



ELSEVIER

Ocean Modelling 3 (2001) 167–192

**Ocean
Modelling**

www.elsevier.com/locate/omodo

A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the North Atlantic ocean

V. Carmillet ^a, J.-M. Brankart ^a, P. Brasseur ^{a,*}, H. Drange ^b, G. Evensen ^b,
J. Verron ^a

^a *Laboratoire des Ecoulements, Géophysiques et Industriels, UMR 5519 CNRS, BP 53 X, F-38041 Grenoble, France*

^b *Nansen Environmental and Remote Sensing Center, Bergen, Norway*

Abstract

Within the European DIADEM project, a data assimilation system for coupled ocean circulation and marine ecosystem models has been implemented for the North Atlantic and the Nordic Seas. One objective of this project is to demonstrate the relevance of sophisticated methods to assimilate satellite data such as altimetry, surface temperature and ocean color, into realistic ocean models. In this paper, the singular evolutive extended Kalman (SEEK) filter, which is an advanced assimilation scheme where three-dimensional, multivariate error statistics are taken into account, is used to assimilate ocean color data into the biological component of the coupled system. The marine ecosystem model, derived from the FDM model [J. Mar. Res. 48 (1990) 591], includes 11 nitrogen and carbon compartments and describes the synthesis of organic matter in the euphotic zone, its consumption by animals of upper trophic levels, and the recycling of detritic material in the deep ocean. The circulation model coupled to the ecosystem is the Miami isopycnic coordinate ocean model (MICOM), which covers the Atlantic and the Arctic Oceans with an enhanced resolution in the North Atlantic basin. The model is forced with realistic ECMWF ocean/atmosphere fluxes, which permits to resolve the seasonal variability of the circulation and mixed layer properties. In the twin assimilation experiments reported here, the predictions of the coupled model are corrected every 10 days using pseudo-measurements of surface phytoplankton as a substitute to chlorophyll concentrations measured from space. The diagnostics of these experiments indicate that the assimilation is feasible with a reduced-order Kalman filter of small rank (of order 10) as long as a sufficiently good identification of the error structure is available. In addition, the control of non-observed quantities such as zooplankton and nitrate concentrations is made possible, owing to the multivariate nature of the analysis scheme. However, a too severe truncation of the error sub-space downgrades the propagation of surface information below the mixed layer. The reduction of the actual state vector to the

* Corresponding author. Tel.: +33-4-7682-5072; fax: +33-4-7682-5271.

E-mail address: pierre.brasseur@hmg.inpg.fr (P. Brasseur).

surface layers is therefore investigated to improve the estimation process in the perspective of sea-viewing wide field-of-view sensor (SeaWiFS) data assimilation experiments. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Coupled models; Kalman filter; Ocean color; North Atlantic

1. Introduction

Three-dimensional models of the marine ecosystems are being used with increasing interest to better understand the space-time distribution of the primary biological properties in the ocean. The basic phenomenology described by these models is the synthesis of organic matter from inorganic materials by autotrophic organisms (i.e., primary production), its consumption by animals of the lower trophic levels, and the recycling of detritic material. The coupling with realistic models of the ocean circulation allows one to take leading factors of the physical environment into account, such as horizontal and vertical mixing processes, upwelling, horizontal advection, and thermodynamic conditions.

The development of high-resolution marine ecosystem models is justified by the need to capture the variability of biological activities at scales which could not be resolved by conventional in situ measurements, and also to predict the evolution of the marine environment at short or medium range. Besides coastal applications, basin-scale ecosystem models are useful to investigate, among others, the role of the phytoplankton in the control of atmospheric carbon dioxide via oceanic uptake in a global change perspective (Sarmiento et al., 1993; Oschlies and Garcon, 1998; Drange, 1996).

The response of a numerical model depends on a large variety of factors: the parameterization of the biological processes, the specification of external forcing functions and the associated hydrodynamic constraints, the discretization of the numerical system in space, the coupling mechanisms between the physical and the biological processes which are sometimes resolved with different time steps, etc. As a result of the many degrees of freedom in the model setup, errors can occur in the numerical simulations which lead to imperfect representations of the reality, and consequently mismatches with observations.

The observation of some facets of the marine ecosystem at the basin scale is now possible with ocean color measurements, which can be made available from satellite missions. For most oceanic regions, the color of the sea is determined primarily by the abundance of phytoplankton and their associated pigments (e.g., chlorophyll *a*): as the phytoplankton concentration increases, ocean color shifts from blue to green. The first space instrument used to measure ocean color was the coastal ocean color scanner (CZCS) operated by NASA from 2 November 1978 to 22 June 1986, with the objective to obtain a better understanding of the temporal and spatial distribution of phytoplankton biomass and primary production. The sea-viewing wide field-of-view sensor (SeaWiFS) project was then developed by NASA as a follow-on sensor to the CZCS; it has delivered ocean color data continuously since July 1997.

