathscr{J}$ will be described, namely gradient steepest descent (GSD) and the Fletcher and Reeves method (FR). Gradient steepest descent is a very simple method, but a slow convergence is seen in many cases, and this is the main reason for comparing it to a conjugate gradient method having better convergence properties.

As in Minoux (1986), global convergence means that no matter starting point, the algorithm converges to a point where a necessary optimality condition is satisfied. This condition is normally $\nabla\mathscr{J} = \mathbf{0}$, as is the case here. Note that if $\mathscr{J}$ is continuously differentiable and convex, the stationarity condition above is both necessary and sufficient to ensure the point to be a global minimum (Minoux, 1986). Remember that the weights in the cost function are positive definite. For linear dynamics, $\boldsymbol{f}$ is linear and therefore all the terms of $\mathscr{J}$ are quadratic. Since all positive definite quadratic forms are convex, they have a unique global minimum. However, when $\boldsymbol{f}$ is nonlinear as here, the dynamical term is non-quadratic.

So, generally the methods used are only capable of finding a stationary point of $\mathscr{J}$. The nonlinear dynamical term may not destroy convergence to the global minimum as long as it is small, i.e. the dynamical weight $\mathbf{W_{qq}}$ is small compared to the others. Whenever $\mathbf{W_{qq}}$ is too large, however, the procedure may only capture a point of stationarity.

The gradient descent procedures are iterative, where a descent or search direction $\mathbf{s}_l$ and a step length λ_l are found at every iteration. The new state is given as $\boldsymbol{\psi}_{l+1} = \boldsymbol{\psi}_l + \lambda_l \mathbf{s}_l$, where $\boldsymbol{\psi}_l$ contains the state variables N_l, P_l and H_l at every time step for a given iteration l. Thus, no model integrations are needed. The search direction is based on the gradient of the cost function with respect to the full state contained in $\boldsymbol{\psi}$. The gradient may be written as

$$\nabla_{\boldsymbol{\psi}} \mathscr{J}_{N'}[\boldsymbol{\psi}], \qquad (10)$$

where the subscript N' denotes the number of discrete points in the time interval. The gradient for our model is given in Appendix B.

Simple finite difference formulas are used for the derivatives in the model equations. For the time derivatives of a scalar function $\phi(t)$, the approximation is centered in the interior, i.e.

$$\left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{t_i} = \frac{\phi_{i+1} - \phi_{i-1}}{2\Delta t} + \mathcal{O}(\Delta t^2). \tag{11}$$

At the end points, one-sided differences are used.

A direction $\mathbf{s}_l$, along which the directional derivative is negative, is called a descent direction. By Taylor expanding the cost function, it is easy to see that $\mathcal{J}_{N'}(\boldsymbol{\psi}_l + \lambda_l \mathbf{s}_l) < \mathcal{J}_{N'}(\boldsymbol{\psi}_l)$ if λ_l is sufficiently small. That is, $\mathcal{J}_{N'}$ is reduced along $\mathbf{s}_l$.

In the gradient steepest descent method (GSD), the search direction $\mathbf{s}_l$ is simply chosen to be in the direction of the negative gradient (see e.g. Minoux (1986) or Kincaid and Cheney (1991)). Further, the so-called "Goldstein criteria" are used to determine the step length λ_l (Goldstein, 1967),

$$\mathcal{J}_{N'}(\boldsymbol{\psi}_l + \lambda_l \mathbf{s}_l) \leq \mathcal{J}_{N'}(\boldsymbol{\psi}_l) + \kappa_1 \lambda_l \left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_l)\right)^T \mathbf{s}_l, \tag{12}$$

$$\mathcal{J}_{N'}(\boldsymbol{\psi}_l + \lambda_l \mathbf{s}_l) \geq \mathcal{J}_{N'}(\boldsymbol{\psi}_l) + \kappa_2 \lambda_l \left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_l)\right)^T \mathbf{s}_l, \tag{13}$$

where $\kappa_1 \in [0,1)$ and $\kappa_2 \in (\kappa_1,1)$. These criteria take into account that a too large λ_l may lead to oscillatory behavior, while a too small may lead to premature convergence.

The conjugate gradient scheme chosen is denoted by the Fletcher and Reeves method (FR) (Fletcher and Reaves, 1964). This particular method corresponds to a particular choice of the parameter β_l below. The search direction at step l is given by

$$\mathbf{s}_l = -\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_l) + \beta_l \mathbf{s}_{l-1}, \tag{14}$$

where β_l in the Fletcher–Reeves method is chosen as

$$\beta_l = \frac{\left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_l)\right)^T \left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_l)\right)}{\left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_{l-1})\right)^T \left(\nabla \mathcal{J}_{N'}(\boldsymbol{\psi}_{l-1})\right)}. \tag{15}$$

The initial search direction $\mathbf{s}_0$ is chosen to be the negative gradient. Further, a one-dimensional minimizer is used to find a λ_l at every iteration step l of FR.

At each step of GSD, a lower value of $\mathcal{J}_{N'}$ is obtained, so the method provides a decreasing sequence. Since $\mathcal{J}_{N'} \geq 0$ (bounded from below), we know that the algorithm must converge globally to a stationary limit point. That is, the routine will always converge no matter starting point (the term "global" refers to this), but it may stop in any stationary point, i.e. in any local minimum or saddle point. Note that the Goldstein criteria provide a step length small enough to ensure the descent property (Minoux, 1986). To ensure global convergence of the FR method, the algorithm is restarted with the value of the negative gradient; global convergence now follows from the GSD case (Minoux, 1986). (Restarting means that the search direction is chosen to be just the negative gradient every time the number of itera-

Table 1
Approximate storage requirements (RAM) for marine ecosystem models implemented with GSD and FR (32-bit machine)

	Numerical scheme[a]	GSD	FR
Zero-dimensional model	Specified storage	$2mn_t$	$3mn_t$
	Number of variables	$6*10^3$	$9*10^3$
	Computer resources	24 Kbytes	36 Kbytes
One-dimensional model	Specified storage	$2mn_t n_z$	$3mn_t n_z$
	Number of variables	$1.2*10^5$	$1.8*10^5$
	Computer resources	$\approx$ 0.5 Mbytes	$\approx$ 0.7 Mbytes
Three-dimensional model	Specified storage	$2mn_t n_x n_y n_z$	$3mn_t n_x n_y n_z$
	Number of variables	$2.7*10^9$	$4.1*10^9$
	Computer resources	$\approx$ 11 Gbytes	$\approx$ 16 Gbytes

[a] m—number of model variables; n_x, n_y, n_z—number of grid points, horizontal and vertical directions; n_t—number of grid points, time grid. It has been assumed that: $m = 3$, $n_x = n_y = 150$, $n_z = 20$ and $n_t = 10^3$.

tions reach the value of a predefined restarting parameter).

It is seen that the GSD method requires a storage (RAM) of order $2mN'$ variables, where m is the number of model equations and N' is the number of grid points. These are needed for the storage of $\boldsymbol{\psi}$ and $\nabla\mathscr{J}_{N'}(\boldsymbol{\psi})$ at each iteration step l. The FR scheme requires $3mN'$ variables for the storage of the search direction $\mathbf{s}_l$, the state vector $\boldsymbol{\psi}_l$ and the gradient $\nabla\mathscr{J}_{N'}(\boldsymbol{\psi}_l)$. The storage requirements are summarized in Table 1 for our zero-dimensional approach, together with eventual extensions to one-dimensional and three-dimensional models. It is observed that the storage requirements for the three-dimensional case are somewhat inconvenient; remember that local variables come in addition. However, a strength of the weak constraint formulation is that it has the property of forgetting very past and very future information. Therefore, there is a possibility of solving a problem in local time intervals, and using the final conditions in one assimilation as initial conditions for the next.

4. Error covariances and weights

The discrete dynamical weight $\mathbf{W}_{\mathbf{qq}}(i, j)$ is generally varying for any pair (i, j). However, it is not a realistic task to find estimates of covariances (weights) between all discrete points i and j. This would also require storage of one 3×3 matrix for each pair (i, j). One approximation is to assume that $\mathbf{W}_{\mathbf{qq}}(i, j)\approx \mathbf{W}_{\mathbf{qq}}(|i-j|)$, i.e. the weights are only dependent of the distance between points i and j. Here we choose to remove the time correlation in the dynamical weight, i.e.

$$\mathbf{W}_{\mathbf{qq}}(i,j) = \hat{\mathbf{W}}_{\mathbf{qq}}\delta(i-j), \tag{16}$$

where $\hat{\mathrm{W}}_{\mathbf{qq}}$ is a constant 3×3 matrix and $\delta(i-j)$ is Dirac's delta function. This must, however, be regarded as an over-simplification, and to increase the time regularity, a smooth field constraint,

$$\mathscr{J}^S_{N'}[\boldsymbol{\psi}] = \Delta t\sum_{i=1}^{N'} \boldsymbol{\eta}_i^T\mathbf{W}_{\boldsymbol{\eta}\boldsymbol{\eta}}\boldsymbol{\eta}_i, \tag{17}$$

is added to the cost functional. In this expression, $\boldsymbol{\eta}_i^T = (\eta_i^N,\ \eta_i^P,\ \eta_i^H)$ and contains the second order time derivatives of the model variables which are approximated by finite difference formulas.

To conclude, the time correlation in the general $\mathbf{W}_{\mathbf{qq}}(i, j)$ has been removed, and the time regularity is accounted for by including the term (Eq. (17)) in the cost function. However, note that we are seeking smooth fields N, P and H instead of smooth errors by using the smoothing norm (Eq. (17)). At least for linear dynamics, there is a unique correspondence between an error covariance matrix and a smoothing norm (McIntosh, 1990). This suggests that the dynamical error terms should have been used in the smoothing norm. However, note that by applying the smoothing norm to the model variables instead, the conditioning of the gradient descent methods will be improved since only smooth functions are searched for (non-smooth N, P and H are not realistic). Further, it is believed that the auto-correlation function for each of the dynamical misfits is similar to the corresponding function for the inverse estimate since the dynamical error terms are strongly dependent on the model solution. Thus, the smoothing weights contained in $\mathbf{W}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ correspond to auto-correlation functions for the model state which in turn should be similar to the auto-correlation functions for the dynamical misfit terms since they are functions of the model state.

To generate measurements, the model equations were integrated forward in time using a standard ODE-solver (Hindmarsh, 1983), and data points were picked directly from the solution curves. To make the experiments more realistic, normally distributed noise was added to the measurements. Since the model equations were solved with a quite high accuracy (10^{-6} mmol m^{-3} for all three variables), the solution curves may be taken as the "true" state. In this way, the error statistics are known in advance, and it is interesting to see how well the data assimilation methods are capable of recreating the "true" solution curves. This procedure is often referred to as a twin experiment in the literature, and it is simply a tool for testing if a data assimilation method is able to capture a "true" (known) state. One might argue that this procedure is not very useful because it does not imply that the data assimilation method will work for real measurements. In particular, one might

ask questions about the statistical hypothesis in the real case. However, these are problems that have to be dealt with after the twin experiments have been executed successfully. Of course, the prior statistical hypothesis may have to be changed for a real data set, and it should be based on an analysis of the data set used. It must be expected that a too poor hypothesis will cause a breakdown of the data assimilation method. The twin experiments give a good indication of how the method will work when the prior statistics are chosen in a consistent way.

