Application of model reduced 4D-Var to a 1D ecosystem model

Joanna S. Pelc[b,c,*], Ehouarn Simon[a], Laurent Bertino[a], Ghada El Serafy[c,b], Arnold W. Heemink[b]

[a]*Nansen Environmental and Remote Sensing Center, Bergen, Norway*
[b]*Delft University of Technology, Delft, The Netherlands*
[c]*Deltares, Delft, The Netherlands*

Abstract

The model reduced 4D-Var (Vermeulen and Heemink, 2006) is investigated to test its feasibility in ecosystem application. Ecological models are known for their high nonlinearity and issues with non-differentiability. However, since the method is performed in the reduced space, the implementation of the adjoint of the tangent linear approximation of the original model is not required.

Twin experiments are conducted in a 1D ecological model. Surface phytoplankton data are used, with 30% log-normally distributed measurement error. Three parameters are chosen for calibration. The method performs very well in the setup where a perfect initial condition is used, as well as for the combined parameter and initial condition estimation. A relatively well calibrated initial condition contributes to accurate parameter estimations. Not accounting for the wrongly assigned initial condition leads to incorrectly calibrated parameters.

Keywords: data assimilation, ecosystem, parameter estimation, initial condition control, model reduced 4D-Var, twin experiment

1. Introduction

In the presence of environmental issues caused by the climate change and eutrophication (Nixon, 1995; Hallegraeff, 1993, 2009, 2010; Pauly et al., 2002; Gregg et al., 2003, 2005; Peperzak, 2003; Paerl and Huisman, 2008; Brush, 2008; Schindler et al., 2008), ecosystem models capable to provide accurate predictions are of high interest. Despite many modeling challenges (Doney, 1999; Jørgensen, 2008), ecological models have evolved to be considered as relevant predictive tools. However, most sophisticated systems are not able to reproduce the reality completely, therefore it is essential to integrate them with the available obser-

[*] Corresponding author
Email address: j.s.pelc@tudelft.nl (Joanna S. Pelc)

vations. Data assimilation techniques serve as a tool to calibrate and improve model accuracy by combining them with the data given by the measurements.

The ecological models though often very advanced and complicated are still missing a unified set of rules, which would govern the whole system (Jørgensen, 2008; Los, 2009; Doron et al., 2011). This is compensated by the use of free parameters in the process parameterizations (Doron et al., 2011). An important source of uncertainties in ecosystem models is assigned to the poorly known parameters. With a given observation set, these could be estimated through laboratory experiments. However, when a large number of parameters is considered it is much harder to manage its calibration manually. Moreover, the calibration needs to be repeated every time a new ecosystem region is considered. Data assimilation methods provide tools to estimate many parameters simultaneously, with the possibility of accounting for their correlations.

There are two distinct classes of data assimilation methods, both capable to account for imperfect parameters. One is the class of variational (inverse) techniques, which searches for an optimal set of control variables, such that a cost function which measures the distance between the model and observations is minimized. The second class is represented by forward methods, which assimilate data sequentially in time. Hence, they are often referred as sequential methods. Their main target is to correct the estimated variables at every time an observation becomes available. The variational techniques are mainly used for parameter estimation, whereas the sequential methods are mainly used for state estimation. However, both approaches can tackle the task of the other one, and are equivalent for linear systems.

The variational techniques were successfully applied to improve the ecosystem predictions. Several approaches have been used, such as the ecosystem state estimation (Natvik et al., 2001), updating the input of a biological model by improving its coupled hydrodynamical model (Fiechter et al., 2011). The most common has been the calibration of the ecological parameters, where the adjoint technique was widely used to obtain the model gradients necessary to minimize the cost function (Fennel et al., 2001; Friedrichs, 2002; Zhao et al., 2005). Obtaining the adjoint of the model may be sometimes complicated, therefore methods which do not require the use of the gradient were also commonly used. This includes techniques such as simulated annealing (Matear, 1995), or genetic algorithms (Ward et al., 2010). However, when the number of parameters to be estimated is increased, these methods become computationally more demanding when compared to the gradient based techniques.

The sequential methods were also applied to perform data assimilation for ecosystems, and were mainly used for state estimation. One of the first sequential methods applied in ecosystem models was the Extended Kalman Filter (Carmillet et al., 2001). Further, applications of the Ensemble Kalman Filter to ecosystems became very common (Allen et al., 2002; Eknes and Evensen, 2002; Natvik and Evensen, 2003; Simon and Bertino, 2009). For a detailed overview of sequential methods used in oceanography and ecology see Bertino et al. (2003). Gaussian anamorphosis extensions of ensemble-based Kalman filters have been suggested by Bertino et al. (2003) to tackle the problems of sub-optimality of the filter raising from the non-Gaussian distributions of most of the biological variables and parameters. These approaches can be easily applied in realistic configurations (Simon and Bertino, 2009) and have been proved to be efficient tools to calibrate poorly known parameters (Doron et al., 2011; Simon and Bertino, 2012) For a detailed overview of data assimilation methods in application to biological models see Gregg (2008).

In this work the aim is to estimate parameters and the initial condition of the model, therefore a variational technique is a suitable choice. The method used in this work is four dimensional variational data assimilation (4D-Var), which is an adjoint method. It was first introduced in meteorology for the initial condition estimation (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987). It proved to be a powerful tool for ecosystem calibration. However, ecological models become more and more sophisticated, and the number of their biological components, as well as their parameters is increasing. Even simple ecosystem models have strong nonlinear behavior. Moreover biogeochemical dynamics often introduce nondifferentiability into the system. For such challenging environment obtaining its adjoint becomes nontrivial. Also the model resolutions are much finer than in the past, which results in large sizes of the model states. This introduces a limitation for using the finite difference gradient approximations, since these ones are not suitable for large problems.

A number of methods has been proposed to deal with this problem by obtaining the adjoint in a reduced space (Vermeulen and Heemink, 2006; Cao et al., 2007; Fang et al., 2009). Although the methods differ in their approach, all are based on the proper orthogonal decomposition (POD), also known as the Karhune-Loéve transform (KLT), principal component analysis (PCA) or the method of empirical orthogonal functions (EOF) (Pearson, 1901; Shlens, 2009). Although EOF-based methods are widely used in ecosystem data assimilation (Nerger and Gregg, 2007; Carmillet et al., 2001; Lermusiaux, 2006), none of the model reduced 4D-Var schemes have yet been used in ecological applications.

Based on a number of simulations of the original model, proper orthogonal decomposition is used to obtain a reduced model. The model-reduced 4D-Var is performed in the reduced space. Therefore, the implementation of the adjoint of the tangent linear approximation of the original model is not required. Instead, it is approximated by the adjoint of the tangent linear approximation of the reduced model. The method is easily extended to the estimation of the initial condition, hence the parameter calibration is coupled together with the initial condition estimation.

Due to the limitations of the adjoint model development for ecosystems, such approach may serve as an attractive tool for these applications. Therefore, the aim of this work is to evaluate the feasibility of a reduced adjoint approach in calibrating ecosystem models. The model-reduced 4D-Var proposed by Vermeulen and Heemink (2006) was effectively used in several applications, such as groundwater flow (Vermeulen et al., 2005), shallow-water flow (Altaf et al., 2009, 2010), oil reservoir optimization (Kaleta et al., 2011), and morphodynamics (Garcia et al., submitted). An advantage was shown especially for models characterized by periodic behavior (Altaf et al., 2009). Since for these type of models, the number of required model simulations to obtain the reduced model is relatively small. Due to the seasonality ecosystem models do show periodic characteristics, therefore this method serves as a potential tool for ecological applications.

The paper is organized as follows. First a brief description of a 1D Ecological model is presented in Section 2. Next the variational assimilation methodology is described in Section 3, with the incremental variational assimilation presented in Section 3.1, and the model reduced 4D-Var methodology described in Section 3.2. Framework of the experiments is presented in Section 4. Results are shown and discussed in Section 5, and finally the conclusions are presented in Section 6.

2. 1D Ecological Model

A 1D ecological model is used to illustrate the capabilities of the model reduced 4D-Var method in ecosystem applications. The model was first introduced by Evans and Parslow (1985). It is a simple differential equation model of nutrients N, phytoplankton P and herbivores H in a mixed layer of varying depth. Eknes and Evensen (2002) extended the model by a vertical mixing term, which resulted in a vertical dimension.

Eknes and Evensen (2002) replaced the terms corresponding to diffusion rates for the concentrations of nutrients, phytoplankton and zooplankton, with a vertical diffusion term of the form $(\partial/\partial z)(K_z(z, M)(\partial/\partial z))$. The K_z is the diffusion coefficient parametrized with respect to the depth z and mixed layer depth M. The smooth version of the mixed layer depth function $M = M(t)$ and its rate of change were used as described in Natvik et al. (2001). Further on, the mixed layer depth is used as a physical input for the ecosystem model.

The equations describing the ecosystem components evolution are given by

$$\frac{\partial N}{\partial t} = -\left(\frac{\alpha(t,z,P)N}{j+N} - r\right)P + \frac{\partial}{\partial z}\left(K_z(z, M(t))\frac{\partial N}{\partial z}\right) \tag{1a}$$

$$\frac{\partial P}{\partial t} = \left(\frac{\alpha(t,z,P)N}{j+N} - r\right)P - \frac{c(P-P_0)H}{K+P-P_0} + \frac{\partial}{\partial z}\left(K_z(z, M(t))\frac{\partial P}{\partial z}\right) \tag{1b}$$

$$\frac{\partial H}{\partial t} = \frac{fc(P-P_0)H}{K+P-P_0} - gH + \frac{\partial}{\partial z}\left(K_z(z, M(t))\frac{\partial H}{\partial z}\right) \tag{1c}$$

where $\alpha = \alpha(t, z, P)$ is the light-limited photosynthetic rate, and $K_z = K_z(z, M(t))$ is the depth dependent diffusion parameter. Both parameters are described in detail in Eknes and Evensen (2002). The other parameters are listed in Table 1.

Their values were calibrated according to Flemish Cap (47°N, east of Newfoundland, Canada).

Symbol	Description	Value	Unit
c	Maximum grazing rate	1.0	1/day
f	Grazing efficiency	0.50	
g	Loss to carnivores	0.07	1/day
j	Uptake half saturation	0.5	mmol N m^{-3}
r	Plant metabolic loss	0.07	1/day
K	Grazing half saturation	1.0	mmol N m^{-3}
P_0	Grazing threshold	0.1	mmol N m^{-3}

Table 1: Physical parameters, appropriate to Flemish Cap, used in the data assimilation experiments. These are the same parameters as used by Eknes and Evensen (2002) and Evans and Parslow (1985)

Figure 1 presents the yearly cycle of the model. The nutrient concentration is constant at the bottom, which serves as an infinite pool of nutrient throughout the whole year. With the vertical mixing term, the nutrients are mixed into the biologically active mixed layer. Next, the nutrients are taken up by the phytoplankton during the spring bloom. Further the phytoplankton abundance and the bloom duration are determined by the nutrient availability and the herbivorous zooplankton grazing.

The three ecosystem components $N(t), P(t), H(t)$ are defined on 20 layers, which divide uniformly a water column of 200 meters deep. All together they form a state vector $\mathbf{x}(t) = [N(t), P(t), H(t)]$, which consists in total of 60 values at any time t. Further on, $\mathbf{x}(t)$ is used to write the model equations (1) in a compact form as follows

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{f}(\mathbf{x}, t) \tag{2}$$

where $\mathbf{f} = [\mathbf{f}_N, \mathbf{f}_P, \mathbf{f}_H]$, with each component corresponding to the right hand side of the model equations (1), respectively.