The comparison of a model response with those data provides a first opportunity for assessment, validation, and more generally for model refinement. A next step is the assimilation of these data with the aim to improve the consistency between observations and model simulations, to

dynamically extrapolate and interpolate the data in space and time, and to make comprehensive interpretation of ocean color measurements from satellites.

The theoretical framework of data assimilation in meteorology and physical oceanography is now well established: variational methods seek to minimize the misfit between data and model simulations by optimization of well-chosen control variables (numerical parameters, initial conditions, forcings, ...) while sequential methods proceed by intermittent blending of observations and model solutions according to their respective accuracy. In spite of their theoretical equivalence in a linear framework, significant differences exist between sequential and variational methods concerning implementation issues and algorithmic solutions.

Advanced algorithms have been developed recently to tackle the oceanic assimilation problems specifically. Reduced-order Kalman filters, for instance, have been developed and applied to academic ocean models, exploring how simplified representations of the estimation error statistics can reduce the computational burden of the conventional Kalman filter while preserving the optimal nature of the method (e.g., Fukumori and Malanotte-Rizzoli, 1995; Cane et al., 1996; Pham et al., 1998). Similarly, ensemble Kalman filters have been proposed, which represent the error statistics by simulating a limited set of model trajectories simultaneously (Evensen, 1994; Burgers et al., 1998).

By contrast, data assimilation into coupled systems is still in its infancy. A review of methods suitable for biogeochemical models can be found in Kasibhatla et al. (2000). A popular approach to assimilate data into marine ecosystem models is by determining poorly known parameters in the biological process equations. Inverse techniques have been applied, quite often in a one-dimensional context, to optimize their values according to local observations, and improve by this way the consistency between the simulated and the observed quantities (e.g., Fasham and Evans, 1995; Matear, 1995; Hurtt and Armstrong, 1996; Prunet et al., 1996). Due to the scarcity of the data, a small number of control variables are optimized in most cases.

These works are all relying on the same hypothesis that a coupled model with a unique parameter set can reproduce the observations, neglecting the possibility of other error sources. Gunson et al. (1999) assimilate ocean color data with a coupled physical–biological model of the North Atlantic ocean, and analyze the sensitivity of the different parameters using twin experiments. Fennel et al. (2000) apply a similar optimization procedure to real measurements of nitrate and chlorophyll at the Bermuda station. Their results indicate that the strong sensitivity of the method can be explained in part by unresolved processes of the model formulation. In principle, it is possible to account for errors in the model formulation with variational algorithms based on a weak constraint or generalized inverse formalism (Natvik et al., 2001). However, the minimization of the cost function of a full 3-D application may give considerable problems with convergence and computer capacity, and sequential algorithms are therefore of interest to assimilate data with imperfect models.

Armstrong et al. (1995) employ the simple nudging method to assimilate monthly averaged CZCS chlorophyll estimates in a circulation model of the Atlantic coupled with the FDM (Fasham et al., 1990) ecosystem model. In spite of its sub-optimal nature, their scheme is able to adjust the phytoplankton concentration over the euphotic zone with some success, demonstrating the interest of an approach which aims at compensating for a variety of possible error sources at once.

Within the EC MAST-III DIADEM project, a data assimilation system for coupled ocean circulation and marine ecosystem model is being implemented for the North Atlantic and the

Nordic Seas. Several novel and advanced assimilation schemes, where proper error statistics is taken into account, are being examined and refined. These involve the ensemble Kalman filter (EnKF) by Evensen (1994), the ensemble Kalman smoother (EnKS) by Evensen and van Leeuwen (2000), and the singular evolutive extended Kalman (SEEK) filter by Pham et al. (1998). The circulation model is the Miami isopycnic coordinate ocean model (MICOM) by Bleck et al. (1989), which has been coupled to the FDM marine ecosystem model by Drange (1994). The DIADEM project is a first attempt to examine these state-of-the-art assimilation methods with the coupled model system in a realistic context, and to lay a foundation for future operational assimilation systems for the North Atlantic and Nordic Seas. This particular paper presents the first results of the SEEK filter implementation to assimilate ocean color data in preparation to those collected from the NASA SeaWiFS program. Focusing on the biological segment of the assimilation system, several key ideas are addressed in this work.

First, the feasibility of a sequential, reduced-order updating scheme is investigated, with the underlying assumption that the mismatch between model predictions and observations can be caused by the combination of several error sources simultaneously (e.g., errors in initial conditions, missing processes, inappropriate forcing functions, inaccurate parameterizations, poorly resolved hydrodynamic constraints, etc.). The reduction of the control dimension is achieved in a multivariate error sub-space describing the dominant modes of the system variability.