The model was first spun up by integrating it for a 3-year period, and the end conditions were then used as initial conditions for all the experiments. In Evans and Parslow (1985), it was verified that the model entered a stable cycle after a long time run, and we found that realistic annual cycle initial conditions are provided at the end of the third year, when a stable cycle is obtained.

The smoothing, the initial and the measurement weight matrices will all be assumed diagonal. That $\mathbf{W}_{\boldsymbol{\eta\eta}}$ is diagonal corresponds to the assumption that each of the variables are to be smooth, but that the smoothness of one variable is not correlated to the smoothness of another. It is believed that such off-diagonal terms are small (compared to the diagonal terms) and therefore may be set to zero. In our twin experiments, the second order derivatives can be found from the "reference solution". Assuming $\mathbf{W}_{\mathbf{aa}}$ to be diagonal implies that the errors of the initial conditions are uncorrelated. That is, the error a^N is not correlated to the errors a^P and a^H. A diagonal $\mathbf{w}$ implies that the measurement errors are uncorrelated. This means that the error $\boldsymbol{\epsilon}^N$ in point i is uncorrelated to $\boldsymbol{\epsilon}^P$ and $\boldsymbol{\epsilon}^H$ in point i and in all other points, and also uncorrelated to $\boldsymbol{\epsilon}^N$ in all points different from i. Choosing $\mathbf{W}_{\mathbf{aa}}$ and $\mathbf{w}$ diagonal is consistent with sampling the initial conditions and the measurements from the reference solution and then adding independent Gaussian errors. As said before, for experiments with real measurements, different prior statistics must be specified through an analysis of the data set used.

We chose the set of weights shown in Table 2. For the measurement and initial weights, corresponding standard deviations are also given, the errors in the measurements may also be viewed in later figures where both the reference solution and the measurements are plotted. The values in Table 2 were obtained in the following way. In our twin experiments, we sampled the measurements by adding Gaussian noise with a known (predefined) variance to the reference solution. In the same way, the initial condition was sampled by adding noise to the first

Table 2
The weights

Weight matrix	Part of weight mat.	Symbol	Value [unit]	Std. dev. [unit]
First guess	N	$\mathbf{W}_{fg}^{11}$	1.0 [(m mol m^{-3})$^{-2}$]	1.0 [m mol m^{-3}]
	P	$\mathbf{W}_{fg}^{22}$	2.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.7 [m mol m^{-3}]
	H	$\mathbf{W}_{fg}^{33}$	1.5 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.8 [m mol m^{-3}]
Smoothing	Smoothing of N	$\mathbf{W}_{\boldsymbol{\eta\eta}}^{11}$	47,000.0 [(m mol m^{-3} day^{-2})$^{-2}$]	–
	Smoothing of P	$\mathbf{W}_{\boldsymbol{\eta\eta}}^{22}$	49,000.0 [(m mol m^{-3} day^{-2})$^{-2}$]	–
	Smoothing of H	$\mathbf{W}_{\boldsymbol{\eta\eta}}^{33}$	49,000.0 [(m mol m^{-3} day^{-2})$^{-2}$]	–
Measurement	N	$\mathbf{w}^{11}$	1700.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.024 [m mol m^{-3}]
	P	$\mathbf{w}^{22}$	1700.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.024 [m mol m^{-3}]
	H	$\mathbf{w}^{33}$	1900.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.023 [m mol m^{-3}]
Initial	N	$\mathbf{W}_{\mathbf{aa}}^{11}$	1700.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.024 [m mol m^{-3}]
	P	$\mathbf{W}_{\mathbf{aa}}^{22}$	1700.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.024 [m mol m^{-3}]
	H	$\mathbf{W}_{\mathbf{aa}}^{33}$	1900.0 [(m mol m^{-3})$^{-2}$]	$\approx$ 0.023 [m mol m^{-3}]
Dynamical	scalar part	$\rho_{\mathbf{qq}}$	Eq. (20)	–
	matrix part	$\hat{\mathbf{W}}'_{\mathbf{qq}}$	Eq. (20)	–

guess initial condition found from the 3-year model spin up. The dynamical weight $\hat{\mathbf{W}}_{\mathbf{qq}}$ will be explained below. Finally, the smoothing weights were adjusted to get the smoothness consistent with the reference solution.

Choosing the dynamical weight $\hat{\mathbf{W}}_{\mathbf{qq}}$ to be diagonal is probably not correct since the system is coupled. A more consistent weight matrix will be based on the "true" state. First, remember that the dynamical error covariance is defined as

$$\hat{\mathbf{Q}}_{\mathbf{qq}} = \overline{\mathbf{q}\mathbf{q}^T}. \tag{18}$$

Further, since the true state is based on a quite accurate ODE-solver (10^{-6} mmol m^{-3} for all three variables), the main contribution for the discrete dynamical errors q_i^N, q_i^P and q_i^H are due to the approximate finite difference formulas used for the derivatives. Now, the errors in the discrete time i may be found by inserting the true state directly into the discrete forms of Eqs. (2), where the derivatives are approximated by finite difference formulas. That is, the reference solution can be inserted into

$$\mathbf{q}_i = \left(\frac{\mathrm{d}\boldsymbol{\Psi}_i}{\mathrm{d}t}\right) - \boldsymbol{f}_i, \tag{19}$$

where a finite difference formula is used for the derivative and $\boldsymbol{f}_i^T = (f_N,\ f_P,\ f_H)_i$. Evaluating the error terms over a long time interval makes it possible to compute $\hat{\mathbf{Q}}_{\mathbf{qq}}$, and the inverse of this should give a consistent weight matrix $\hat{\mathbf{W}}_{\mathbf{qq}}$. A 5-year run of the model from the annual cycle initial conditions was performed, and the dynamical errors were estimated at every time step of the assimilation interval, i.e. $\Delta t = 0.5d$ (the covariance is clearly dependent on the time step). Finally, the covariance matrix was inverted, and the resulting weight matrix for our model was given by

$$\hat{\mathbf{W}}_{\mathbf{qq}} = \hat{\mathbf{W}}'_{\mathbf{qq}}\,\rho_{\mathbf{qq}} = \begin{pmatrix} 0.3270 & 0.6743 & 1.1246 \\ 0.6743 & 1.5019 & 2.4735 \\ 1.1246 & 2.4735 & 4.0837 \end{pmatrix} * 10^{10}. \tag{20}$$

Mainly the constant $\rho_{\mathbf{qq}}$ outside the matrix was changed if Δt was changed (see Section 5.4). For real experiments, the system noise would probably be much higher, i.e. the dynamical weights would probably be much smaller.

The error terms q^N, q^P and q^H are plotted for the first 2 years in Figs. 1 and 2. It is verified that the errors are large in the neighborhoods of the peak values of the variables (compare with the plots of the state variables, e.g. in Case 1, Section 5.1). Of course, this is also expected since the changes of the derivatives are large in these areas, and since the finite difference formulas therefore probably lead to larger errors.

We have done several (statistical) assumptions for obtaining the final weights. In our twin experiments, where we know the "reference" solution, we could of course evaluate a general $\mathbf{W}_{\mathbf{qq}}(i, j)$. However, it is not practical to store the $\mathscr{O}(N'^2)$ elements required for a general $\mathbf{W}_{\mathbf{qq}}(i, j)$, and it is not realistic to find good estimates of all the cross-covariances for a real application, as mentioned before. A more realistic task is to find the (time) mean of the covariances $\hat{\mathbf{Q}}_{\mathbf{qq}}$. However, it is probably an over-simplification to have a constant model error covariance matrix, and to increase the time regularity, we defined the smoothing norm (Eq. (17)). Again, an accurate time dependent smoothing matrix $\mathbf{W}_{\boldsymbol{\eta\eta}}$ could have been computed in our twin experiments, but in a real experiment it would be more realistic to find an average smoothness of each model variable. To conclude, the statistical hypothesis and the choices of weights are only approximations of the truth. In our twin experiments, we will see that the approximations we have done for obtaining the final weights are sufficient for recreating the "reference" solution to a high degree of accuracy.

For real experiments, a reference solution is not available, and the prior error statistics are therefore not known in advance. Error statistics for the measurements and initial conditions should be available. However, one must also provide estimates of the prior dynamical error covariances. One interesting methodology is to use Monte Carlo simulations. By perturbing N_{ens} model states initially according to a Gaussian distribution, and integrating them forward in time, one can sample any statistical moment from the ensemble. For a nonlinear model, the ensemble of model states will become non-Gaussian during the integration. To include the effect of external error growth, one can force the model with (white in time) pseudo random fields with mean zero and specified covariances (see Evensen (1994) for details). As said

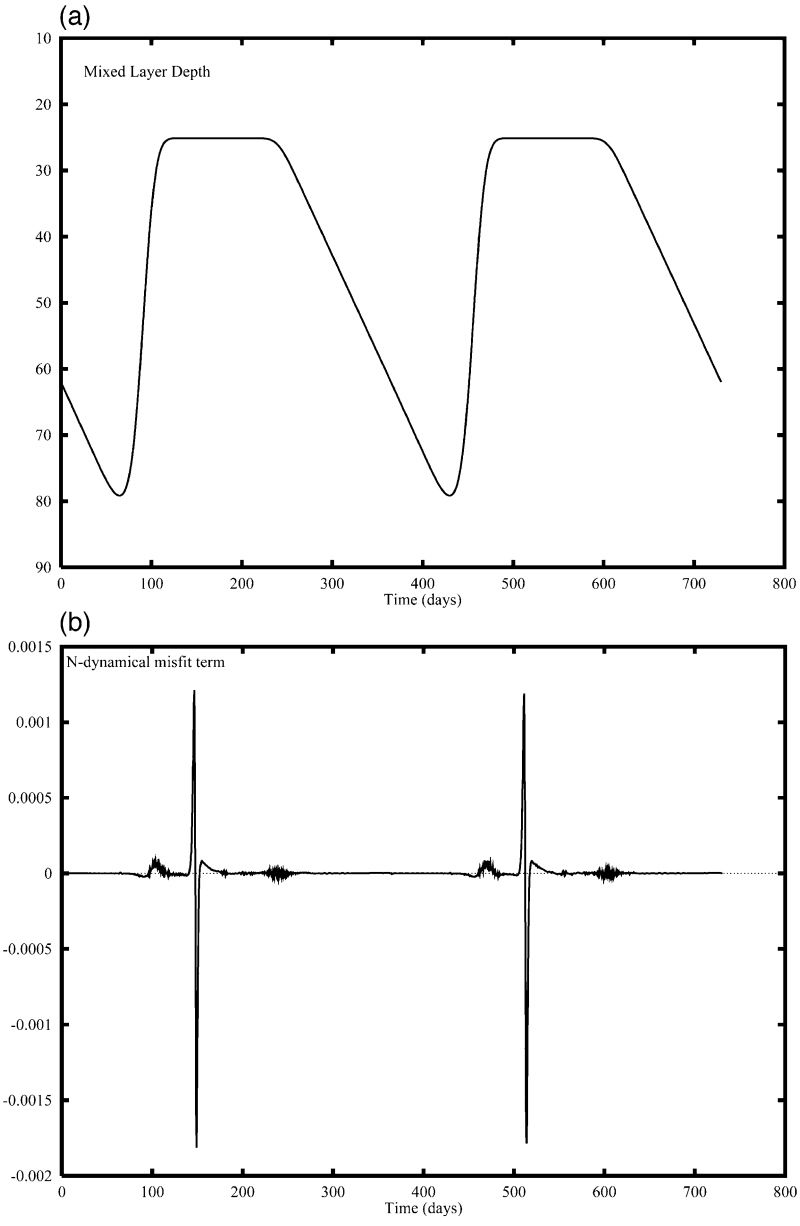

Fig. 1. The mixed layer depth M [m]—smooth version (top) and the evaluated misfit term for the N-equation [mmol m^{-3} day^{-1}] (bottom). Molar units are always in terms of nitrogen, although this is not important to the discussion.