3. Variational data assimilation

The main purpose of data assimilation methods is to combine theoretical knowledge, given by the equations based on the laws of physics, together with practical knowledge represented by measurements of the system. The theoretical knowledge is represented in form of a model, which in this work is represented by the three component ecological system given by (2). Further on the model is discretized and its numerical approximation results in the following forward model

$$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}, \boldsymbol{\alpha}) \tag{3}$$

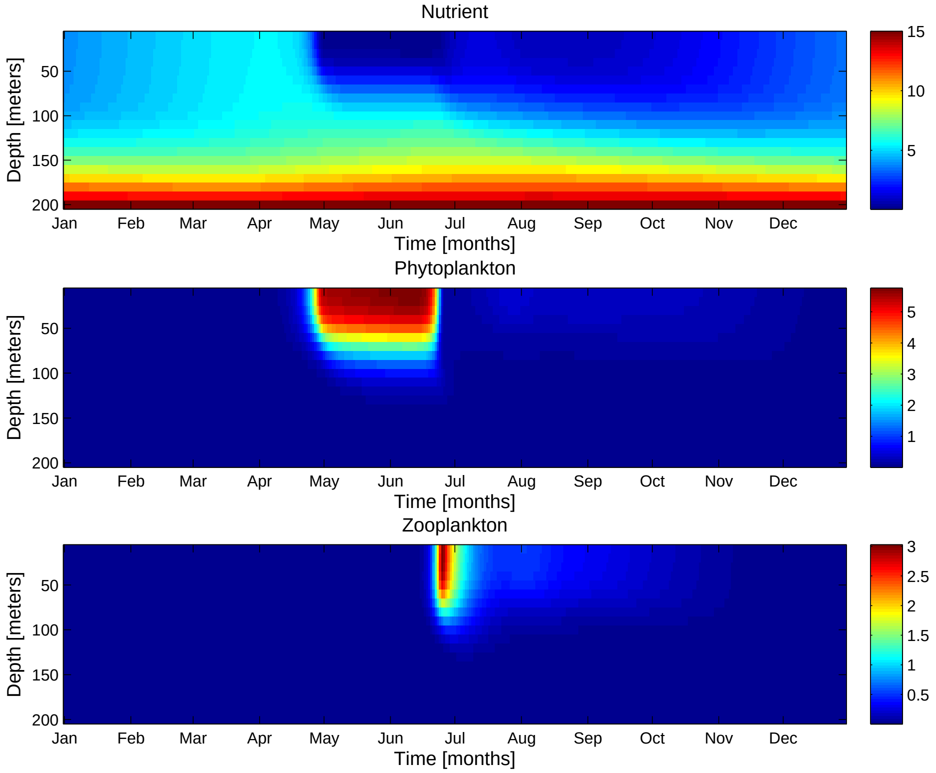

Figure 1: Yearly annual cycle: temporal evolution of Nutrients, Phytoplankton and Herbivorous Zooplankton concentrations in a water column. All variables are in units: mmol N m^{-3}.

where $\mathbf{x}_i$ stands for the model state at time t_i, $\mathcal{M}_i$ is the model that forecasts the state from time t_{i-1} to t_i, and $\boldsymbol{\alpha}$ is a set of parameters which model $\mathcal{M}_i$ depends on. The state vector $\mathbf{x}_i$ represents the values of the modeled variables specified for each of the grid cells. The main uncertainty is assumed to come from unknown parameters and initial conditions, therefore no model error terms were introduced. Namely, the model is assumed to be perfect.

The practical knowledge is represented by measurements of events which relate to the model output $\mathbf{x}_i$. The observational operator, which translates the output given by the model into an observational space, is given as follows

$$\mathbf{y}_i = \mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) + \boldsymbol{\epsilon}_i \tag{4}$$

where $i = 1, .., n_{\mathbf{y}}$, $\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha})$ is the observation operator at time t_i. The output of $\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha})$ is a theoretical counterpart of the measurement $\mathbf{y}_i$ for a given model state $\mathbf{x}_i$ and set of parameters $\boldsymbol{\alpha}$, at the time t_i. $n_{\mathbf{y}}$ stands here for a number of discrete times at which an observation occurred. The observations rarely represent the reality accurately, very often they carry a significant measurement error, as well as a representativeness error. Therefore, an observational error at time t_i is introduced, and it is denoted as $\boldsymbol{\epsilon}_i = \boldsymbol{\epsilon}_i^o + \boldsymbol{\epsilon}_i^r$. $\boldsymbol{\epsilon}_i^o$ stands for the "instrument" measurement error, and $\boldsymbol{\epsilon}_i^r$ stands for the observational error due to the representativeness. In the considered setup the measurements are generated and assimilated at the model grid points, therefore in this work the representativeness error is known, and it is equal to zero, i.e. $\boldsymbol{\epsilon}_i^r = 0$.

Once both sources of information has been brought into one space, it is possible to establish a distance between the model trajectory and the given observations. The idea is to find a set of optimal values of parameters, states, initial conditions, inputs etc, which minimize these distances. The variables chosen to be estimated, are usually referred to as *control variables* or *a control vector*. Next, a cost function is defined which is a measure of all the distances along the assimilation time for a given set of control variables $\mathbf{c}$

$$J(\mathbf{c}) = \frac{1}{2}(\mathbf{c} - \mathbf{c}^b)^T \mathbf{B}_{\mathbf{c}}^{-1}(\mathbf{c} - \mathbf{c}^b) + \frac{1}{2}\sum_{i=1}^{n_{\mathbf{y}}}(\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) - \mathbf{y}_i) \quad (5)$$

where $\mathbf{c}^b$ is a prior (background) knowledge about the control variables $\mathbf{c}$, $\mathbf{B}_{\mathbf{c}}$ is a background error covariance matrix, which represents uncertainty of the prior (background) information about the control variables $\mathbf{c}$, $\mathbf{R}_i$ is an observation error covariance matrix representing uncertainty about the observations taken at time t_i and uncertainty about the observation operator at time t_i. Sometimes, the observations are not enough to explain the entire control variable domain. Therefore the distance from their background values need to be controlled to keep estimated variables in their range.

In order to minimize J, the gradient of J with respect to the control variables is needed. In the following section the cost function J is formulated in an incremental approach, and by means of the adjoint technique, its gradient with respect to the control variables is derived. The incremental technique is an essential link in understanding of the model-reduced 4D-Var technique, which is explained in the consecutive section.

3.1. Incremental 4D-Var

In the incremental 4D-Var approach (Courtier et al., 1994) the model is linearized with respect to the control vector around its best guess. Next the cost function is minimized being constrained on the linearized version of the model (3). The set of control variables which minimizes the cost function is the optimal solution in the linearized space of solutions. However, if the full model is nonlinear the same solution will only be suboptimal for the full space. Therefore the procedure needs to be repeated: the suboptimal solution becomes a new best guess, and the model is again linearized. The procedure repetitions are called outer loops. The minimization iterations within each outer loop are called inner loops. With stronger nonlinearities of the model, the number of outer loops need to be increased.

An advantage of the incremental approach over the classical one, is that once the model is linearized, then the gradient of the model remains constant within each outer loop. Hence, it does not need to be recalculated at every inner loop, unlike in the classical approach.

First the model equations (3) are linearized with respect to the control vector, which in this work consists of parameter vector $\boldsymbol{\alpha}$ and the initial condition $\mathbf{x}_0$. The linearization is considered around the best guess which in this case is the background value of the control vector $\mathbf{c}^b = \left[\mathbf{x}_0^b, \boldsymbol{\alpha}^b\right]^T$

$$\delta\mathbf{x}_i = \frac{\partial\mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^b)}{\partial\mathbf{x}_{i-1}}\delta\mathbf{x}_{i-1} + \frac{\partial\mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^b)}{\partial\boldsymbol{\alpha}}\delta\boldsymbol{\alpha} \tag{6}$$

where $\mathbf{x}_i^b$ stands for the background model output, i.e. originated at the background initial condition $\mathbf{x}_0^b$ and background parameters $\boldsymbol{\alpha}^b$. By introducing the following notation

$$\mathbf{M}_j^{\boldsymbol{\alpha}} = \frac{\partial\mathcal{M}_j(\mathbf{x}_{j-1}^b, \boldsymbol{\alpha}^b)}{\partial\boldsymbol{\alpha}} \tag{7a}$$

$$\mathbf{M}_j^{\mathbf{x}} = \frac{\partial\mathcal{M}_j(\mathbf{x}_{j-1}^b, \boldsymbol{\alpha}^b)}{\partial\mathbf{x}_{j-1}} \tag{7b}$$

and

$$\mathbb{M}_{i,j}^{\mathbf{x}} = \mathbf{M}_i^{\mathbf{x}} \cdot \mathbf{M}_{i-1}^{\mathbf{x}} \cdot ... \cdot \mathbf{M}_j^{\mathbf{x}}, \qquad j = 1, ..., i \tag{8}$$

the linearized model can be rewritten as follows

$$\delta\mathbf{x}_i = \mathbb{M}_{i,1}^{\mathbf{x}}\delta\mathbf{x}_0 + \sum_{j=1}^{i}\mathbb{M}_{i,j+1}^{\mathbf{x}}\mathbf{M}_j^{\boldsymbol{\alpha}}\delta\boldsymbol{\alpha} \tag{9}$$

where $\mathbb{M}_{i,i+1}^{\mathbf{x}} = \mathbb{I}$, and $\mathbb{I}$ is an identity operator of the same size. For simplicity the following notation is introduced

$$\mathbf{G}_i^{\boldsymbol{\alpha}} = \sum_{j=1}^{i}\mathbb{M}_{i,j+1}^{\mathbf{x}}\mathbf{M}_j^{\boldsymbol{\alpha}} \tag{10}$$

Then the linearized model (9) can be rewritten as

$$\delta\mathbf{x}_i = \mathbb{M}_{i,1}^{\mathbf{x}}\delta\mathbf{x}_0 + \mathbf{G}_i^{\boldsymbol{\alpha}}\delta\boldsymbol{\alpha} \tag{11}$$

Notice, that $\mathbf{G}_i^{\boldsymbol{\alpha}}$ is the sensitivity of the model with respect to the parameter change, and $\mathbb{M}_{i,1}^{\mathbf{x}}$ is the sensitivity of the model to the initial condition change, both at the time t_i.

Often it happens that the observations are nonlinearly related with the model state or parameters. In that case also the observation operator needs to be linearized with respect to the control vector. Hence

$$\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) \simeq \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b) + \frac{\partial \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b)}{\partial \mathbf{x}_i}\delta\mathbf{x}_i + \frac{\partial \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b)}{\partial \boldsymbol{\alpha}}\delta\boldsymbol{\alpha} \tag{12}$$

Next, by introducing extra notation

$$\mathbf{H}_i^{\mathbf{x}} = \frac{\partial \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b)}{\partial \mathbf{x}_i} \tag{13a}$$

$$\mathbf{H}_i^{\boldsymbol{\alpha}} = \frac{\partial \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b)}{\partial \boldsymbol{\alpha}} \tag{13b}$$

and combining the equation (12) with the relation (11), the linearization of the observation operator can be written in terms of the increments $\delta\mathbf{x}_0$ and $\delta\boldsymbol{\alpha}$, i.e.

$$\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) \simeq \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b) + \mathbf{H}_i^{\mathbf{x}}\mathbb{M}_{i,1}^{\mathbf{x}}\delta\mathbf{x}_0 + (\mathbf{H}_i^{\mathbf{x}}\mathbf{G}_i^{\boldsymbol{\alpha}} + \mathbf{H}_i^{\boldsymbol{\alpha}})\delta\boldsymbol{\alpha} \tag{14}$$

Then, let $\delta\mathbf{c} = [\delta\mathbf{x}_0, \delta\boldsymbol{\alpha}]^T$, and let the sensitivity of the observational operator with respect to $\delta\mathbf{c}$ be defined as

$$\mathbf{H}_i^{\mathbf{c}} = [\mathbf{H}_i^{\mathbf{x}}\ \mathbb{M}_{i,1}^{\mathbf{x}},\ \ \mathbf{H}_i^{\mathbf{x}}\mathbf{G}_i^{\boldsymbol{\alpha}} + \mathbf{H}_i^{\boldsymbol{\alpha}}] \tag{15}$$

Thus, the approximation of the observational operator becomes

$$\mathcal{H}_i(\mathbf{x}_i, \boldsymbol{\alpha}) \simeq \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b) + \mathbf{H}_i^{\mathbf{c}}\delta\mathbf{c} \tag{16}$$

Now, by using the approximation of the observational operator (16) and linearized control vector $\mathbf{c} = \mathbf{c}^b + \delta\mathbf{c}$, the cost function (5) is approximated by

$$\hat{J}(\delta\mathbf{c}) = \frac{1}{2}\delta\mathbf{c}^T\mathbf{B}_{\mathbf{c}}^{-1}\delta\mathbf{c} + \frac{1}{2}\sum_{i=1}^{n_{\mathbf{y}}}\big(\mathbf{H}_i^{\mathbf{c}}\delta\mathbf{c} + \mathbf{d}_i\big)^T\mathbf{R}_i^{-1}\big(\mathbf{H}_i^{\mathbf{c}}\delta\mathbf{c} + \mathbf{d}_i\big) \tag{17}$$

where $\mathbf{d}_i = \mathcal{H}_i(\mathbf{x}_i^b, \boldsymbol{\alpha}^b) - \mathbf{y}_i$, and $\hat{J}$ is the incremental 4D-Var cost function. The value $\hat{J}(\delta\mathbf{c})$ is approximating the value of the nonlinear cost function around the background control vector $J(\mathbf{c}^b + \delta\mathbf{c})$. Note, the incremental cost function $\hat{J}$ depends directly on $\delta\mathbf{c}$, which is therefore considered the control variable in the incremental formulation.

The next step is the gradient formulation. To this end the incremental change of the cost function (17) has to be investigated with respect to the change of the increment $\delta\mathbf{c}$. According to the adjoint technique for calculating gradients (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987), it results in the following gradient of the cost function $\hat{J}$ with respect to the increment $\delta\mathbf{c}$

$$\nabla_{\delta\mathbf{c}}\hat{J}(\delta\mathbf{c}) = \mathbf{B}_{\mathbf{c}}^{-1}\delta\mathbf{c} + \sum_{i=1}^{n_{\mathbf{y}}}(\mathbf{H}_i^{\mathbf{c}})^T\mathbf{R}_i^{-1}(\mathbf{H}_i^{\mathbf{c}}\delta\mathbf{c} + \mathbf{d}_i) \tag{18}$$

The incremental 4D-Var is a very useful tool to estimate parameters and initial conditions, as well as other types of inputs of the model. However, even though the model is linearized in this approach, which significantly improves the efficiency of the method, the gradient of the model states still needs to be evaluated for every outer loop. The following section describes a technique, where the gradient is approximated using a reduced model approach.

3.2. Model Reduced 4D-Var

The model reduced 4D-Var (Vermeulen and Heemink, 2006) is a method proposed to avoid the implementation of the adjoint of the tangent linear approximation of the original model. The key idea of the method is the way it tackles the evaluation of the derivative of the model with respect to the state. With an orthogonal projection matrix $\mathbf{P}$, the derivative in question is mapped into a smaller subspace. Then it is considered as the directional derivative in the directions, specified by the columns of $\mathbf{P}$. Construction of the matrix $\mathbf{P}$ assures that essential information is captured in a small number of columns, hence the finite difference approximation becomes feasible.

$$\frac{\partial\mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial\mathbf{x}_{i-1}}\mathbf{P} \simeq \frac{\mathcal{M}_i(\mathbf{x}_{i-1}^b + \epsilon\,\mathbf{P}, \boldsymbol{\alpha}^{\boldsymbol{b}}) - \mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\epsilon} \tag{19}$$

In order to incorporate the property (19) into the linearized model equations (11), the equations (11) as well need to be projected into the reduced space. Hence the following is obtained

$$\mathbf{P}^T\delta\mathbf{x}_i = \mathbf{P}^T\frac{\partial\mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial\mathbf{x}_{i-1}}\mathbf{P}\mathbf{P}^T\delta\mathbf{x}_{i-1} + \mathbf{P}^T\frac{\partial\mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial\boldsymbol{\alpha}}\delta\boldsymbol{\alpha} \tag{20}$$

where $\mathbf{P}^T$ maps the model states into a smaller subspace, whereas $\mathbf{P}$ maps the reduced states back into the full space. The columns of $\mathbf{P}$ are referred as *patterns.* $\mathbf{P}$ is of the size $n_{\mathbf{x}} \times n_{\mathbf{P}}$, where $n_{\mathbf{x}}$ is the size of the state vector and $n_{\mathbf{P}}$ is the size of the reduced space (number of patterns). Matrix $\mathbf{P}$ is constructed using the POD method which guarantees the best reconstruction of the model states in the root mean squared error sense (Antoulas, 2005; Kaleta, 2011). Hence, the following values $||\delta\mathbf{x}_i - \mathbf{P}\mathbf{P}^T\delta\mathbf{x}_i||_2$ are minimal, where $||\cdot||_2$ denotes the L_2 norm. Appendix A explains how the projection matrix $\mathbf{P}$ is obtained.

The projected model increments are denoted as $\delta\mathbf{z}_i = \mathbf{P}^T\delta\mathbf{x}_i$, where $\delta\mathbf{z}_i$ is the reduced model increment. Hence the equation (20) is rewritten as follows

$$\delta\mathbf{z}_i = \mathbf{P}^T \frac{\partial \mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial \mathbf{x}_{i-1}} \mathbf{P}\delta\mathbf{z}_{i-1} + \mathbf{P}^T \frac{\partial \mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial \boldsymbol{\alpha}} \delta\boldsymbol{\alpha} \tag{21}$$

Matrix $\mathbf{P}$ is used to map the variables $\delta\mathbf{z}_i$ from the reduced space back into the full space, i.e. $\delta\hat{\mathbf{x}}_i = \mathbf{P}\delta\mathbf{z}_i$, where $\delta\hat{\mathbf{x}}_i$ is an approximation of the full space increment, and $||\delta\mathbf{x}_i - \delta\hat{\mathbf{x}}_i||_2$ is relatively small. Next, by introducing the following notation

$$\mathbf{N}_{\mathbf{i}}^{\mathbf{x}} = \mathbf{P}^T \frac{\partial \mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial \mathbf{x}_{i-1}} \mathbf{P} \tag{22a}$$

$$\mathbf{N}_{\mathbf{i}}^{\boldsymbol{\alpha}} = \mathbf{P}^T \frac{\partial \mathcal{M}_i(\mathbf{x}_{i-1}^b, \boldsymbol{\alpha}^{\boldsymbol{b}})}{\partial \boldsymbol{\alpha}} \tag{22b}$$

the equation (21) is rewritten as follows

$$\delta\mathbf{z}_i = \mathbf{N}_{\mathbf{i}}^{\mathbf{x}} \delta\mathbf{z}_{i-1} + \mathbf{N}_{\mathbf{i}}^{\boldsymbol{\alpha}} \delta\boldsymbol{\alpha} \tag{23}$$

Next, analogically to (8) the following product is introduced

$$\mathbb{N}_{i,j}^{\mathbf{x}} = \mathbf{N}_{\mathbf{i}}^{\mathbf{x}} \cdot \mathbf{N}_{\mathbf{i-1}}^{\mathbf{x}} \cdot \ldots \cdot \mathbf{N}_{\mathbf{j}}^{\mathbf{x}} \tag{24}$$

and the reduced tangent linearized model (23) is rewritten as follows

$$\delta\mathbf{z}_i = \mathbb{N}_{i,1}^{\mathbf{x}}\ \delta\mathbf{z}_0 + \sum_{j=1}^{i} \mathbb{N}_{i,j+1}^{\mathbf{x}} \mathbf{N}_{\mathbf{j}}^{\boldsymbol{\alpha}} \delta\boldsymbol{\alpha} \tag{25}$$

where $\mathbb{N}_{i,i+1}^{\mathbf{x}}$ is an identity operator in the appropriate size. Analogically to the previous section the following notation is introduced

$$\hat{\mathbf{G}}_i^{\boldsymbol{\alpha}} = \sum_{j=1}^{i} \mathbb{N}_{i,j+1}^{\mathbf{x}} \mathbf{N}_{\mathbf{j}}^{\boldsymbol{\alpha}} \tag{26}$$

Using the property $\delta\hat{\mathbf{x}}_i = \mathbf{P}\delta\mathbf{z}_i$, the linearized model can be rewritten as follows

$$\delta\hat{\mathbf{x}}_i = \mathbf{P}\mathbb{N}_{i,1}^{\mathbf{x}}\ \delta\mathbf{z}_0 + \mathbf{P}\hat{\mathbf{G}}_i^{\boldsymbol{\alpha}} \delta\boldsymbol{\alpha} \tag{27}$$

The cost function in the model reduced approach takes the same form as the incremental cost function in the non-reduced approach, see equation (17). However, the control variable increment is now expressed as $\delta\mathbf{c} = [\delta\mathbf{z}_0,\ \delta\boldsymbol{\alpha}]^T$, and the linearized observation operator is formulated accordingly to the reduced space as follows

$$\mathbf{H}_i^{\mathbf{c}} = [\mathbf{H}_i^{\mathbf{x}}\mathbf{P}\mathbb{N}_{i,1}^{\mathbf{x}},\ \mathbf{H}_i^{\mathbf{x}}\mathbf{P}\hat{\mathbf{G}}_i^{\boldsymbol{\alpha}} + \mathbf{H}_i^{\boldsymbol{\alpha}}] \tag{28}$$

The gradient of the cost function in the model reduced approach is just as expressed in equation (18), with appropriate $\delta \mathbf{c}$ and $\mathbf{H}_i^{\mathbf{c}}$.

4. Framework of the experiments

Model reduced four dimensional variational data assimilation is implemented for the 1D ecological model. Three parameters of interest are considered for estimation: grazing efficiency (f), loss to carnivores (g) and plant metabolic loss (r), the choice as in Simon and Bertino (2012). In the experiments where the initial condition is assumed to be unknown, it is also included as one of the control variables.

4.1. Twin experiment setup

Three twin experiments are presented. Each is designed, such that the same true solution is used for all setups, as well as the same set of observations. For each setup the true parameter set is obtained by shifting the background parameter values f, g, and r, by 50%, 43% and 43% of each prior value respectively. Hence, the following set of the true parameters is obtained $f = 0.75$, $g = 0.10$, $r = 0.10$. The background parameter values used are equal with the ones listed in the Table 1, and they are used as the starting values for each data assimilation experiment. Five year assimilation window is used for all experiments, where each is preceded with one year of spin up, used to obtain the initial conditions. The same true initial condition is used for all experiments. Whereas the initial condition assigned within data assimilation varies depending on the experiment setup.

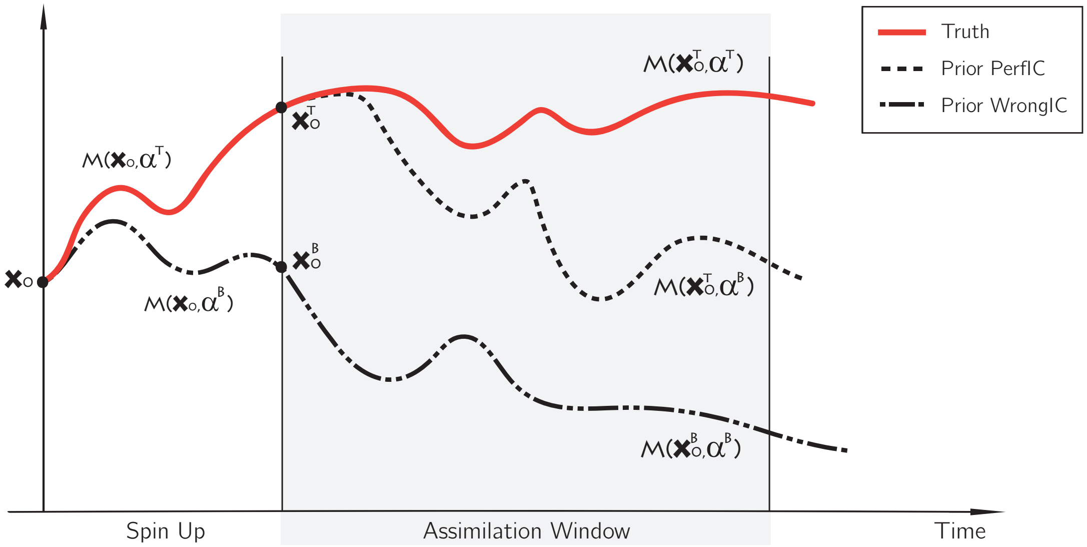
Truth
Prior PerfIC
Prior WrongIC
$M(\mathbf{x}_0^T, \alpha^T)$
$\mathbf{x}_0^T$
$M(\mathbf{x}_0, \alpha^T)$
$\mathbf{x}_0^B$
$\mathbf{x}_0$
$M(\mathbf{x}_0, \alpha^B)$
$M(\mathbf{x}_0^T, \alpha^B)$
$M(\mathbf{x}_0^B, \alpha^B)$
Spin Up
Assimilation Window
Time

Figure 2: Scheme of the twin experiment design

Two initial conditions are used to set up the data assimilation experiments: the prior and the true initial condition. Both are generated by running the model for a period of time (the spin up), one year in this case, where each is originated with the same values. After one year of simulation they arrive at different solutions. The discrepancy comes from distinct parameters assigned. The true initial condition is obtained using the true parameter set, whereas the prior initial condition is generated with the background parameters, see the scheme of the experiment design in Figure 2. Since the prior initial condition differs from the true one, it is also referred as wrong initial condition.