before, by using a constant in time dynamical weight matrix, one does only have to store one 3×3 matrix (an average is taken over the time interval). However, the above procedure will require N_{ens} model integrations. Experiments indicate that 100–150 members are sufficient (Evensen, 1994), and the CPU time required should be moderate at least for zero- and one-dimensional problems. An average smoothness can also be evaluated from the ensemble, or the smoothing weights can be adjusted manually.

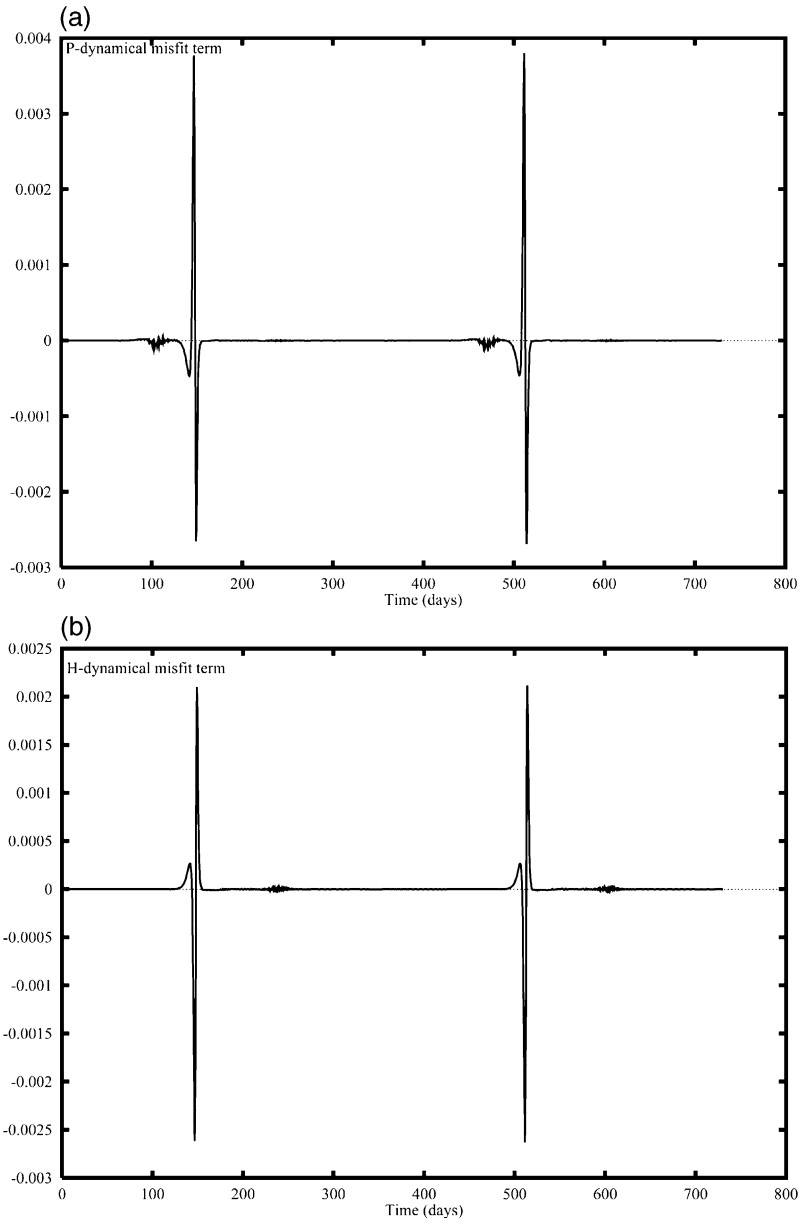

Fig. 2. The evaluated misfit terms for the P-equation [mmol m^{-3} day^{-1}] (top) and for the H-equation [mmol m^{-3} day^{-1}] (bottom).

In the next section, results from several data assimilation experiments will be presented.

5. Assimilation experiments

As emphasized before (Section 3), the gradient descent methods may converge to a local minimum of the penalty function $\mathcal{J}_{N'}$. However, if one starts the algorithms close to a global minimum, there is a good chance of not getting trapped in possible local ones. So, one should try to find a starting evolution that is likely to be close to the "true" solution. An objective analysis (OA) estimate will provide such a starting evolution. This is done by performing a

spline smoothing algorithm, which can be equivalent to objective analysis, as explained in McIntosh (1990) or Bennett (1992). The OA estimate is found by minimizing a cost function $\mathscr{I}_{N'}$, which is defined as

$$\mathscr{I}_{N'}[\boldsymbol{\Psi}] = \sum_{i=1}^{N'} \left(\boldsymbol{\Psi}_i - \overline{\boldsymbol{\Psi}}\right)^T \mathbf{W}_{fg} \left(\boldsymbol{\Psi}_i - \overline{\boldsymbol{\Psi}}\right) + \mathbf{a}^T \mathbf{W}_{\mathbf{aa}} \mathbf{a} + \boldsymbol{\epsilon}^T \mathbf{w} \boldsymbol{\epsilon} + \Delta t \sum_{i=1}^{N'} \boldsymbol{\eta}_i^T \mathbf{W}_{\boldsymbol{\eta}\boldsymbol{\eta}} \boldsymbol{\eta}_i, \tag{21}$$

where $\overline{\boldsymbol{\Psi}}$ is the mean and $\mathbf{W}_{fg}$, which is symmetric and positive definite, is the first guess weight matrix (the values are given in Table 2). Note that the new term $\sum_{i=1}^{N'} (\boldsymbol{\Psi}_i - \overline{\boldsymbol{\Psi}})^T \mathbf{W}_{fg} (\boldsymbol{\Psi}_i - \overline{\boldsymbol{\Psi}})$ is quadratic, and that $\mathscr{I}_{N'}$ is positive definite and quadratic and therefore has a unique global minimum. Minimizing $\mathscr{I}_{N'}$ leads to an estimate which is close to the data, the initial conditions and the first guess or mean. For more information on objective analysis and smoothing splines, see e.g. McIntosh (1990) or Bennett (1992).

In all experiments described in this section, an OA estimate is used as a starting evolution when minimizing the weak constraint cost functional $\mathscr{J}_{N'}$. The different cases are summarized in Table 3.

5.1. Case 1: the basis case

The time interval was chosen to cover 2 years, i.e. $t \in [0,730]$, and a resolution of $0.5d$ was used. Further, the data set was chosen such that the characteristic patterns of the variables were recognized, i.e. all variables were measured at 44 discrete time points, which means $\Delta t_{\text{obs}} \approx 17d$.

Both GSD and FR were used for finding the OA estimate. Fig. 3 shows the evolution of the cost function for the two methods, note that the FR algorithm has much better convergence properties than the GSD algorithm. Therefore, the FR method was preferred in the rest of the assimilation experiments (see also Minoux (1986)).

From Fig. 3 it is seen that the estimates start far from the measurements and the initial conditions (the routine was started from constant values of N, P and H). Further, the figure suggests that the OA estimates are closer to the measurements and the initial conditions than the starting state, but further from the first guess (we used the mean as a first guess, see Eq. (21)). The wave pattern that can be seen in the cost function, especially for the initial part, is caused by the 400th iteration restarts in the FR algorithm.

The weak constraint cost function and the assimilation results are shown in Figs. 4 and 5, and the results are found to be quite satisfactory. The weak constraint estimate is very close to the reference solution, and the information from the dynamics is used to reproduce the peaks in the spring bloom of phytoplankton and zooplankton. Further, it is seen that the largest errors are at locations where the second derivatives are large, and this may be expected since the errors of using the finite difference formulas are larger in these areas. The FR method was able to overcome the strong nonlinearities in the dynamics. When starting the algorithm from the reference solution instead of from the OA estimates, it converged to the same value of the cost as found before, and thus the global minimum had been found.

The weak constraint estimates are further from the reference solution close to $t = 0$ than in the rest of the time interval. This is, however, not regarded as a major weakness, since the reduced accuracy only affect the boundary domain of the assimilation interval. This can be more easily seen in Case 4 (and also in Case 2A), see Fig. 6 or Fig. 9.

Table 3
Description of the cases

Case	Measured	Δt_{obs}	T'	Resolution (Δt)
1	N, P, H	$17d$	$730d$	$0.5d$
2A	N, P, H	$20d$	$730d$	$0.5d$
2B	N, P, H	$50d$	$730d$	$0.5d$
3	P	$5d$	$730d$	$0.5d$
4	N, P, H	$20d$	$2190d$	$0.7d$

5.2. Case 2: reducing the number of measurements

In this case, the sensitivity with respect to the data density is investigated. It must be expected that the dynamics will dominate for a poor data set, and convergence to a local minimum of the cost may be the result.

Instead of investigating the sensitivity with respect to data density, we could of course have fo-

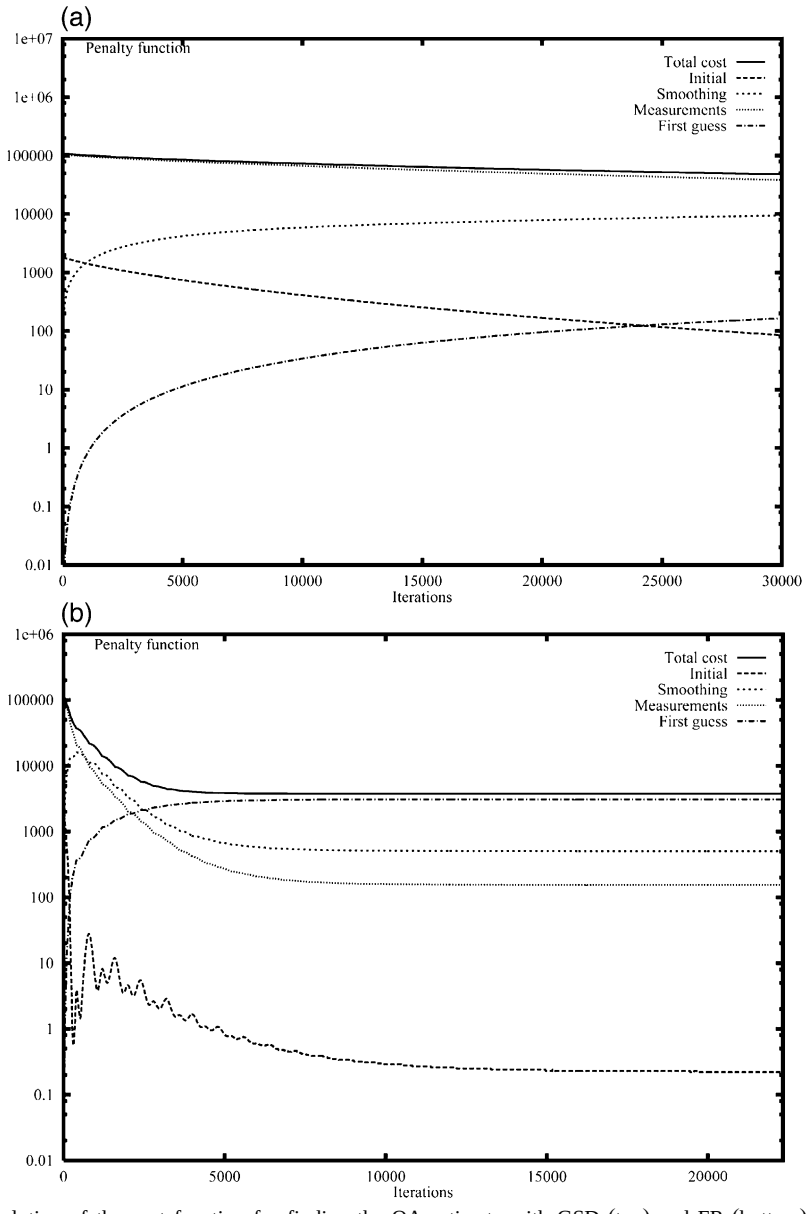

Fig. 3. Case 1; The evolution of the cost function for finding the OA estimate with GSD (top) and FR (bottom). Notice the very slow convergence of GSD.

cused more on the sensitivity with respect to the amount of noise added to the data. Both approaches will eventually cause the dynamics to dominate, and thus have an equivalent effect on the gradient descent routine. We chose the former approach for convenience.

5.2.1. Case 2A

First, the measurement density was reduced by about 20%, i.e. $\Delta t_{\text{obs}} \approx 20\,d$. Apart from this, the case is similar to Case 1. Of course, whenever the measurement density is changed, a new OA estimate has to be found. Further, this estimate must be

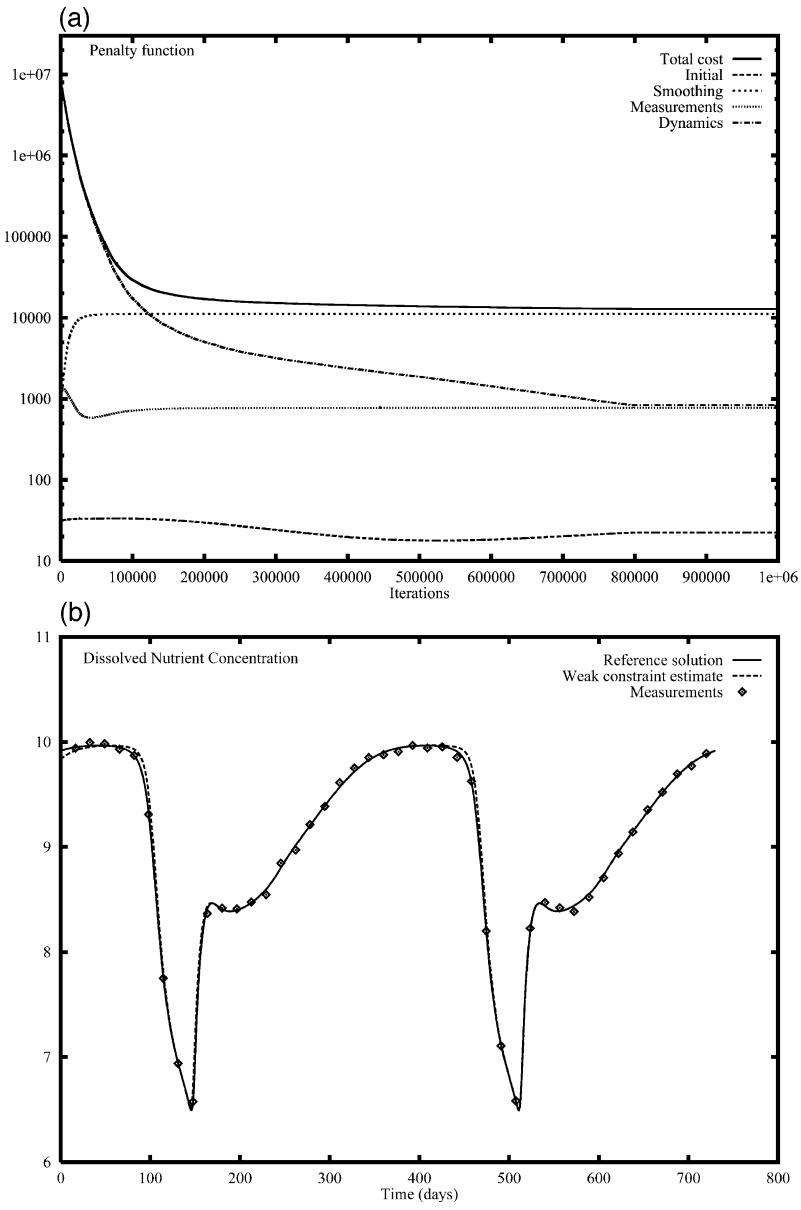

Fig. 4. Case 1; The evolution of the weak constraint cost function (top) and the weak constraint estimate for nutrients N [mmol m^{-3}] (bottom).

expected to be poorer because the density has been reduced.

The results were, in fact, quite good also in this case, and convergence was obtained from the OA estimate. The final value of the cost function was compared to the corresponding value after some iterations when using the reference solution as the starting state. These values were found equal, i.e. the algorithm converged to the global minimum. The inverse estimates of phytoplankton and zooplankton

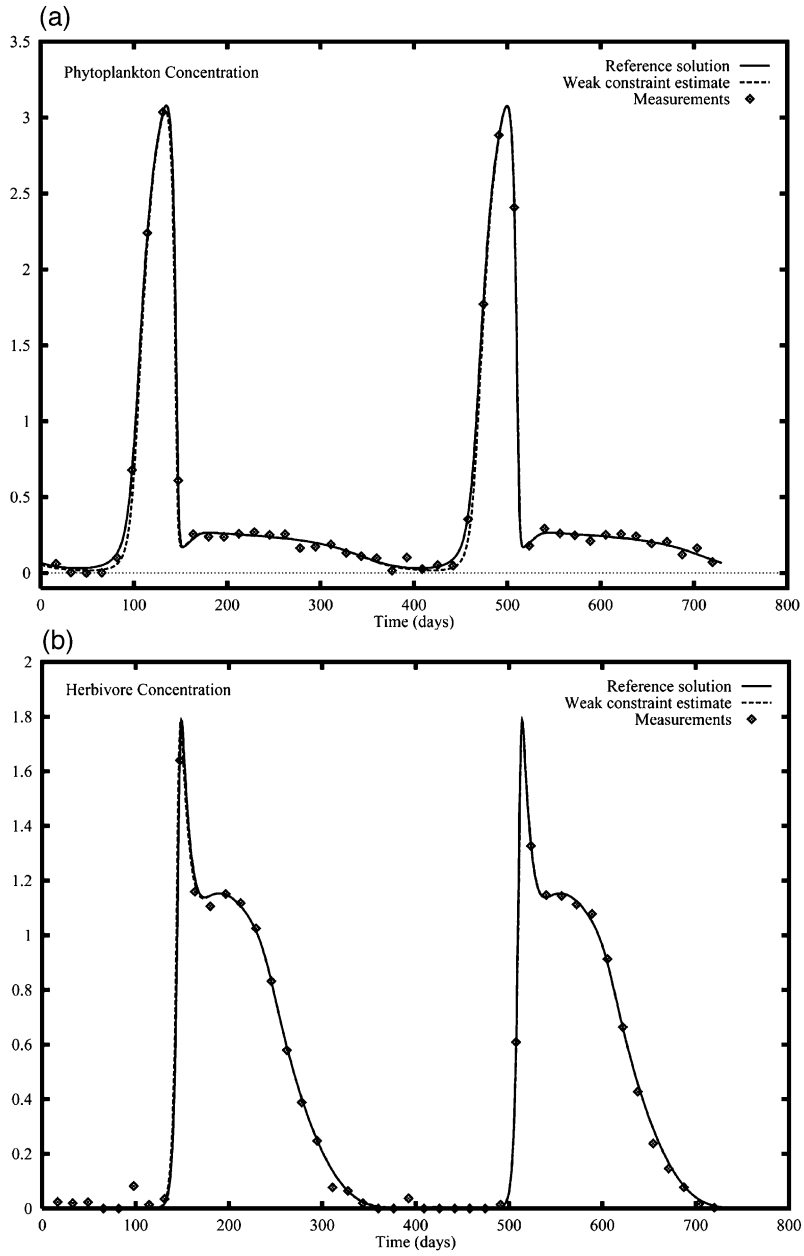

Fig. 5. Case 1; The weak constraint estimate for phytoplankton P [mmol m^{-3}] (top) and herbivores H [mmol m^{-3}] (bottom).

are given in Fig. 6. The convergence was a little slower than in Case 1 (10–20%), and the weak constraint estimates were a little poorer, especially around the peaks (see Fig. 6). Still, the results are found to be quite satisfactory; the weak constraint estimate is very close to the reference solution.

Also when we chose $\Delta t_{\text{obs}} \approx 35d$, satisfactory results were found.