In the first experiment only parameters are estimated, whereas the initial condition is assumed to be perfect. Therefore, the true initial condition is assigned within data assimilation experiment. The purpose of this experiment is to examine whether the method calibrates the parameters correctly in the environment where the initial condition does not introduce extra uncertainty. This experiment is referred as 4DVar-Par&PerfIC.

In the two other experiments the prior initial condition is used to set up data assimilation. This way the calibration of the parameters needs to be coupled with the control of the initial condition in order to obtain an optimal update of the parameters. This approach is covered by the second experiment, and it is referred as 4DVar-Par&IC. The prior initial condition is used here to give starting values for the estimation of the initial condition. It is expected, that the control of the initial condition is essential. However, to see its importance, and compare to a situation where it was not accounted for, the third experiment is performed. In this setup only parameters are estimated while the prior initial condition is assigned, hence its name 4DVar-Par&WrongIC.

Typically the initial condition has an impact on the model only at the beginning of the simulation, then it is "forgotten" and the rest of the time is governed only by the parameters. Therefore, estimating the initial conditions under assumption of wrong parameters is not expected to work, unless it is performed for the assimilation window covering a period when the initial condition has an influence on the model (here it is about one year, although it may vary for other models). To illustrate that a fourth experiment is presented, where parameters are fixed at their prior values, and only the initial condition is estimated during the assimilation. The experiment is called 4DVar-WrongPar&IC.

4.2. Observations

The common source of measurements in ecology are satellite images of chlorophyll, which is further processed into the surface phytoplankton concentration. Hence, in the experiments the observations are generated only for phytoplankton concentration, for the first and second layer (the surface layers) of the model. The measurements are generated using the true solution with log-normally distributed observational noise (Campbell, 1995).

Commonly the observation error in satellite surface chlorophyll data is assumed to be around 30% of the data (Gregg and Casey, 2004). In these experiments, we assume that the error in the surface phytoplankton concentration follow the same rate. Observations are assumed to be available every four days of the assimilation window. The measurement error is assumed to be known within the experiments. Hence, the error covariance matrix $\mathbf{R}$ contains the real observation errors on its diagonal, whereas the off diagonal terms are equal to zero.

4.3. Background error covariance matrices

The background error covariance matrix $\mathbf{B_c}$ used in the experiments is defined as follows

$$\mathbf{B_c} = \begin{bmatrix} \mathbf{B}_{\boldsymbol{\alpha}} & \mathbf{0} \\ \mathbf{0} & \beta \mathbf{B}_{\mathbf{x}_0} \end{bmatrix} \tag{29}$$

where $\mathbf{B}_{\boldsymbol{\alpha}}$ is the background error covariance matrix for the parameters, $\mathbf{B}_{\mathbf{x}_0}$ is the background error covariance matrix for the initial condition, $\mathbf{0}$ is used symbolically to represent zero matrices in appropriate sizes, β is used as a scaling factor to balance the importance between the two background components while minimization, in this work $\beta = 0.01$. $\mathbf{B}_{\boldsymbol{\alpha}}$ contains the parameter variances on its diagonal, and its off diagonal terms are equal to zero. Hence, the parameters are treated independently within the data assimilation experiments. $\mathbf{B}_{\mathbf{x}_0}$ is a block diagonal matrix defined as follows

$$\mathbf{B}_{\mathbf{x}_0} = \begin{bmatrix} \mathbf{B}_{N_0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{P_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{H_0} \end{bmatrix} \tag{30}$$

where $\mathbf{B}_{N_0}, \mathbf{B}_{P_0}, \mathbf{B}_{H_0}$ stands for the error background covariance matrices for Nutrients, Phytoplankton and Herbivorous Zooplankton respectively. Each of these matrices were generated using the Gaussian covariance function with characteristic length scale equal to 50 meters (”squared exponential” in Rasmussen and Williams (2006)).

4.4. Minimization

The minimization of the cost function was performed using M1QN3 Quasi-Newton method described and implemented by Gilbert and Lemaréchal (1989). Since it is unconstrained type of minimization, the control variables were truncated in case of exceeding their domain after each outer loop minimization. The parameter bounds were chosen as follows: $0.1 < f < 1$, and $0.01 < g, r < 0.15$, whereas the initial values of the states were constrained by 0 and 20 mmol N m^{-3}. In this relatively simple model the control variables rarely needed to be truncated. However, in larger applications it is advisable to use more sophisticated tools to constrain the estimates, for example the anamorphosis function (Simon and Bertino, 2009).

4.5. Reduced model setup

In order to obtain the reduced model described in Section 3.2 the projection matrix $\mathbf{P}$ is created (see Appendix A). To this end, an ensemble of model dynamics is generated by perturbing each parameter of interest separately, and collecting the corresponding model output. Each of the estimated parameters is perturbed by a fixed amount of 25% of the parameter value, and for each perturbation a separate five years model simulation is performed. Then the snapshots are collected at every 2 days, which results in 913 snapshots in time, for each parameter. The snapshot collection is used to create a covariance matrix. Next, its most significant eigenvectors create projection matrix $\mathbf{P}$.

In order to account for the model dynamics due to initial condition variability, also a perturbation of the initial condition should be included in the snapshots. However, in this work the misfit in the initial condition was generated by perturbing the parameters and running the model for a certain amount of time (see Figure 2). This way the perturbation in the initial condition remained physical. Therefore, here the snapshots obtained from the perturbed parameters are enough to represent the dynamics for both the initial condition and the parameters.

If the state variables have different magnitudes, then taking the snapshots for all of them together to create the covariance matrix will lead to a loss of information. Namely, the eigenvalue decomposition will detect the dynamics of the state concentration with the highest magnitude as the most relevant, whereas the rest of the model state dynamics will be neglected. There are two ways to overcome this problem: (1) transform the states variables such that each of them is scaled to be of the same magnitude, (2) create the covariance matrix for each of the model constituents separately (Kaleta et al., 2011). The first approach is likely to result in a smaller number of the significant eigenvectors, and since it links all the substances together, it is also expected to improve the initial concentration estimations. The second approach is more generic since it does not need the information about the magnitudes of the model states. Although, its construction does not account for the correlations between the model states, it does not cause the loss in the accuracy of the adjoint approximation. Note, both approaches are expected to perform equivalently for the directional derivative approximations, since these are calculated for each substance separately in either case. Therefore, both approaches are equivalent in their performance of the adjoint approximation. In this work we follow the second approach.

The snapshot collection is used to create a covariance matrices for each of the model concentration separately. Next a selection of their most significant eigenvectors create projection matrices $\mathbf{P}_N, \mathbf{P}_P, \mathbf{P}_H$. Further on, the general projection matrix $\mathbf{P}$ is created by placing each matrix block-wise on a diagonal, resulting in the following pseudo-orthogonal matrix

$$\mathbf{P} = \left[\begin{array}{ccc} \mathbf{P}_N & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_H \end{array} \right] \tag{31}$$

The reduced model is chosen to carry 99% of the information (A.8) of the full model. With every outer loop the reduced model is recreated, hence the number of patterns corresponding to the selected amount of information may vary. However, in this work the number of patterns does not fluctuate much. For Nutrients it always results in 4 patterns, for Phytoplankton in most of cases it is 3 patterns, while very rarely 2 or 4, and the number for Herbivorous Zooplankton varies between 2 and 4 patterns, with predominance of 4. The converging number of patterns in all presented experiments is 4, 3 and 4 respectively for N, P and H. Thus the total number of patterns results in most of cases to 11.

The 60 values corresponding to the model state vector are represented in the reduced space by about 11 values (the exact number depends on the total number of patterns). Therefore, the total size of the control vector consisting of parameters and initial condition is reduced from 63 to only 14 (11 patterns and 3 parameters). Number of model runs needed to perform each outer loop is equal to the total number of patterns (about 11), plus the number of parameters (3) and plus one background run, which sums up to about 15 model runs per one outer loop.

5. Results and Discussion

The results of the experiment 4DVar-Par&PerfIC shows that the method deals very well with the calibration of the parameters, when the initial condition is known. The challenge of this experiment was the recovery of the true parameters, since these were strongly biased from their prior counterparts. Within 11 outer loops the method converged obtaining a good accuracy of the estimated parameters, their percentage error was reduced from about 30% down to 0.06% 0.07% 0.097% respectively for f, g and r. Since the parameters were the only source of misfit in this experiments, it is expected that their well calibrated values will assure good model match with the observations (Figure 7). This is also confirmed by the final value of the cost function, which converges approximately to its expected minimum given by 456, which is the half of the observation number (Tarantola, 1987), see Figure 4. Well calibrated parameters assured also good match of the model with the truth for the layers and variables which were not measured, for the reference see the water column profiles in Figure 5, and the root mean square error of the state variables plotted as the time series in Figure 9.

Based on the experiment 4DVar-Par&PerfIC it can be concluded that once the

	Parameters			Cost Fun	#OL
	f	g	r		
Prior	0.5000	0.0700	0.0700	4.29e+5	
	+50%	+43%	+43%		
Truth	**0.7500**	**0.1000**	**0.1000**		
Perfect Initial Condition Setup					
4DVar-Par&PerfIC	0.7496	0.0999	0.1001	436.13	11
Perturbed Initial Condition Setup					
4DVar-Par&WrongIC	0.9933	0.1526	0.1049	1196.49	9
4DVar-Par&IC	0.7266	0.0947	0.0997	451.27	30
4DVar-WrongPar&IC	-	-	-	3.19e+5	1
Number of observations				2 × 456	
	Parameter Percentage Error				
Prior	33.33	30.00	30.00		
Perfect Initial Condition Setup					
4DVar-Par&PerfIC	0.06	0.07	0.097		
Perturbed Initial Condition Setup					
4DVar-Par&WrongIC	32.44	52.65	4.94		
4DVar-Par&IC	3.12	5.27	0.34		

Table 2: Summarized results from the experiments

initial condition is known, the parameters can be easily calibrated. Therefore, in the next experiment, where both parameters and initial condition are perturbed, the main challenge was the control of the initial condition. The results of its estimation are presented in Figure 6, which shows quite good performance. The best match can be seen for the phytoplankton initial concentration, the zooplankton also is performing quite well, however, the initial values of nutrients moved slightly away from its prior towards the wrong direction. The best performance of the phytoplankton is expected, since this variable is the only one measured. The performance of the initial condition estimation for the other state variables could be improved by accounting for their correlations in the error covariance matrix $\mathbf{B}_{\mathbf{x}_0}$ defined in equation (30).