5.2.2. Case 2B

Now, the measurement density was reduced to $\Delta t_{\text{obs}} \approx 50\,d$. The weak constraint results found in

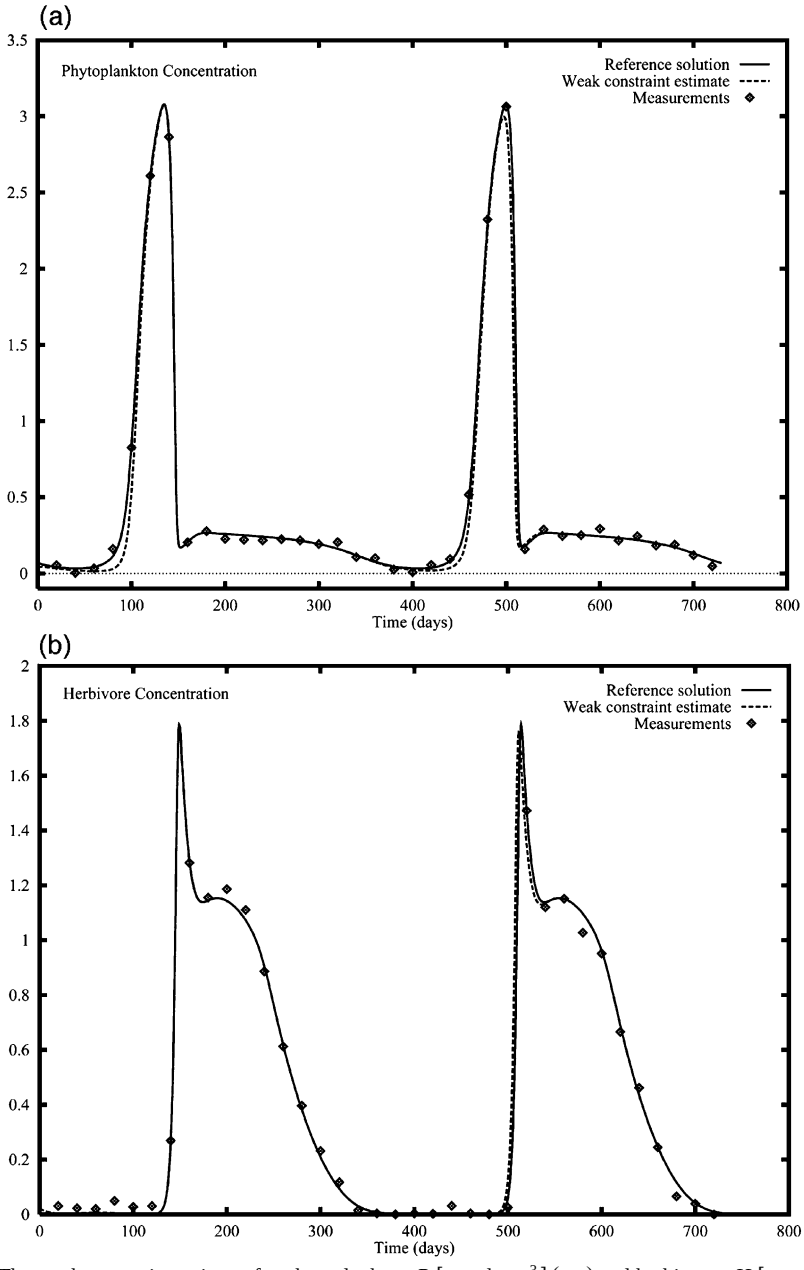

Fig. 6. Case 2A; The weak constraint estimate for phytoplankton P [mmol m^{-3}] (top) and herbivores H [mmol m^{-3}] (bottom).

this case are much poorer than the previous results, phytoplankton and zooplankton are plotted in Fig. 7.

Starting FR from the reference solution and from the OA estimates gave in this case different values of the cost function. Thus, the FR algorithm had only discovered a local minimum of the cost function when starting from the OA estimate. By the nature of the FR search, it must be expected that this method will not be able to find the global minimum when the dynamical term dominate in the penalty function, as mentioned before (Section 3). A too poor data set will clearly allow the nonlinear dynamics to domi-

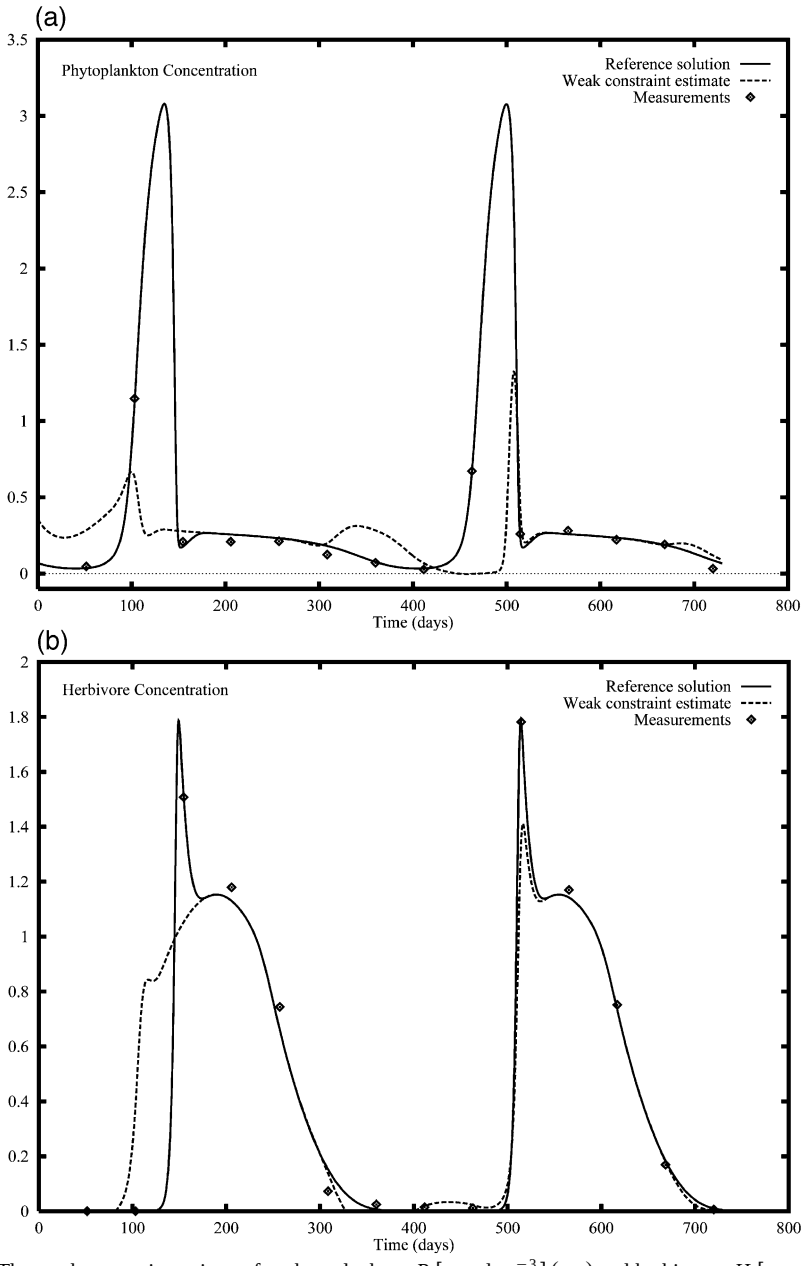

Fig. 7. Case 2B; The weak constraint estimate for phytoplankton P [mmol m^{-3}] (top) and herbivores H [mmol m^{-3}] (bottom).

nate and convergence to a local minimum may be the result.

It is worth mentioning that the OA estimate was much poorer in this case because of the lack of data points in the peaks of the reference solution, especially for phytoplankton. This might have caused the weak constraint cost function to converge to a local instead of to the global minimum.

5.3. Case 3: measuring only P

In this case, only one ecosystem variable, P, was measured. As stated before (Abstract), there is a

large amount of ocean colour data available from earth observation satellites, e.g. SeaWiFS and MOS. These observations give a measure of the colour of the ocean, from which we can deduce chlorophyll concentrations (O'Reilly et al., 1998). This may be used to provide measurements of phytoplankton, see e.g. Semovski and Woźniak (1995) for information on relationships between chlorophyll *a*, carbon and nitrogen. Because measurements of zooplankton and nutrients may be difficult to provide, it is interesting

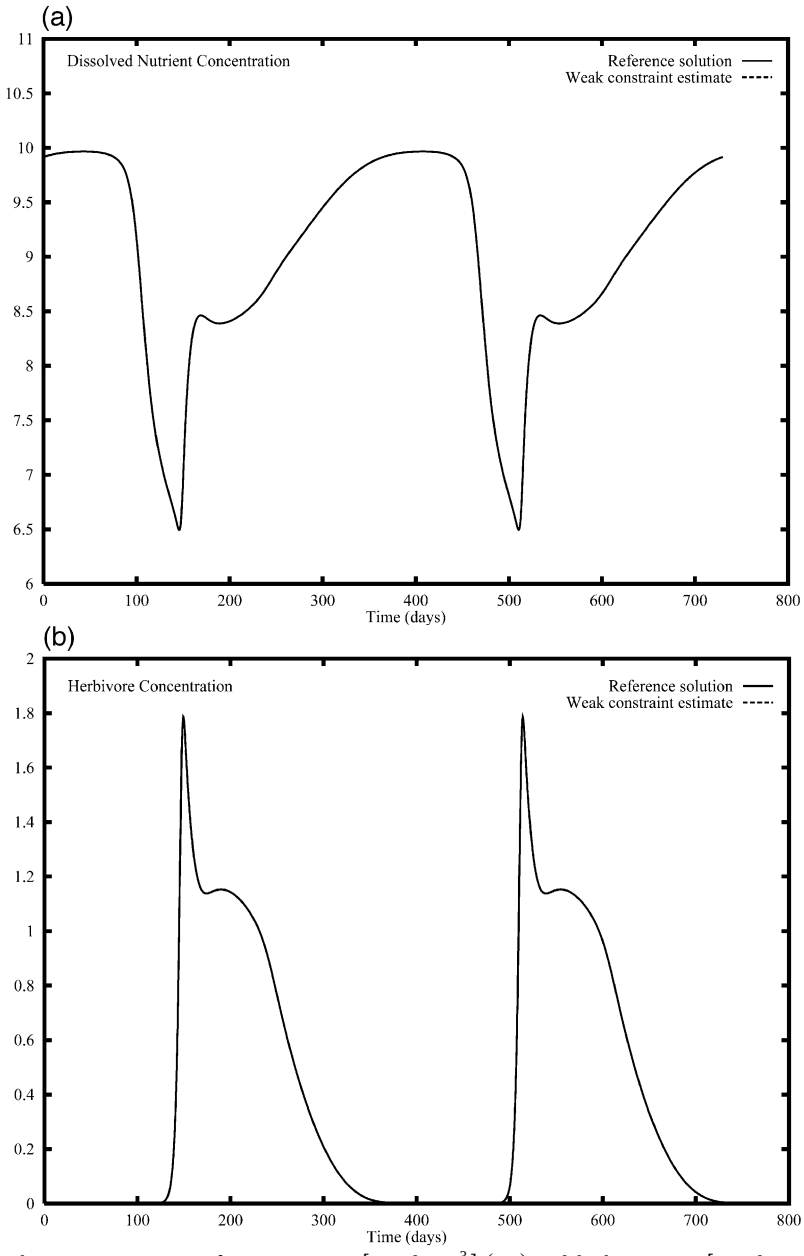

Fig. 8. Case 3; The weak constraint estimate for nutrients N [mmol m^{-3}] (top) and herbivores H [mmol m^{-3}] (bottom). The weak constraint estimate is hidden behind the reference solution.

to investigate an assimilation for which only the phytoplankton concentration is measured.

Note that the OA estimates cannot be found for the N and H variables, since these are not measured. The algorithm was started from the OA estimate of P and from constant values (mean of the reference solution) of N and H.