The model matched the observations quite well, as shown in blue in Figure 7. Only during the first assimilation year there is a loss in the bloom magnitude, which results from the underestimated initial nutrient concentration. However, the general accuracy of the initial condition was good enough to contribute to good calibration of the parameters, as shown in Figure 3. They were not as precise as in the experiment 4DVar-Par&PerfIC, although they obtained relatively good accuracy, with the parameter percentage errors 3.12%, 5.27% and 0.34% respectively for f, g, and r. Similarly the model performance outside of the measured areas is also slightly worse than the previous experiment, however it does

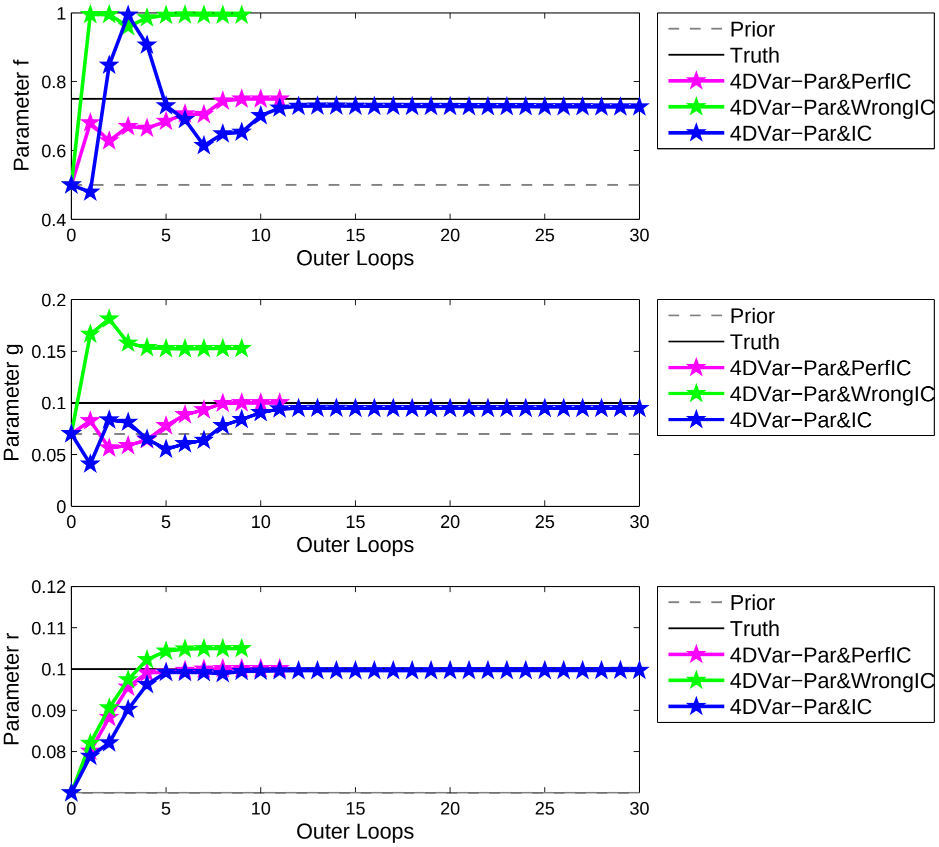

Figure 3: Parameter convergence shown within the outer loops for all three experiments together.

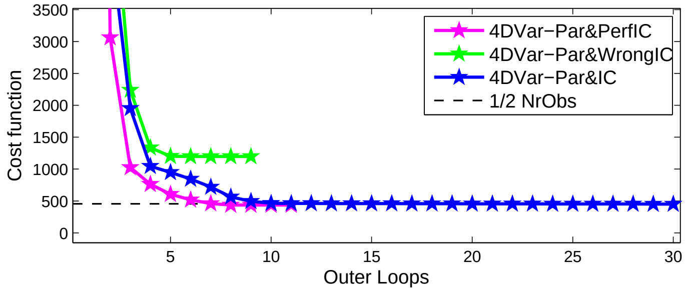

Figure 4: The cost function shown within the outer loops for all three experiments together (it was enlarged for clarity). The cost function of the experiment 4DVar-WrongPar&IC was too large, therefore it was removed for the clarity of the image.

perform quite well as shown by water column plots in Figure 5, and at the root mean square errors plots in Figure 9.

Although 30 outer loops are plotted for this experiment, the cost function stabilizes around its expected minimum much earlier. Already after the tenth outer

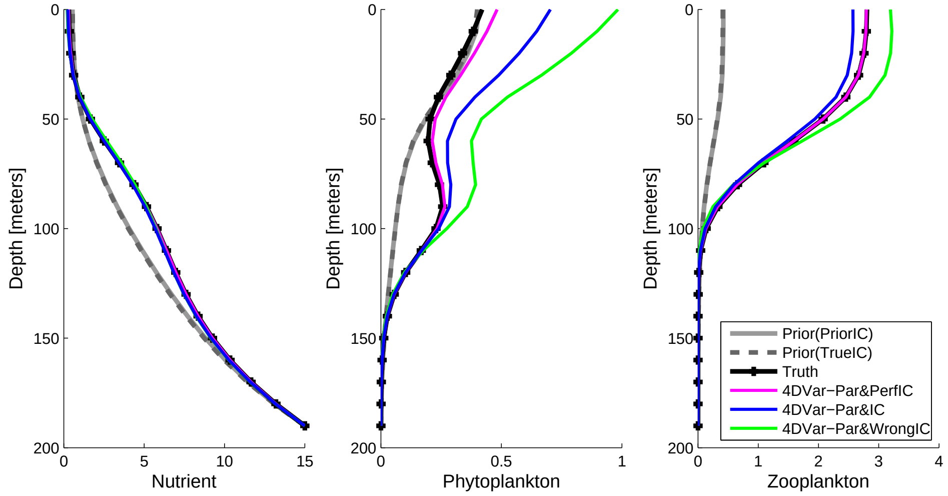

Figure 5: Water columns representing the concentrations of the state variables for each experiment. The results are presented at the end of the bloom of the second year cycle (day 545). In gray is the prior (no data assimilation) model output, in black is the truth, in magenta is the experiment 4DVar-Par&PerfIC, in blue is 4DVar-Par&IC and in green is 4DVar-Par&WrongIC.

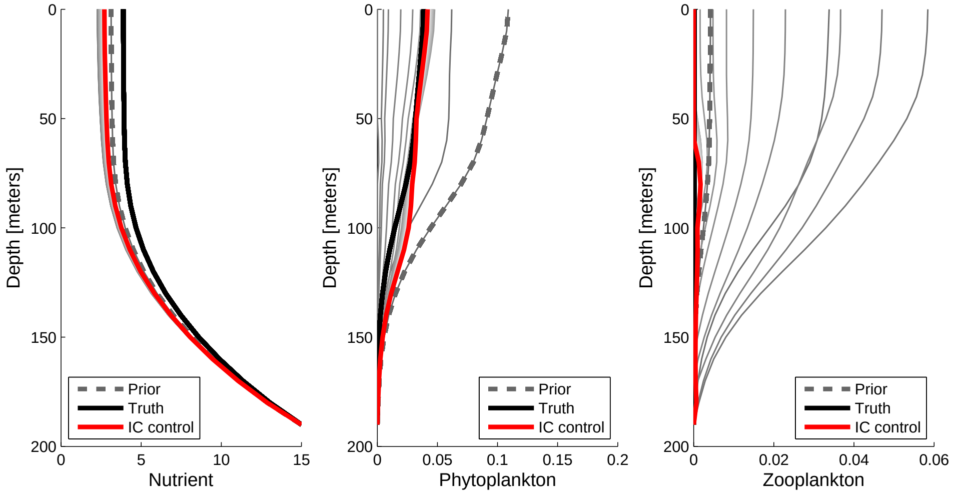

Figure 6: The performance of the initial condition estimation in the experiment 4DVar-Par&IC. The dashed gray line represents the prior initial condition, in black is the true one, the thin gray lines are the initial conditions resulting from the consecutive outer loops, the final outer loop corresponds to the 4D-Var estimation, and it is shown in red.

loop the cost function attains a value of 463.42, after which the cost function drop is not anymore that significant, see Figure 4. Thus, the outer loop minimization could have been terminated already around 13 - 17 outer loops, without loss in the control variable accuracy, since these also attained relatively rigid estimations already in the early stage of the minimization.

For a comparison the experiment 4DVar-Par&WrongIC was conducted where

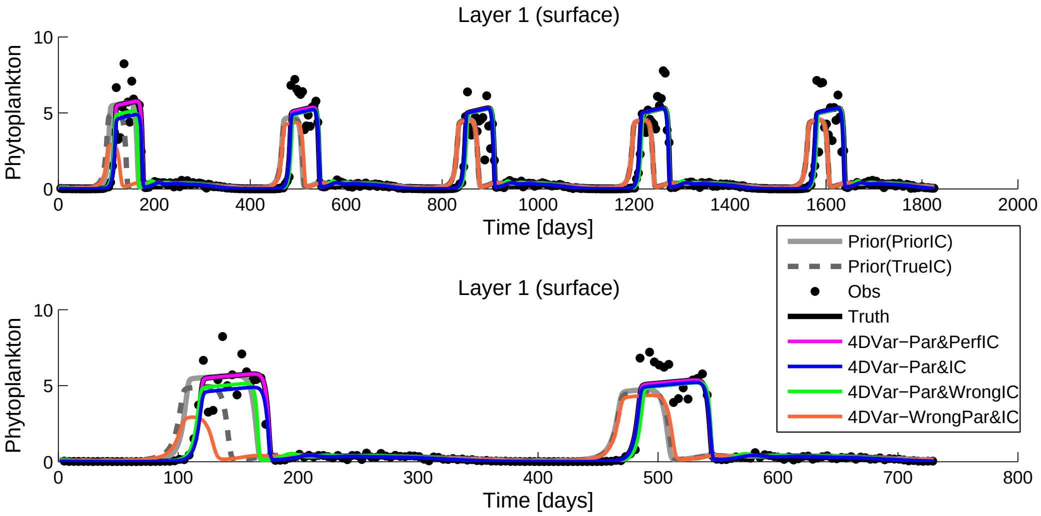

Figure 7: Phytoplankton concentration shown at the surface layer of the water column. The upper plot presents the whole five year assimilation window, and the lower plot magnifies the first two years for a better view.

only parameters were estimated within the wrong initial condition framework. Within 9 outer loops the cost function converged to its minimum. However, the parameters did not converge to their true values. Their percentage errors are 32.44%, 52.65%, and 4.94% respectively for f, g and r, whereas the starting percentage error was around 30% for all three parameters, as listed in Table 2. Only parameter r performs reasonably well, whereas the other two parameters obtain convergence at wrong values (see Figure 3). The incorrectly adjusted parameters compensate for the wrong initial condition to match the misfit between the model and the measurements, which results in quite good fit, as shown in Figure 7. However, the model performance for the unobserved areas, is much less accurate when compared to the experiment 4DVar-Par&IC. For reference see the water column plots in Figure 5, and where it is particularly visible, at the root mean square errors shown for the last assimilation year in Figure 9. It confirms that the control of the initial condition along with the parameters plays an important role in order to accurately calibrate the parameters. See Table 2 for the summary of the results from all three experiments.

Estimation of the initial condition alone with wrongly assigned parameters (4DVar-WrongPar&IC) failed, as shown by the root mean square error of the initial condition in Table 3, as well as indicated by very high cost function value shown in Table 2. Also the root mean square error of the initial condition, as shown in Table 3, illustrates that the method performs much worse when parameter estimation is excluded. That confirms the importance of the parameter calibration, particularly for experiments for which the assimilation window cover

RMSE Initial Condition	N	P	H
4DVar-Par&IC	1.4160	0.1310	0.2579
4DVar-WrongPar&IC	4.1400	0.5325	105.8000

Table 3: Comparison of the initial condition estimation for 4DVar-Par&IC and 4DVar-WrongPar&IC. RMSE stands here for the root mean square error of the initial condition estimation, and it has been normalized with respect to the prior RMSE.

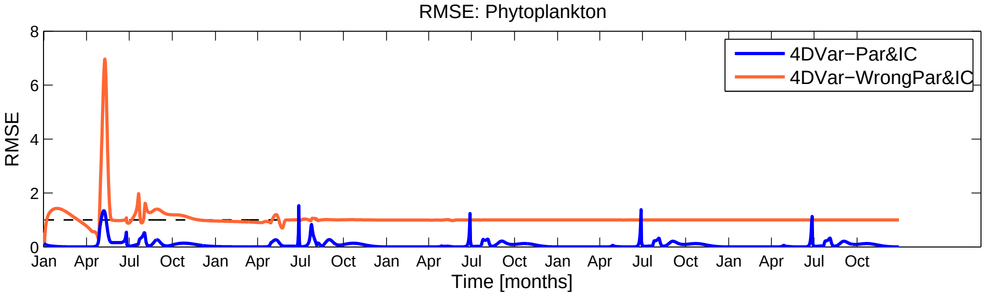

Figure 8: Root mean square error (RMSE) of Phytoplankton (depth averaged) normalized with respect to the prior RMSE. Two experiments are compared: 4DVar-Par&IC and 4DVar-WrongPar&IC. The dashed line indicates value equal to one, and it is plotted for a reference. The results are plotted as a time series for five years assimilation window.

a longer period of time. That can be explained by the limited memory of the initial condition. In this model the initial concentrations have an impact on the model simulation mainly during the first year, after that it is ”forgotten” and the model is governed only by the parameters. This phenomena can be well observed in Figure 7, as well as in Figure 8 where the model output of the experiment 4DVar-WrongPar&IC is merging together with the background model simulations already starting from the second year. Therefore, the initial condition is incapable of explaining the measurements assimilated during the last four years of the assimilation window. Hence, the method is unable to lower the cost function substantially in this case. That also results in wrong initial condition updates, which in consequence gives a very poor model output during the first year. Such experiment is expected to perform better for a shorter assimilation window (particularly in this case it would be one year). Due to very poor performance of this experiment, from now on the focus will be only on the remaining three experiments.