When the measurement density and the measurement error variance were chosen as in Case 1, the FR algorithm converged to a local minimum. Therefore, we chose $\Delta t_{\text{obs}} \approx 5d$. Also, we had to reduce the measurement error variance by almost two orders of magnitude to obtain convergence to the global minimum of the cost function $\mathcal{J}$.

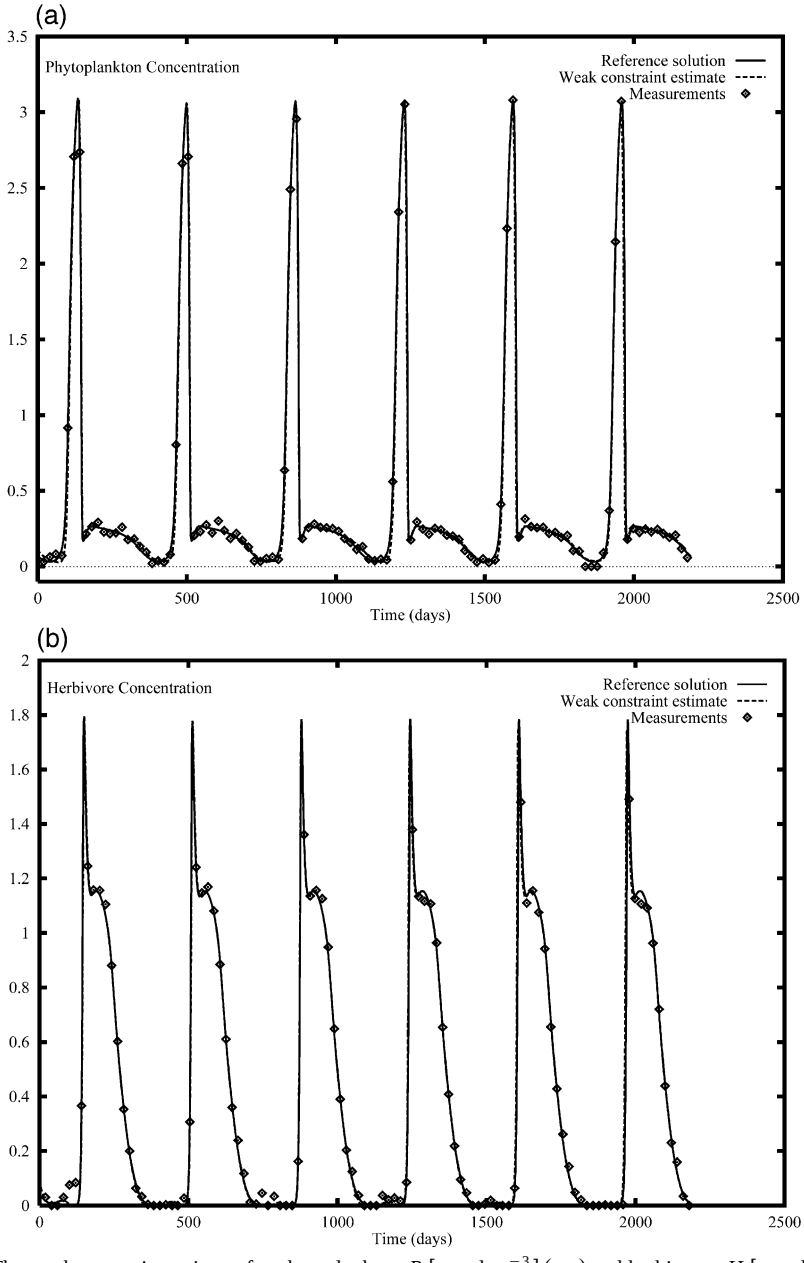

Fig. 9. Case 4; The weak constraint estimate for phytoplankton P [mmol m^{-3}] (top) and herbivores H [mmol m^{-3}] (bottom).

The results for nutrients and zooplankton are shown in Fig. 8. The results of this case are highly satisfactory and it is clearly verified that as long as the measurement density is good and the measurement error variance is small for the phytoplankton component, accurate estimates may be provided also for the others.

Our assumptions about the measurement errors and the measurement density are probably not the same as for some real data set, e.g. the measurement errors we have used are probably smaller than the errors of real ocean colour measurements. As said before (Section 5.2), larger errors could cause a gradient descent routine to converge to a local minimum of the cost function. However, if the real data set at the same time is denser than ours, the routine may still be able to locate the global minimum.

5.4. Case 4: increasing the time interval

In this case, it was investigated how sensitive the inverse estimate was with respect to a larger time interval; it was now increased to cover 6 years, i.e. $t \in [0,2190]$. A data density of $\Delta t_{\text{obs}} \approx 20\,d$ and a resolution of $\Delta t = 0.7d$ was chosen. Remember that the dynamical weight matrix being dependent on the time step had to be re-evaluated in this case. As said before (Section 4), mainly the constant part $\rho_{\mathbf{qq}}$ changed for a new resolution.

Again, FR converged to the global minimum. The results for phytoplankton and zooplankton are given in Fig. 9. It is observed that the results are good also in this case, and this shows the main benefit of using a weak constraint formulation. The strong sensitivity with respect to the initial conditions which is seen in a strong constraint approximation is removed. Miller et al. (1994) used the so-called adjoint method, which solves a variational problem using the model as a strong constraint, with the nonlinear Lorenz model. They proved that the cost function was strongly dependent on even small perturbations in the initial conditions. Further, when increasing the interval over which the minimization was carried out, they found an increasing number of minima in the cost function with respect to the same perturbations in the initial conditions. Thus, the strong constraint approach showed a strong dependence on the initial conditions, and this dependence became increasingly stronger for larger assimilation intervals. This is the main reason for choosing the weak constraint approach here. The variables are by including the dynamics as a weak constraint, allowed to recover from disturbances, and this makes it possible to obtain good results for large intervals. One may say that the dynamics only include information from points not too far away, i.e. a local behavior is seen (Evensen and Fario, 1997).

It has been seen that he quality of the weak constrain estimate was dependent on the data density. However, by studying the figures in the current case, it is believed that the data density could be poorer as long as a reasonably good density is ensured in the intervals of rapid change, i.e. in the spring and early summer.

6. Discussion

The following conclusions are based on the twin experiments presented in Section 5.

The Fletcher and Reeves method (FR) converged in the data assimilation experiments with our simplified ecosystem model. However, it was seen that an ordinary gradient steepest descent search (GSD) converged far too slow to be practical, even for finding the OA estimate in Case 1. The fact that the convergence of GSD probably will be even slower for a one-dimensional model means that this method is inappropriate for future developments.

Our twin experiments demonstrated that satisfactory results may be expected as long as the measurement density is reasonably good. If an assimilation domain is well covered by measurements of reasonably good quality, the penalty function is essentially quadratic (and convex), and the global minimum of the cost function can be reached. To be more specific, the penalty function can only be expected to be essentially quadratic in a domain around the global minimum, so this minimum can be captured only if we start the FR search not too far away from it (in our experiments, an objective analysis estimate was used as a starting state). However, whenever there are few measurements or the measurements contain large errors, the dynamics will dominate and there will be a large possibility of reaching a local minimum or saddle point only. This could have been

solved by using statistical methods, which have a higher probability of tracking the global minimum, see e.g. Kirkpatrick et al. (1983). However, they are much more expensive; often too expensive to be practical (Evensen and Fario, 1997). All data assimilation techniques will of course have problems tracking the true state when a too poor data set is used.

A quite satisfactory result is that the weak constraint estimate did not lose any quality when the time interval was tripled. This is due to the local behavior of the weak constraint formulation, that is, the estimate is only dependent on points not too far away. Thus, a weak constraint estimate is not as crucially dependent on the initial conditions as a strong constraint approximation is, see Miller et al. (1994) and Evensen and Fario (1997).

An inverse estimate close to the reference solution was also obtained when only one of the ecosystem components, phytoplankton, was measured and used (see Case 3). Future marine observation systems will probably mainly consist of remotely sensed ocean colour data. These observations may be used to measure the water content of chlorophyll, which again may be used to deduce phytoplankton concentrations. Thus, for this particular model, it seems possible to obtain satisfactory estimates also for nutrients and zooplankton, by assimilating measurements of the phytoplankton component only. This result is believed to be of major interest when trying to develop future data assimilation systems for the marine environment in the ocean. This example demonstrates the importance of the coupling between the model variables when assimilating measurements of one type. In other models, some variables may be more weakly coupled, and the data assimilation system may only be able to control some ecosystem compartments. One could in this case improve the assimilation system by utilizing measurements of other types. However, for the strongly coupled model variables, measurements of one type should be sufficient.

The reader may have observed that the gradient descent methods do not provide any estimates of the error statistics. However, this could be provided by using a statistical sampling method (Evensen, 1997).

It is believed that FR also will perform well for a one-dimensional ecosystem model. The storage and CPU requirements are moderate also in this case (see Table 1) although one can expect a worsening in the conditioning of the problem and thus more iterations are required. However, a full three-dimensional application may give considerable problems with computer capacity. The state variable becomes extremely large in this case, and the convergence will probably require too much CPU-time. There is, however, a possibility of solving the problem for small "local" time intervals, and use the final estimates from one interval as initial conditions for the next. However, for operational three-dimensional systems, other techniques should probably be used, e.g. ensemble Kalman filter (Evensen, 1994) or representer techniques (Bennett et al., 1997).

Several of the expressions used in marine ecosystem models may be referred to as semi-empirical, which means that a given curve form is assumed, and certain parameters have to be estimated. It is worth mentioning that it is possible to formulate a combined inverse and parameter estimation problem. For example, assume that the parameter for the maximum grazing rate of zooplankton varies throughout the year, i.e. $c = c(t)$. Further, by defining c as $c(t) = c_0(t) + \epsilon_c(t)$, where c_0 is a first guess and ϵ_c is an unknown error, we may include the term $\int_0^{T'}\int_0^{T'} \epsilon_c(t_1) W_{cc}(t_1, t_2) \epsilon_c(t_2) \mathrm{d}t_1 \mathrm{d}t_2$ in the penalty function. Parameter estimation experiments are, however, not performed here. In Eknes and Evensen (1997), such a combined inverse and parameter estimation formulation was used to estimate the wind drag and the diffusion coefficient in a one-dimensional Ekman flow model. Evensen et al. (1998) give a general discussion of parameter estimation problems in dynamical models.

Acknowledgements

This work has been supported by the EC-MAST-III BASYS project (MAS3-CT96-0058), and also by the research Council of Norway (Programme for Supercomputing) through a grant of computing time.

Appendix A. The model

This section contains a brief description of the dynamical model, which is a model by Evans and Parslow (1985).

The model has already been referred to as zero-dimensional. This means that we assume that the ocean can be divided into two layers; one upper and one lower. In each of these layers all ecosystem components are completely mixed, so their concentrations are constant within each layer and vertical variations are averaged.

Further, our model contains one physical and three ecosystem components. The mixed layer depth M is the variable for the thickness of the upper layer, which is the layer we are interested in. Plankton types included are phytoplankton and herbivorous zooplankton, whose concentrations are denoted by P and H. The last variable N holds the concentration of dissolved nutrients, such as nitrogen, silicate and phosphate. The upper layer is the biologically active layer containing both nutrients and plankton, while the lower one is inactive and contains no plankton.

The rate of change of the mixed layer depth is prescribed by a function $\zeta(t)$. As the mixed layer deepens, water from the lower layer is mixed into the upper, and nutrients from the two layers are mixed, and phytoplankton and zooplankton in the upper layer experiences a dilution. As the upper layer shallows, no mixing of water takes place in this layer, and nutrient and phytoplankton concentrations do not change while herbivores will be concentrated because they actively "follow" the shallowing of the layer. This N, P asymmetrical behavior is captured by defining $\zeta^+(t) = \max(\zeta(t), 0)$. Diffusion of phytoplankton and nutrients across the mixed layer is modelled by a constant m'.

The coupled system of model equations for M, N, P and H is:

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \zeta(t), \tag{22}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\left(\frac{\alpha(t,M,P)\,N}{j' + N} - r\right)P + \frac{m' + \zeta^+(t)}{M}(N_0' - N), \tag{23}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \left(\frac{\alpha(t,M,P)\,N}{j' + N} - r\right)P - \frac{c(P - P_0')\,H}{K + P - P_0'} - \frac{m' + \zeta^+(t)}{M}P, \tag{24}$$

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{fc(P - P_0')\,H}{K + P - P_0'} - gH - \frac{\zeta(t)}{M}H, \tag{25}$$

where $\alpha(t, M, P)$ is the photosynthetic rate of phytoplankton described in Evans and Parslow (1985). (In the calculation of α, light intensity is included as a physical forcing). The parameters are appropriate to Flemish Cap (east of Newfoundland), they are also given in Evans and Parslow (1985). (Short definitions are given in Appendix C).

The variable M above is representing a physical input to the ecosystem variables N, P and H. Normally, the physical part would by itself be a coupled set of differential equations. Further, a data assimilation system should be created for this physical part, and the resulting variables should be used as input for the biochemical equations. Remember that $\zeta(t)$ is the prescribed function for the deepening or shallowing of the mixed layer depth M. This function is only specified to get a simple model describing the annual cycle of the ecosystem components N, P and H.

Evans and Parslow (1985) chose ζ as a simple step function of time, and the mixed layer depth M was found by a forward integration of $\mathrm{d}M/\mathrm{d}t = \zeta(t)$. We used a shapiro filter to find a smooth version of $M = M(t)$, the result is plotted in Fig. 1. Further, we used second order finite difference formulas to estimate ζ.

Thus, only the three ecosystem components will be assimilated while the mixed layer depth M is regarded as an input parameter dependent of time. By defining $\boldsymbol{\Psi}^T(t) = [N(t), P(t), H(t)]$, the model Eqs. (23)–(25) may be written as

$$\frac{\mathrm{d}\boldsymbol{\Psi}}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{\Psi}, t), \tag{26}$$

where $\boldsymbol{f}^T = [f_N, f_P, f_H]$ and the vector components represent the right hand sides of each of Eqs. (23)–(25).

Appendix B. The gradient

This appendix contains some comments on the calculation of the gradient, which was needed for the search direction $\mathbf{s}_l$ in the gradient descent routines described earlier. The details will be left for the

reader. Only useful hints, definitions and some prior calculations are given here. As explained in Section 3, the gradient with respect to the discrete N, P and H is needed. Thus, we are now using discrete equations.

First, define an initial part, a model or dynamical part, a measurement part and a smooth field constraint part of the discretized cost function:

$$\mathscr{J}^I_{N'} = \mathbf{a}^T \mathbf{W}_{\mathbf{aa}} \mathbf{a}, \quad \mathscr{J}^D_{N'} = \Delta t^2 \sum_{i=1}^{N'} \sum_{j=1}^{N'} \mathbf{q}_i^T \mathbf{W}_{\mathbf{qq}}(i,j) \mathbf{q}_j, \tag{27}$$

$$\mathscr{J}^M_{N'} = \boldsymbol{\epsilon}^T \mathbf{w} \boldsymbol{\epsilon}, \quad \mathscr{J}^S_{N'} = \Delta t \sum_{i=1}^{N'} \boldsymbol{\eta}_i^T \mathbf{W}_{\boldsymbol{\eta}\boldsymbol{\eta}} \boldsymbol{\eta}_i. \tag{28}$$

Note that the discrete version of the measurement operator $\mathscr{L}$ is a matrix $\mathbf{L}$, which is multiplied with the full discretized state vector $\boldsymbol{\psi}$. Now, the derivatives with respect to the full state are evaluated for each of the above parts.

B.1. The $\nabla \mathscr{J}^I_{N'}$ part of the gradient

By taking the weight matrix $\mathbf{W}_{\mathbf{aa}}$ as symmetric, one gets for the derivative of $\mathscr{J}^I_{N'}$ with respect to the concentration of phytoplankton at the discrete point k (time), P_k:

$$\frac{\partial \mathscr{J}^I_{N'}}{\partial P_k} = 2 \frac{\partial \mathbf{a}^T}{\partial P_k} \mathbf{W}_{\mathbf{aa}} \mathbf{a} = \begin{cases} 2\mathbf{e}_2^T \mathbf{W}_{\mathbf{aa}} \mathbf{a} & \text{if } k = 1 \\ 0 & \text{if } k > 1. \end{cases} \tag{29}$$

The derivatives with respect to N_k and H_k are similar, but with $\mathbf{e}_2$ interchanged by $\mathbf{e}_1$ and $\mathbf{e}_3$, respectively ($\mathbf{e}_1^T = (1, 0, 0)$, $\mathbf{e}_2^T = (0, 1, 0)$, $\mathbf{e}_3^T = (0, 0, 1)$).

B.2. The $\nabla \mathscr{J}^M_{N'}$ part of the gradient

Again, by using that $\mathbf{w}$ is symmetric, one gets for the derivative with respect to P_k:

$$\frac{\partial \mathscr{J}^M_{N'}}{\partial P_k} = -2 \left(\mathbf{L} \frac{\partial \boldsymbol{\psi}}{\partial P_k} \right)^T \mathbf{w}(\mathbf{d} - \mathbf{L}\boldsymbol{\psi}) = -2 \left(\mathbf{L} \boldsymbol{\delta}_{\boldsymbol{\psi}, P_k} \right)^T \mathbf{w}(\mathbf{d} - \mathbf{L}\boldsymbol{\psi}), \tag{30}$$

where the delta function represents a state vector in which the kth element of P is one, and all other elements are zero. To be more specific, this delta function chooses the kth element of the P part of the full state vector $\boldsymbol{\psi}$ to be one, and the rest to be zero. The other derivatives are similar, but now with the delta functions $\boldsymbol{\delta}_{\boldsymbol{\psi}, N_k}$ and $\boldsymbol{\delta}_{\boldsymbol{\psi}, H_k}$.

B.3. The $\nabla \mathscr{J}^D_{N'}$ part of the gradient

We use that $\mathbf{W}_{\mathbf{qq}}$ is symmetric:

$$\frac{\partial \mathscr{J}^D_{N'}}{\partial \psi_k} = 2\Delta t^2 \sum_{i=1}^{N'} \sum_{j=1}^{N'} \frac{\partial \mathbf{q}_i^T}{\partial \psi_k} \mathbf{W}_{\mathbf{qq}}(i,j) \mathbf{q}_j = 2\Delta t^2 \sum_{i=1}^{N'} \frac{\partial \mathbf{q}_i^T}{\partial \psi_k} \sum_{j=1}^{N'} \mathbf{W}_{\mathbf{qq}}(i,j) \mathbf{q}_j, \tag{31}$$

Now, remember that $\mathbf{q}_n = ((\mathrm{d}\boldsymbol{\Psi}_n)/(\mathrm{d}t)) - \boldsymbol{f}_n$, where $\boldsymbol{\Psi}_n^T = (N, P, H)_n$ and $\boldsymbol{f}_n^T = (f_N, f_P, f_H)_n$, where f_N, f_P and f_H are the right hand sides of the model Eqs. (23)–(25). Now, the differentiation of $\mathbf{q}_i$ may be split up as follows:

$$\frac{\partial \mathbf{q}_i^T}{\partial \psi_k} = \frac{\partial}{\partial \psi_k} \left(\frac{\mathrm{d}\boldsymbol{\Psi}_i}{\mathrm{d}t} \right)^T - \frac{\partial}{\partial \psi_k} (\boldsymbol{f}_i)^T. \tag{32}$$

For models with equations of the form $\mathrm{d}x/\mathrm{d}t = f(x)$, we see that the first part will always be the same, while the second is different for different models.

For the first part of Eq. (32), remember that $\mathrm{d}\boldsymbol{\Psi}_i/\mathrm{d}t$ is approximated by a finite difference formula. Therefore, taking the derivative with respect to a variable in a particular discrete point will lead to Dirac delta functions with bases at the differentiation point. For example, for $i \in [2, N' - 1]$,

$$\frac{\partial}{\partial P_k} \left(\frac{\mathrm{d}P}{\mathrm{d}t} \right) \approx \frac{\partial}{\partial P_k} \left(\frac{P_{i+1} - P_{i-1}}{2\Delta t} \right) = \frac{\delta_{i+1,k} - \delta_{i-1,k}}{2\Delta t}. \tag{33}$$

Note that one sided differences must be used at the end points $i = 1$ and $i = N'$.

The second part of Eq. (32) involves differentiation of the right hand sides of the model equations. The details are left for the reader.

Note also that in the expressions above, a general weight function $\mathbf{W_{qq}}(i, j)$ has been used. When the time dependency in the weight is removed, as done here, i.e. $\mathbf{W_{qq}}(i, j) = \hat{\mathbf{W}}_{\mathbf{qq}}\delta(i-j)$ (see Section 4), all the expressions containing the dynamical weight are simplified in the way that one summation is removed and also a Δt due to the Dirac delta function. For example, the dynamical part of the cost function (27) is transformed to

$$\mathscr{J}^D_{N'} = \Delta t \sum_{i=1}^{N'} \mathbf{q}_i^T \hat{\mathbf{W}}_{\mathbf{qq}} \mathbf{q}_i. \tag{34}$$

B.4. The $\nabla \mathscr{J}^S_{N'}$ part of the gradient

Symmetry of $\mathbf{W}_{\boldsymbol{\eta\eta}}$ is used, and we get for the derivative with respect to P_k:

$$\begin{aligned}\frac{\partial \mathscr{J}^S_{N'}}{\partial P_k} &= 2\Delta t \sum_{i=1}^{N'} \frac{\partial \boldsymbol{\eta}_i^T}{\partial P_k} \mathbf{W}_{\boldsymbol{\eta\eta}} \boldsymbol{\eta}_i \\ &= 2\Delta t \sum_{i=1}^{N'} \frac{\partial \eta_i^P}{\partial P_k} \mathbf{e}_2^T \mathbf{W}_{\boldsymbol{\eta\eta}} \boldsymbol{\eta}_i.\end{aligned} \tag{35}$$

Finite difference approximations are used for the second derivatives, and again one gets Dirac delta functions with bases at point k. The details are left for the reader.