There are tendencies which are common for all three experiments. While fitting the model curve to the given phytoplankton data, the first feature to be fitted is the beginning of the bloom. Afterwards, the magnitude and the time duration of the event can be adjusted. In the considered case, the plant metabolic loss (r) is the only parameter responsible for the start of the bloom. Therefore, the parameter r is always the first parameter to converge. The other two parameters will converge, after the beginning of the bloom has been captured. In

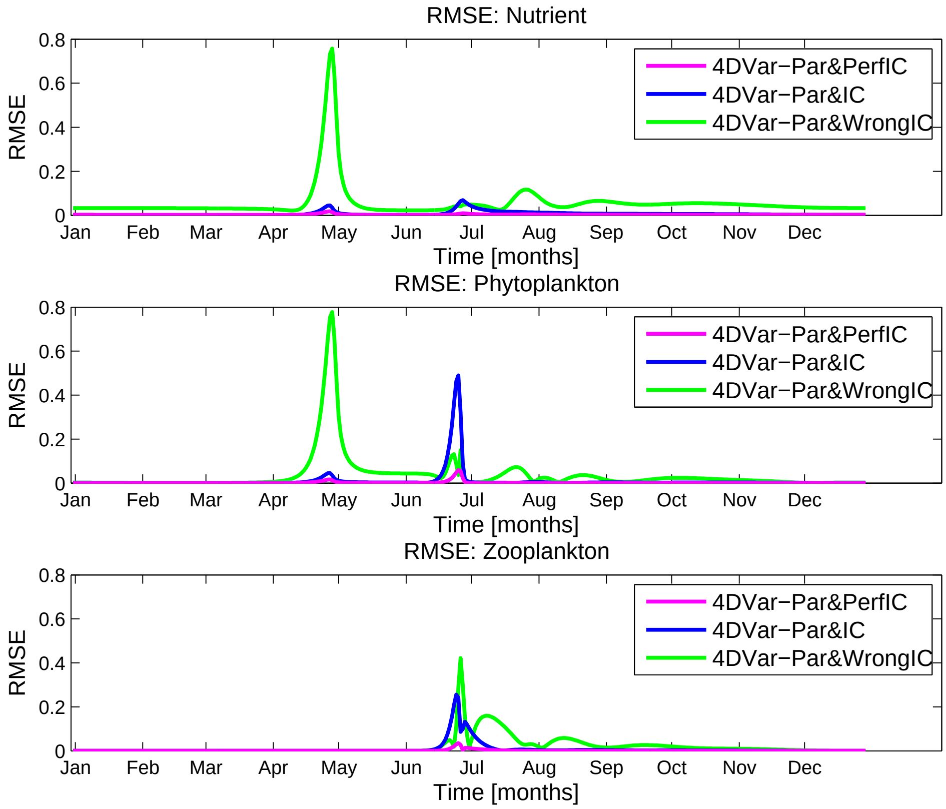

Figure 9: Root mean square error for each of the state variables plotted as a time series (depth averaged), shown for the last assimilation year, all three experiments together.

the given setup of the twin experiment the start of the bloom is quite well pronounced despite the measurement noise. Therefore, the parameter r achieves a good accuracy even though observations contain 30% of error. Furthermore, the parameters f and g influence mainly the magnitude and shape of the phytoplankton bloom, which is not easily detected from a noisy data. Moreover, these two parameters correspond to the Zooplankton concentration (see model equations (1)), therefore, they both influence indirectly the Phytoplankton concentration. Hence, as only the Phytoplankton data is given, the two parameters become conditionally dependent, and thus it is harder to estimate their two values in the right proportions, especially when very noisy measurements are used. Accounting for their dependence in the background error covariance matrix (29) could address this problem to improve the accuracy, as well as the speed of convergence of these two parameters.

Until now all experiments have been performed using 99% of information to build the reduced model. Additional experiments have been performed to show the performance of the method with different amount of information recovered, see Table 4. The twin experiments were performed using the following true

	Parameters			Cost	#OL		
	f	g	r	Function			
Prior	0.5	0.07	0.07	4.24e+5			
Truth	**0.7**	**0.1**	**0.1**				
4DVar-60%	0.4810	0.0717	0.1032	787.94	2		
4DVar-75%	0.5737	0.0776	0.0920	508.74	3		
4DVar-90%	0.4724	0.0700	0.0996	469.02	6		
4DVar-95%	0.6737	0.0934	0.0995	454.82	11		
4DVar-99%	0.6880	0.0962	0.0996	447.77	30		
Number of observations				2 × 456			
	RMSE Initial Condition			# Patterns (OL-average)			
	N	P	H	Total	N	P	H
4DVar-60%	1.01	0.32	0.04	**3**	1	1	1
4DVar-75%	1.02	0.37	4.49	**3.89**	1.67	1.22	1.00
4DVar-90%	1.37	0.19	3.63	**5.67**	2.42	1.92	1.33
4DVar-95%	2.02	0.10	0.25	**7.06**	3.00	2.06	2.00
4DVar-99%	1.72	0.16	0.35	**10.92**	4.00	3.16	3.76

Table 4: Summary of the experiments where different amount of information is recovered to build the reduced model. The number of patterns may be different for each outer loop, therefore, here it has been averaged over all outer loops. RMSE stands here for the root mean square error of the initial condition estimation, and it has been normalized with respect to the prior RMSE.

parameters: $f = 0.7$, $g = 0.1$ and $r = 0.1$. The *Perturbed Initial Condition Setup* was used, where the parameters were estimated along with the initial condition (4DVar-Par&IC).

As it was expected, when the amount of information decreases, then less outer loops are needed to attain convergence. Whereas the cost function consistently increases when the quality of the reduced model decreases, which indicates less accurate model match with the data. The results of the control variable estimation also encourages to use as high amount of information as affordable. Although there are some fluctuations along the general trend, the overall tendency is such that the more information is retrieved, the better they are calibrated. Similarly to the previous experiments, the most accurate estimations are shown for the parameter r. The remaining two parameters f and g are more difficult to estimate, which is again caused by their interfering effects in Equation (1c). Although their estimates maintain the tendency to move towards the correct values, there remains more fluctuations in their updates.

The estimation of the initial condition was presented in terms of the root mean square error normalized with respect to its corresponding prior error. This way values below one correspond to an improvement in the initial condition, and values above one indicate deteriorations. Here, similarly as in previous experiments, there is always a good performance for the estimation of the initial Phytoplankton concentration. To a certain extent, the initial Phytoplankton estimates are consistent with respect to the amount of the recovered information. However, the control of the other two values fluctuate more with respect to the quality of the reduced model. Such behavior is expected, since these two state variables are not observed, and they were already difficult to estimate when 99% information was recovered.

To illustrate the robustness of the method an experiment has been performed for which an ensemble of 15 different prior parameter sets has been randomly generated. Each of them was used as the prior set of parameters to initialize 15 different data assimilation experiments. Each twin experiment used the same reference solution, which was generated with the following true parameter set: $f = 0.7$, $g = 0.1$ and $r = 0.1$. The experiments were performed using the *Perturbed Initial Condition Setup*, where the parameters were estimated along with the initial condition (4DVar-Par&IC). The measurements were regenerated for each experiment in order to show robustness of the results with respect to the randomness in the observation noise.

	Parameters			Cost Function		#OL
	f	g	r	Prior	4D-Var	
Truth	**0.7**	**0.1**	**0.1**			
4DVar-1	0.5757	0.07503	0.09981	1.2223e+3	479.66	7
4DVar-2	0.6829	0.09207	0.09905	5.4997e+6	536.34	40
4DVar-3	0.6836	0.09591	0.09973	4.5837e+5	438.63	40
4DVar-4	0.6474	0.08820	0.09923	6.7982e+6	505.99	14
4DVar-5	0.6870	0.09634	0.09972	5.8290e+5	439.58	40
4DVar-6	0.6498	0.08755	0.09888	3.8433e+5	516.33	12
4DVar-7	0.6700	0.09405	0.09984	3.4332e+6	444.73	59
4DVar-8	0.6239	0.08491	0.09943	1.7548e+5	435.34	9
4DVar-9	0.6709	0.09387	0.10030	2.4162e+4	484.31	30
4DVar-10	0.6787	0.09097	0.09899	1.3607e+7	545.24	36
4DVar-11	0.6278	0.08531	0.09946	8.5052e+4	465.78	9
4DVar-12	0.7290	0.10290	0.10030	1.2229e+5	439.03	61
4DVar-13	0.6979	0.09976	0.10040	1.2038e+3	470.36	60
4DVar-14	0.7654	0.11150	0.10010	4.8201e+5	470.79	60
4DVar-15	0.7160	0.10060	0.09976	1.0553e+8	529.14	60
Number of observations					2 × 456	

Table 5: Summarized results from the ensemble of 15 experiments.

The prior parameters were drawn according to log-normal distribution with the mean equal to the background parameter values as listed in Table 1, and the error variance equal to 25% of the parameter value.

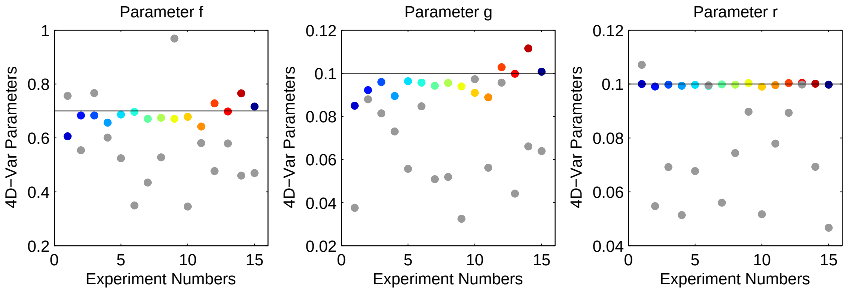

Figure 10: Parameter estimations shown for the ensemble of 15 experiments. The randomly generated background parameters are plotted in gray, the 4DVar estimates are shown in color (each experiment in a different color), the black line represents the true parameter values.

The Figure 10 shows the resulting estimations of the parameters for the ensemble of 15 twin experiments. Almost in every experiment the parameters converge to their true values, showing an improvement with respect to their prior parameter values. The experiment number 10 is an exception, where estimated value of parameter g is slightly worse compared to its background value. However, both the prior and the estimated values are still relatively close to the true parameter value. Similarly as in the previous experiments, the parameter r is much more accurately estimated, than the parameters f and g. This behavior is expected, and it was explained along with the previous experiments.

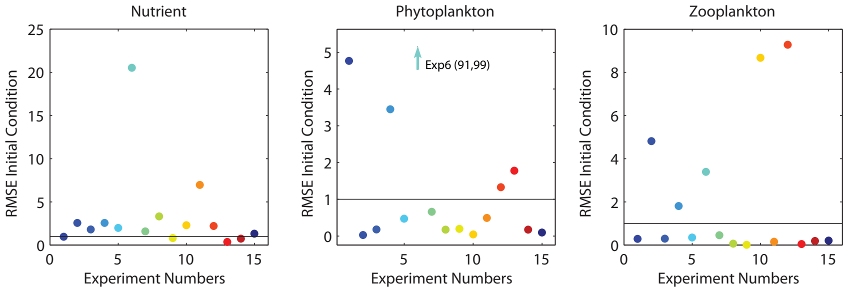

Figure 11: Root mean square error (RMSE) of the initial condition estimations normalized with respect to the prior RMSE in the initial condition. The results are shown for the ensemble of 15 experiments. Each color represents the resulting RMSE corresponding to a different experiment number. The black line is drawn at the value equal to one for reference.

The good performance of the parameter estimation does not always correspond to the well estimated initial conditions. Figure 11 shows the root mean square error (RMSE) obtained in the initial condition estimations. The RMSE was normalized with respect to the RMSE in the prior initial condition. Therefore, the RMSE value below one represents an improvement with respect to the prior initial condition, and the values above one represent the initial condition estimation which was worse than its prior. Similarly to previous experiments, the best performance is shown for the Phytoplankton initial concentrations. It is expected, since it is the only observed model variable. Also very good performance is shown for the estimation of the initial concentration of the Zooplankton. However, the initial condition estimates for the Nutrients are relatively bad, which is also consistent with the previous experiments.