Appendix C. List of symbols

M_{obs} Number of observations.
M Mixed layer depth.
N Concentrations of nutrients.
P Concentrations of phytoplankton.
H Concentrations of herbivores.
$\boldsymbol{\Psi}$ State vector containing N, P and H.
q^N, q^P, q^H Unknown dynamical misfit terms for the model equations.
$\mathbf{q}$ Vector containing the unknown dynamical misfit terms q^N, q^P and q^H.
f^N, f^P, f^H The right hand sides of the model equations.
$\boldsymbol{f}$ Vector containing the right hand sides of the model equations.
N_0, P_0, H_0 First guess initial conditions.
$\boldsymbol{\Psi}_0$ Vector containing N_0, P_0 and H_0.
a^N, a^P, a^H Unknown error terms for the initial conditions.
$\mathbf{a}$ Vector containing a^N, a^P and a^H.
$\mathbf{d}$ Vector of measurements.
$\mathscr{L}$ Vector of linear measurement functionals.
$\boldsymbol{\epsilon}$ Vector of unknown measurement errors.
$\mathbf{Q_{qq}}(t_1,t_2)$ Model error covariance matrix between t_1 and t_2.
$\mathbf{Q_{aa}}$ Error covariance matrix for the initial conditions.
$\mathbf{w}^{-1}$ Error covariance matrix for the measurements.
$\mathbf{W_{qq}}(t_1,t_2)$ Dynamical weight matrix, functional inverse of $\mathbf{Q_{qq}}(t_1,t_2)$.
$\mathbf{W_{aa}}$ Initial weight matrix, inverse matrix of $\mathbf{Q_{aa}}$.
$\mathbf{w}$ Measurement weight matrix, inverse matrix of $\mathbf{w}^{-1}$.
$\mathscr{J}[\boldsymbol{\Psi}]$ Weak constraint cost function (penalty function).
$\boldsymbol{\delta}(t_1-t_3)$ Identity matrix whenever $t_1 = t_3$, zero matrix whenever $t_1 \neq t_3$.
$p(\mathscr{J}[\boldsymbol{\Psi}])$ Probability function for $\mathscr{J}[\boldsymbol{\Psi}]$.
$A_{\mathscr{J}}$ Normalizing constant for the probability function $p(\mathscr{J}[\boldsymbol{\Psi}])$.
$\boldsymbol{\psi}$ Full state containing N, P and H in every discrete point.
l Iteration number of the gradient descent routines.
$\mathbf{s}_l$ Search direction of the lth iteration of a gradient descent routine.
λ_l Step length of the lth iteration of a gradient descent routine.
N' Number of discrete points (time).
$\mathscr{J}_{N'}$ Discrete version of $\mathscr{J}$.
κ_1, κ_2 Tolerance parameters in the Goldstein criteria.
β_l A weight parameter in the conjugate gradient routine.
m Number of model equations.
$\delta(i-j)$ Dirac's delta function.
$\mathscr{J}^S_{N'}[\boldsymbol{\psi}]$ The smooth field constraint term of the discrete cost function.
$\mathscr{J}^D_{N'}[\boldsymbol{\psi}]$ The dynamical term of the discrete cost function.
$\mathscr{J}^M_{N'}[\boldsymbol{\psi}]$ The measurement term of the discrete cost function.
$\mathscr{J}^I_{N'}[\boldsymbol{\psi}]$ The initial term of the discrete cost function.

$\hat{\mathbf{W}}_{\mathbf{qq}}$ Constant dynamical weight matrix.
$\hat{\mathbf{Q}}_{\mathbf{qq}}$ Constant dynamical error covariance matrix.
$\hat{\mathbf{W}}'_{\mathbf{qq}}$ Part of $\hat{\mathbf{W}}_{\mathbf{qq}}$.
$\rho_{\mathbf{qq}}$ Part of $\hat{\mathbf{W}}_{\mathbf{qq}}$.
η^N, η^P, η^H Second order derivatives of the model variables.
$\boldsymbol{\eta}$ Vector which contains η^N, η^P and η^H.
$\mathbf{W}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ Matrix containing the smoothing weights.
$\mathscr{K}_{N'}$ Spline smoothing cost function.
$\mathbf{W}_{fg}$ First guess weight matrix.
Δt Time between discrete points.
Δt_{obs} Time between observations.
$\zeta(t)$ The prescribed change of the mixed layer depth M.
$\zeta^+(t)$ $\zeta^+(t) = \max(\zeta(t), 0)$.
$\alpha(t, M, P)$ Photosynthetic rate of phytoplankton.
m' Diffusion parameter.
$\mathbf{L}$ Discrete measurement operator (matrix).
$\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ $\mathbf{e}_1^T = (1, 0, 0)$, $\mathbf{e}_2^T = (0, 1, 0)$, $\mathbf{e}_3^T = (0, 0, 1)$.
$\boldsymbol{\delta}_{\boldsymbol{\psi}, N_k}$, $\boldsymbol{\delta}_{\boldsymbol{\psi}, P_k}$, $\boldsymbol{\delta}_{\boldsymbol{\psi}, H_k}$ Delta functions which choose one of N_k, P_k or H_k of $\boldsymbol{\psi}$.
T' End of time interval.
ψ_k State vector element.
j' Uptake half saturation (phytoplankton).
r Plant metabolic loss.
N'_0 Deep nutrients.
c Maximum grazing rate (herbivores).
f Grazing efficiency (herbivores).
K Grazing half saturation (herbivores).
P'_0 Grazing threshold.
g Loss to carnivores.

References

Apostol, T.M., 1969. Calculus, 2nd edn., vol. 2. Wiley.

Bennett, A.F., 1992. Inverse Methods in Physical Oceanography. Cambridge Univ. Press, ISBN 0-521-38568-7.

Bennett, A.F., Thorburn, M.A., 1992. The generalized inverse of a nonlinear quasigeostrophic ocean circulation model. J. Phys. Oceanogr. 22, 213–230.

Bennett, A.F., Chua, B.S., Leslie, L.M., 1997. Generalized inversion of a global numerical weather prediction model. Meteorol. Atmos. Phys. 60, 165–178.

Courant, R., Hilbert, D., 1953. Methods of Mathematical Physics, vol. 1. Wiley-Interscience, New York.

Daley, R., 1991. Atmospheric Data Analysis. Cambridge Univ. Press.

Eknes, M., Evensen, G., 1997. Parameter estimation solving a weak constraint variational formulation for an Ekman model. J. Geophys. Res. 102 (C6), 12479–12491.

Evans, G.T., Parslow, J.S., 1985. A model of annual plankton cycles. Biol. Oceanogr. 3 (3), 327–347.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (C5), 10143–10162.

Evensen, G., 1997. Advanced data assimilation for strongly nonlinear dynamics. Mon. Weather Rev. 125, 1342–1354.

Evensen, G., Fario, N., 1997. Solving for the generalized inverse of the Lorenz model. J. Meteorol. Soc. Jpn. 75, 229–243.

Evensen, G., Dee, D.P., Schröter, J., 1998. Parameter estimation in dynamical models. In: Chassignet, E.P., Verron, J. (Eds.), Ocean Modeling and Parameterizations. NATO ASI.

Fasham, M.J.R., Evans, G.T., 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47°N 20°W. Trans. R. Soc. Lond. Ser. B 348, 203–209.

Fletcher, R., Reaves, C.M., 1964. Function minimization by conjugate gradients. Comput. J. 7, 149–154.

Gelb, A., 1979. Applied Optimal Estimation. 5th edn. The Analytic Sciences, ISBN 0-262-70008-5.

Gill, P.E., Murray, W., Wright, M.H., 1982. Practical Optimization. Academic Press, London.

Goldstein, A.A., 1967. Constructive Real Analysis. Harper, New York.

Harmon, R., Challenor, P., 1996. A Markov chain Monte Carlo method for estimation and assimilation into models. Ecological Modelling 101, 41–59.

Hindmarsh, A.C., 1983. Odepack, a systematized collection of ode solvers. In: Stepleman, R.S. (Ed.), Scientific Computing. North-Holland, Amsterdam, pp. 55–64.

Hurtt, G.C., Armstrong, R.A. et al., 1996. A pelagic ecosystem model calibrated with BATS data. Deep-sea Res. II 43 (2–3), 653–683.

Jazwinski, A.H., 1970. Stochastic Processes and Filtering Theory. Academic Press, New York.

Kincaid, D., Cheney, W., 1991. Numerical Analysis. Brooks/Cole, ISBN 0-534-13014-3.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.

Lawson, L.M., Spitz, Y.H., Hoffmann, E.E., Long, R.B., 1995. A data assimilation technique applied to a predator–prey model. Bull. Math. Biol. 57 (4), 593–617.

Lawson, L.M., Hofmann, E.E., Spitz, Y.H., 1996. Time series sampling and data assimilation in a simple marine ecosystem model. Deep-sea Res. II 43 (2–3), 625–651.

Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: a case study at station P. J. Mar. Res. 53, 571–607.

McIntosh, P.C., 1990. Oceanographic data interpolation: objective analysis and splines. J. Geophys. Res. 95 (C8), 13529–13541.

McOwen, R.C., 1996. Partial Differential Equations. Prentice-Hall, ISBN 0-13-121880-8.

Miller, R.N., Ghil, M., Gauthiez, F., 1994. Advanced data assimilation in strongly nonlinear dynamical systems. J. Atmos. Sci. 51 (8), 1037–1056.

Minoux, M., 1986. Mathematical Programming—Theory and Algorithms. Wiley.

O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. 103 (C11), 24397–24953.

Prunet, P., Minster, J.-F., Ruiz-Pino, D., Dadou, I., 1996a. Assimilation of surface data in a one-dimensional physical–biogeochemical model of the surface ocean: 1. Method and preliminary results. Global Biogeochem. Cycles 10 (1), 111–138.

Prunet, P., Minster, J.-F., Echevin, V., Dadou, I., 1996b. Assimilation of surface data in a one-dimensional physical–biogeochemical model of the surface ocean: 2. Adjusting a simple trophic model to chlorophyll, temperature, nitrate and pCO_2 data. Global Biogeochem. Cycles 10 (1), 139–158.

Semovski, S.V., Woźniak, B., 1995. Model of the annual phytoplankton cycle in the marine ecosystem—assimilation of monthly satellite chlorophyll data for the North Atlantic and Baltic. Oceanologia 37 (1), 3–31.

Shanno, D.F., 1978. Conjugate gradient methods with inexact line searches. Math. Operations Res. 3, 244–256.

Spitz, Y.H., Moisan, J.R., Abbot, M.R., Richman, J.G., 1998. Data assimilation and a pelagic ecosystem model: parameterization using time series observations. J. Mar. Syst. 16, 51–68.