The initial condition has the impact mainly on the first year of the assimilation window, whereas the remaining years are governed by the model parameters. Therefore, the effect of not very accurate initial condition estimations can be seen only within the first bloom cycle (see Figure 12). Relatively well calibrated parameters contribute to very good model match for the rest of the assimilation window. It is also confirmed by the final values of the cost functions for all 15 experiments, as summarized in the Table 5. Although some experiments reached 60 outer loops, for all of them 10-15 outer loops were sufficient for the cost function to converge.

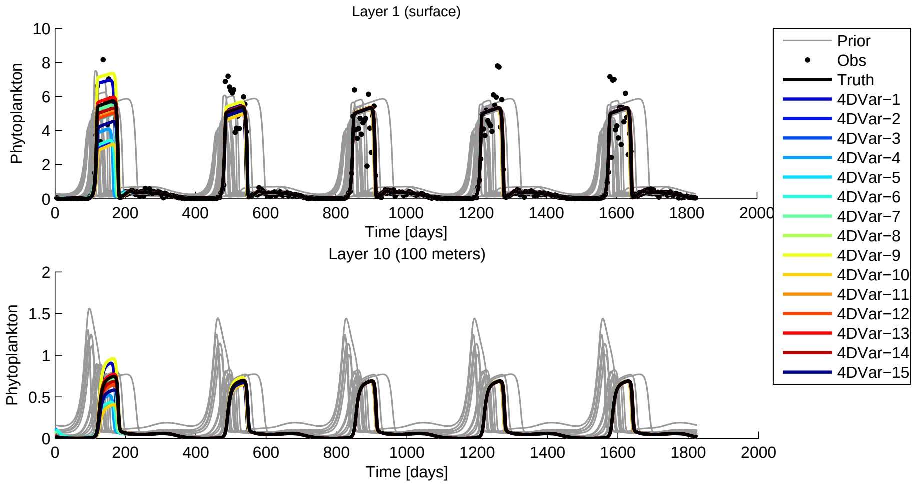

Figure 12: Phytoplankton concentration shown within the whole five year assimilation window. The upper plot presents the surface layer of the water column, and the lower plot presents the concentration at the 100 meters depth (the 10th layer). Results are shown for the ensemble of 15 experiments. The concentrations resulting from the data assimilation experiments are plotted in color. The prior concentrations are shown in gray. One arbitrarily chosen observation set is presented.

6. Conclusions

A model-reduced 4D-Var technique (Vermeulen and Heemink, 2006) was applied to a one dimensional ecological model (Eknes and Evensen, 2002). The system describes a water column with concentrations of nutrients, phytoplankton and zooplankton in time. Three ecosystem parameters were calibrated: grazing efficiency (f), loss to carnivores (g) and plant metabolic loss (r). The parameter estimation was combined with the initial condition estimation. The performance of the method was evaluated by means of twin experiments.

Four experiments were presented, one with parameter estimation only, where a perfect initial condition was used for data assimilation. In the second experiment, the initial condition was perturbed, and it was included in the control vector to be estimated along with the parameters. For a comparison, the third experiment was conducted, where the initial condition was also perturbed, however only the parameters were calibrated. Hence, it was performed within a wrong initial condition assumption. The goal of the fourth experiment was to estimate the initial condition in a framework with wrongly assigned parameters.

The first experiment has shown that the model-reduced 4D-Var works very well for the parameter calibration, where the initial condition is already known. Despite that the true parameters were strongly biased from their prior values, the calibration technique handled the problem with high accuracy. Also the measurement error provided extra difficulty since its standard deviation was assumed to be equivalent to 30% of the measurement value.

In the second experiment the parameter and initial condition estimation were combined. The same setup was used, except that the initial condition was perturbed, hence not perfect anymore. Therefore, the main challenge next to parameter estimation was the control of the initial condition, which in general performed quite well. Beside the initial Nutrient concentrations, very accurate estimations were shown for the other two initial concentrations. As expected the best result was shown for the control of the initial Phytoplankton, which is the only concentration for which measurements are available. There is a lot of interaction between all three state variables. Therefore, there is a potential to improve the control of their initial concentrations, particularly Nutrients. It could be done by accounting for their correlations in the background error covariance matrix $\mathbf{B}_{\mathbf{x}_0}$ (30), for example by generating it from an ensemble of model outputs. Furthermore, despite treating the substances independently in the projection matrix $\mathbf{P}$ (31) does not affect the accuracy in the approximation of the adjoint, it nevertheless could be beneficial for the initial condition estimation. The initial concentrations are estimated as the linear combination of the columns of $\mathbf{P}$, therefore extra information about how the state variables are correlated is expected to improve the initial condition update.

The relatively good performance of the initial condition control resulted in a very good accuracy of the estimated parameters. This could not be achieved without the initial condition control, which was confirmed by the third experiment, where the same setup was used, except that the initial condition was not calibrated. Wrongly assigned initial state values resulted in inaccurate parameter estimations, as well as larger root mean square errors for the concentrations at the unobserved areas.

The estimation of the initial condition with wrongly assigned parameters has shown to be unsuccessful. Due to the short memory of the initial condition, such experiments have more potential to work for shorter assimilation windows. On the other hand, this experiment illustrated the importance of the parameters, particularly for assimilation covering longer time periods.

The reduced model setup ensured its high accuracy, however, it is not necessary to always target at 99% of the information recovered. Experiments for different amounts of information recovered were performed to investigate this. It was shown that even 60% of the information can be sufficient to obtain satisfactory results. Moreover, decreasing the quality of the reduced model resulted in faster convergence of the cost function (less outer loops were needed). This shows that already rough model approximations are good enough to minimize the cost function effectively. Furthermore, due to reduction the model dynamics are simplified, which helps to avoid local minima.

For less nonlinear cases it may be possible to use the same reduced model for all outer loops. Then the model outputs would not have to be re-simulated for every outer loop, which would benefit in significantly less computational time required. Moreover, in this case the reduced model did not need snapshots covering the whole 5 year period. Since the original model consists of yearly cycles it would be sufficient to collect snapshots only from one year, preferably more frequently during the bloom period. This way also some of the computational time could be saved.

An advantage of estimating the initial condition in the reduced space is that it results in a smaller size of the control vector. Instead of calibrating the initial state in the full space, it is estimated as a linear combination of the columns of the projection matrix $\mathbf{P}$. Therefore, in this case instead of 60 values representing the initial condition, only about 11 need to be included as the control variables, which reduces the number of estimated values from 63 to 14, when taking into account also the three biological parameters.

To illustrate the robustness of the method an experiment has been performed for which an ensemble of 15 different prior parameters has been randomly generated. Each of them was used as the prior set of parameters to initialize 15 different data assimilation experiments. The measurements were regenerated for each experiment in order to also show robustness of the results with respect to the randomness in the observation noise. The ensemble of 15 experiments has shown consistent results, which proves the robustness of the method.

The model reduced 4D-Var technique performed well in the simple 1D ecosystem model, hence there is a potential in the method for real case ecological applications. However, more advanced tools might be needed to assure good performance of the technique. Such as a proper transformations to constrain the parameters within their domains, anamorphosis or log-transformations for the model states (Campbell, 1995; Simon and Bertino, 2009), as well as more sophisticated background covariance matrix also might be necessary. Moreover, tackling the problem for larger systems will need a lot more patterns even to recover just 60% of information. This will require to plan more carefully how to simulate and select the model snapshots. Larger applications also will result in more parameters which need to be estimated. Already in this small 1D ecosystem model it was noticed that some parameters could not always be uniquely defined given only the Phytoplankton observations. As a consequence, there might be more than one parameter set resulting in a satisfactory value of the cost function. Estimating more parameters in ecological models is likely to give similar issues, as it was already experienced in Ward et al. (2010).

Although the system used to investigate the method is small, it is representative of yearly cycles of phytoplankton bloom. It allows to face important issues when applying data assimilation methods in ecosystem models, such as the nonlinearity of the model, the strong impact of parameters on the model dynamics, the positiveness of the variables, as well as the relatively sparse observations with large error. It's a good framework for a first assessment of methodological developments in data assimilation before integration in real complex configurations. Based on the relatively good results obtained in this 1D configuration, the next step is to apply the model reduced 4D-Var method and assess its performances in realistic 3D configurations.

Acknowledgments The authors wish to thank the editor A. Oschlies and the two anonymous referees for their helpful and constructive comments. A grant of CPU time from the Norwegian Supercomputing Project (NOTUR) has been used.

Appendix A. Construction of projection matrix P

The matrix **P** is expected to maintain the most important dynamics of the model. It is needed to project the linearized model equations (20). Therefore, the reduced space corresponding to **P** should span a subspace of all the possible model sensitivities. To this end, an ensemble of possible model changes is generated by perturbing the parameters of interest; the same ones around which the model is linearized. Each of the control variables is perturbed separately, and next the corresponding model outputs are collected at different time locations. The more snapshots are taken, the more details are captured.

The snapshots are centered and normalized, and next collected in a matrix denoted as $\mathbf{\Delta X}$

$$\mathbf{\Delta X} = [\mathbf{e}_1(t_1), ..., \mathbf{e}_1(t_{n_s}), \cdots, \mathbf{e}_{n_{\boldsymbol{\alpha}}}(t_1), ..., \mathbf{e}_{n_{\boldsymbol{\alpha}}}(t_{n_s})] \tag{A.1}$$

where $t_1, t_2, ..., t_{n_s}$ are selected times at which snapshots are taken, with n_s defining the number of snapshots, $n_{\boldsymbol{\alpha}}$ is the number of estimated parameters, and $\mathbf{e}_k(t_j)$ is a normalized model change for each parameter perturbation $k = 1, 2, .., n_{\boldsymbol{\alpha}}$ and for each snapshot $j = 1, 2, .., n_s$, i.e.

$$\mathbf{e}_k(t_j) = \frac{\mathcal{M}_j(\mathbf{x}^b_{j-1}, \boldsymbol{\alpha}^b + \Delta\boldsymbol{\alpha}_k) - \mathcal{M}_j(\mathbf{x}^b_{j-1}, \boldsymbol{\alpha}^b)}{\|\mathcal{M}_j(\mathbf{x}^b_{j-1}, \boldsymbol{\alpha}^b + \Delta\boldsymbol{\alpha}_k) - \mathcal{M}_j(\mathbf{x}^b_{j-1}, \boldsymbol{\alpha}^b)\|} \tag{A.2}$$

With such constructed snapshots, the collection $\mathbf{\Delta X}$ is an ensemble of increments $\delta \mathbf{x}_i$. The size of $\mathbf{\Delta X}$ is $n_{\mathbf{x}} \times n_{\mathbf{e}}$, where $n_{\mathbf{e}}$ is a total number of ensembles, i.e. $n_{\mathbf{e}} = n_{\boldsymbol{\alpha}}\, n_s$. Next the ensemble is used to create a covariance matrix $\mathbf{C_X}$

$$\mathbf{C_X} = \frac{1}{n_{\mathbf{e}} - 1} \mathbf{\Delta X \Delta X}^T \tag{A.3}$$

The covariance matrix captures the variability of the model sensitivities with respect to the change of the parameters $\boldsymbol{\alpha}$. The large values in the diagonal terms correspond to interesting dynamics (high variance). The large magnitudes in the off-diagonal terms correspond to high redundancy (high correlations). Thus, ideal would be to find a projection matrix $\mathbf{P}$ such that the covariance of the projected state $\mathbf{\Delta Z} = \mathbf{P \Delta X}$ is a diagonal matrix. Hence, a projection matrix $\mathbf{P}$ is needed such that the following covariance matrix is diagonal

$$\mathbf{C_Z} = \frac{1}{n_{\mathbf{e}} - 1} \mathbf{\Delta Z \Delta Z}^T \tag{A.4}$$

where $\mathbf{C_Z}$ is the covariance matrix of the state in the reduced space. Notice that the covariance matrix $\mathbf{C_Z}$ is related with $\mathbf{C_X}$ as follows

$$\mathbf{C_Z} = \frac{1}{n_{\mathbf{e}} - 1} \mathbf{P \Delta X} (\mathbf{P \Delta X})^T = \mathbf{P} (\frac{1}{n_{\mathbf{e}} - 1} \mathbf{\Delta X \Delta X}^T) \mathbf{P}^T = \mathbf{P C_X P}^T$$

Since the construction of a covariance matrix assures it to be always symmetric, it can be diagonalized by an orthogonal matrix of its eigenvectors. Hence $\mathbf{C_X} = \mathbf{E D E}^T$, where $\mathbf{D}$ is diagonal matrix of eigenvalues of $\mathbf{C_X}$, and columns of matrix $\mathbf{E}$ are the eigenvectors. The next step is to choose $\mathbf{P} = \mathbf{E}^T$, then

$$\mathbf{C_Z} = \mathbf{P C_X P}^T = \mathbf{P}(\mathbf{E D E}^T)\mathbf{P}^T = (\mathbf{P P}^T)\mathbf{D}(\mathbf{P P}^T) = (\mathbf{P P}^{-1})\mathbf{D}(\mathbf{P P}^{-1}) = \mathbf{D} \tag{A.5}$$

Hence by choosing the eigenvectors of $\mathbf{C_X}$ as the columns of the projection matrix $\mathbf{P}$, the covariance matrix of the reduced model state $\mathbf{C_Z}$ is equal to the diagonal matrix of eigenvalues $\mathbf{D}$. The eigenvectors of $\mathbf{C_X}$ are the principal components of $\mathbf{\Delta X}$.

In order to get the eigenvectors of $\mathbf{C_X}$ solving a reduced eigenvalue problem is much more efficient, since the size of $\mathbf{\Delta X}^T \mathbf{\Delta X}$ is $n_{\mathbf{e}} \times n_{\mathbf{e}}$, whereas the size of $\mathbf{\Delta X \Delta X}^T$ is $n_{\mathbf{x}} \times n_{\mathbf{x}}$, and in most of the cases $n_{\mathbf{e}} \ll n_{\mathbf{x}}$. Thus

$$\mathbf{\Delta X}^T \mathbf{\Delta X} \mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{A.6}$$

where λ_i are the eigenvalues of $\mathbf{\Delta X}^T \mathbf{\Delta X}$ and $\mathbf{v}_i$ are its eigenvectors. After the eigenvectors $\mathbf{v}_i$ are established, the eigenvectors of $\mathbf{\Delta X \Delta X}^T$ are obtained as follows

$$\mathbf{p}_i = \frac{1}{\sigma_i} \mathbf{\Delta X} \mathbf{v}_i \tag{A.7}$$

where $\sigma_i = \sqrt{\lambda_i}$ are the singular values of $\mathbf{\Delta X}$. Notice that, since $\|\mathbf{\Delta X v}_i\| = \sigma_i$, then $\|\mathbf{p}_i\| = 1$. Hence, the vectors $\mathbf{p}_i$ are not only orthogonal, but also orthonormal, which is essential for evaluation of the directional derivatives (19).

In order to construct the projection matrix $\mathbf{P}$, the leading principal components $\mathbf{p}_i$ of $\mathbf{\Delta X}$ have to be selected. To this end the vectors $\mathbf{p}_i$ are chosen such that, the variance of $\mathbf{\Delta X}$ along that direction is maximized. Notice from the calculations (A.5), that the ith diagonal value of $\mathbf{D}$ is the variance of $\mathbf{\Delta X}$ along direction $\mathbf{p}_i$. Hence the values of the eigenvalues λ_i represent the magnitude of spread of $\mathbf{\Delta X}$ in the direction of eigenvector $\mathbf{p}_i$. In other words the eigenvalues λ_i correspond to the variance of the model in the reduced space. Therefore, choosing the eigenvectors corresponding to the largest eigenvalues, will create an orthonormal basis, which will map the full model space into space spanning the most important dynamics of the system. For more details about principal component analysis see Shlens (2009).

The more eigenvectors are added to the basis, the closer the reduced space gets to the original one. There is a measure to evaluate the percentage of the information carried by the reduced space

$$\mathcal{I} = \frac{\sum_{i=1}^{n_{\mathbf{P}}} \lambda_i}{\sum_i \lambda_i} \tag{A.8}$$

where $n_{\mathbf{P}}$ is the number of selected eigenvectors, and $\sum_i \lambda_i$ is the total sum of all eigenvalues. The number $n_{\mathbf{P}}$ is chosen according to how much information $\mathcal{I}$ needs to be recovered. Usually a relatively small number of vectors $\mathbf{p}_i$ is enough to represent over 90% of information. The $n_{\mathbf{P}}$ selected principal components $\mathbf{p}_i$ arranged in a matrix as columns, create the projection matrix $\mathbf{P}$. The principal components $\mathbf{p}_i$ are often referred as *patterns*.

References

Allen, J. I., Eknes, M., , Evensen, G., 2002. An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. Annales Geophysicae 20, 1–13.

Altaf, M. U., Heemink, A. W., Verlaan, M., 2009. Inverse shallow-water flow modeling using model reduction. International Journal for Multiscale Computational Engineering 7, 577–594.

Altaf, M. U., Heemink, A. W., Verlaan, M., 2010. Model-reduced variational data assimilation for shallow water flow modeling. V European Conference on Computational Fluid Dynamics, ECCOMAS CFD 2010, J. C. F. Pereira and A. Sequeira (Eds), Lisbon, Portugal, 14-17 June 2010.

Antoulas, A. C., 2005. Approximation of Large-Scale Dynamical Systems. SIAM, Philadelphia.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential Data Assimilation Techniques in Oceanography. International Statistical Review 71, 223–241.

Brush, G. S., 2008. Historical land use, nitrogen, and coastal eutrophication: A paleoecological perspective. Estuaries and Coasts 32, 18–28.

Campbell, J. W., 1995. The lognormal distribution as a model for bio-optical variability in the sea. J. Geophys. Res. Oceans 100, 13 237–13 254.

Cao, Y., Zhu, J., Navon, I., Luo, Z., 2007. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. International Journal for Numerical Methods in Fluids 53 (10), 1571–1583.

Carmillet, V., Brankart, J.-M., Brasseur, P., Drange, H., Evensen, G., 2001. A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical - biochemical model of the North Atlantic. Ocean Modeling 3, 167–192.

Courtier, P., Paw, J.-N., Hollingsworth, A., 1994. A strategy for operational implementation of 4d-var, using an incremental approach. Q.J.R. Meteorol. Soc. 120, 1367–1387.

Doney, S. C., 1999. Major challenges confronting marine biogeochemical modeling. Global Biogeochemical Cycles 13, 705–714.

Doron, M., Brasseur, P., Brankart, J.-M., 2011. Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model: Twin experiments. Journal of Marine Systems 87, 194–207.

Eknes, M., Evensen, G., 2002. An Ensemble Kalman filter with 1D- marine ecosystem model. Journal of Marine Systems 36, 75–100.

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

Fang, F., Pain, C., Navon, I., Piggott, M., Gorman, G., Farrell, P. E., Allison, P., Goddard, A., 2009. A POD reduced order 4D-Var adaptive mesh ocean modeling approach. International Journal for Numerical Methods in Fluids 60 (7), 709–732.

Fennel, K., Losch, M., Schröter, J., Wenzel, M., 2001. Testing a marine ecosystem model: sensitivity analysis and parameter optimization. Journal of Marine Systems 28, 45–63.

Fiechter, J., Broquet, G., Moore, A. M., Arango, H. G., 2011. A data assimilative, coupled physical-biological model for the Coastal Gulf of Alaska. Dynamics of Atmospheres and Oceans 52, 95–118.

Friedrichs, M. A., 2002. Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. Deep-Sea Research II 49, 289–31.

Garcia, I. D., El Serafy, G., Heemink, A., Schuttelaars, H., submitted. Bathymetric data assimilation for wave properties estimation: Towards a comprehensive morphodynamic data assimilation system. Ocean Dynamics.

Gilbert, J. C., Lemaréchal, C., 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. Mathematical Programming 45, 407–435.

Gregg, W. W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. Journal of Marine Systems 69, 205–225.

Gregg, W. W., Casey, N. W., 2004. Global and regional evaluation of the SeaWiFS chlorophyll data set. Remote Sensing of Environment 93, 463–479.

Gregg, W. W., Casey, N. W., McClain, C. R., 2005. Recent trends in global ocean chlorophyll. Geophysical Research Letters 32, (L03606)1–5.

Gregg, W. W., Conkright, M. E., Ginoux, P., O´Reilly, J. E., Casey, N. W., 2003. Ocean primary production and climate: Global decadal changes. Geophysical Research Letters 30, (3)1–4.

Hallegraeff, G. M., 1993. A review of harmful algal blooms and their apparent global increase. Phycologia 32 (2), 79–99.

Hallegraeff, G. M., 2009. Impacts of climate change on the harmful algal blooms.

Hallegraeff, G. M., 2010. Ocean climate change, phytoplankton community responses, and harmful algal blooms: A formidable predictive challenge. Jurnal of Phycology 46, 220–235.

Jørgensen, S. E., 2008. Overview of the model types available for development of ecological models. Ecological Modelling 215, 3–9.

Kaleta, M. P., 2011. Model-reduced gradient-based history matching. Ph.D. thesis, Delft University of Technology.

Kaleta, M. P., Hanea, R. G., Heemink, A. W., Jansen, J. D., 2011. Model-reduced gradient-based history matching. Computational Geosciences 15, 135–153.

Le Dimet, F.-X., Talagrand, O., 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus 38A, 97–110.

Lermusiaux, P. F., 2006. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. Journal of Computational Physics 217, 176–199.

Los, H., 2009. Eco-hydrodynamic modeling of primary production in coastal waters and lakes using BLOOM. Ph.D. thesis, Wageningen University.

Matear, R. J., 1995. Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P. Journal of Marine Research 53, 571–607.

Natvik, L. J., Eknes, M., Evensen, G., 2001. A weak constraint inverse for a zero-dimensional marine ecosystem model. Journal of Marine Systems 28, 19–44.

Natvik, L. J., Evensen, G., 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic Part 1. Data assimilation experiments. Journal of Marine Systems 40–41, 127–153.

Nerger, L., Gregg, W., 2007. Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. Journal of Marine Systems 68, 237–254.

Nixon, S. W., 1995. Coastal marine eutrophication: A definition, social causes and future concerns. Ophelia 41, 199–219.

Paerl, H. W., Huisman, J., 2008. Blooms like it hot. Science 320, 57–58.

Pauly, D., Christensen, V., Gunette, S., Pitcher, T., Sumaila, U., Walters, C., Watson, R., Zeller, D., 2002. Toward sustainability in world fisheries. Nature 418, 689–695.

Pearson, K., 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2(6), 559–572.

Peperzak, L., 2003. Climate change and harmful algal blooms in the North Sea. Acta Oecologica 24, 139–144.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian Processes for Machine Learning. The MIT Press, 2006. ISBN 0-262-18253-X.

Schindler, D. W., Hecky, R. E., Findlay, D. L., Stainton, M. P., Parker, B. R., Paterson, M. J., Beaty, K. G., Lyng, M., Kasian, S. E. M., 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. In: Proceedings of the National Academy of Sciences of the United States of America. Vol. 105. pp. 11254–11258.

Shlens, J., 2009. A Tutorial on Principal Component Analysis. Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Insitute for Biological Studies La Jolla, CA 92037, Dated: April 22, 2009; Version 3.01.

Simon, E., Bertino, L., 2009. Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment. Ocean Science 5, 495–510.

Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the DEnKF for combined state and parameter estimation: application to a 1D ocean ecosystem model. Journal of Marines Systems 89, 1–18.

Talagrand, O., Courtier, P., 1987. Variational assimilation of meteorological observations with adjoint vorticity equation. Part I. Theory. Q.J.R. Meteorol. Soc. 113, 1311–1328.

Tarantola, A., 1987. Inverse problem theory: Methods for data fitting and model parameter estimation. Elsevier Science Pub. Co. Inc., New York, NY.

Vermeulen, P. T. M., Heemink, A. W., 2006. Model-Reduced Variational Data Assimilation. Monthly Weather Review 134, 2888.

Vermeulen, P. T. M., Heemink, A. W., Valstar, J. R., 2005. Inverse modeling of groundwater flow using model reduction. Water Resources Research 41.

Ward, B. A., Friedrichs, M. A., Anderson, T. R., Oschlies, A., 2010. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. Journal of Marine Systems 81, 34–43.

Zhao, L., Wei, H., Xub, Y., Feng, S., 2005. An adjoint data assimilation approach for estimating parameters in a three-dimensional ecosystem model. Ecological Modelling 186, 234–249.