Second, the capacity of an observing system based on surface chlorophyll measurements only, to control other biological compartments through a multivariate assimilation scheme, is examined. Such observing systems are fairly novel at this time and will be developed in the coming years. Assessments in terms of observability and controllability properties are thus needed to get practical experience and retrieve a maximum benefit from their use.

Finally, sensitivity assimilation experiments are performed to determine the best possible implementation options of the SEEK filter in the DIADEM configurations, in the perspective of real data assimilation experiments and pre-operational exercises. The definition of a suitable multivariate state vector is a particularly delicate issue with the present system, as two distinct ecosystem formulations are present in the coupled model: the FDM dynamical model in the upper, euphotic layers, and the diagnostic regeneration model proposed by Sarmiento et al. (1993) in the deep ocean. A twin experiment approach is considered as a first step to tackle the main difficulties of the sequential assimilation problem. These simplified experiments, however, will not allow us to be conclusive about the capacity of the method to compensate for errors in the model formulation, and the effectiveness of the scheme to assimilate real data.

The paper is organized as follows: in Section 2, we describe the physical and ecosystem sectors of the coupled system as it is implemented in the DIADEM configurations; the basic elements of the SEEK assimilation scheme are discussed in Section 3; the protocol of the twin assimilation experiments set up to validate the concept is given in Section 4; the analysis of these experiments is reported in Section 5, and a final discussion concludes the paper with Section 6.

2. Model description

The numerical code used in the assimilation experiments described in Section 5 is an Atlantic version of the MICOM coupled to the FDM ecosystem model of the seasonal nitrogen cycle

derived from Fasham et al. (1990) and further modified by Drange (1994) to include a description of the carbonic acid system in sea water. We first provide an overview of the model setup, with some emphasis on the novelties introduced in the DIADEM implementations.

2.1. *The ocean circulation model*

The circulation model is based on MICOM, the isopycnic coordinate ocean circulation model developed and described by Bleck and Boudra (1986), Bleck et al. (1990), Smith et al. (1990) and Bleck et al. (1992). MICOM is designed using advanced numerical techniques and advection routines, and can be run with arbitrary bottom topography and fairly low diffusion. It handles stratification over steep bottom topography, a feature which poses problems in conventional level models such as the GCM model employed by Sarmiento et al. (1993) or in models based on sigma coordinates in the vertical. In this particular version, the vertical discretization is achieved on 17 isopycnic layers. A mixed layer model based on the Kraus–Turner bulk formulation (Bleck et al., 1989) is implemented in the upper layer, which is treated specifically to permit horizontal variations of temperature and salinity. A minimum thickness of 20 m is specified for this layer. In all sub-surface layers, the temperature is chosen as the prognostic variable while the salinity is diagnosed according to the prescribed density.

From a dynamical point of view, the advantage of using the density as vertical coordinate is justified by the fact that eddy mixing in the ocean mainly takes place along neutral surfaces, while diapycnal mixing is several orders of magnitude lower. With an isopycnic coordinate system, the numerical diffusion associated with the advection scheme will be directed along density surfaces too. The benefit of this approach for physical tracers such as temperature and salinity, also holds for the advection and diffusion of the biological variables of the ecosystem model. Another advantage of the isopycnic formulation is its capacity to resolve sharp vertical density gradients at the base of the mixed layer: when the mixed layer gets deeper during winter, the density surfaces are deepened too, and the model can maintain large density gradients over the pycnocline.

The model domain covers the Atlantic ocean, with an enhanced focus of the orthogonal curvilinear grid (Bentsen et al., 1999) to better resolve the areas of strong biological activity in the northern hemisphere (Fig. 1). The Arctic basin is included in the model domain too, where a dynamic-thermodynamic ice model is implemented (Lisæter, 2000). The use of a variable grid resolution over the domain extended to the South Atlantic allows the prescription of closed boundaries far from the area of interest, where relaxation to climatology is enforced. There are four additional open-sea boundaries: the Bering Strait, the Gibraltar Strait, a cut in the Gulf of Mexico and the Kattegat Strait excluding the Baltic Sea. The bathymetry has been prepared by interpolation of the ETOPO5 data set onto the model grid. There are currently two versions of the DIADEM model, one coarse resolution which has been used during the testing phase of the project, and another version on exactly the same domain but with double resolution which will be used in the pre-operational experiment. The version used here is the coarser one, with a resolution of approximately $1/2^\circ$ in the regions of interest in the North Atlantic.

The circulation model is initialized using the Levitus (1994) climatology redistributed onto the model grid. In the context of twin experiments, the atmospheric forcing fields of heat, freshwater and momentum are derived from the ECMWF climatological data sets using bulk formulation

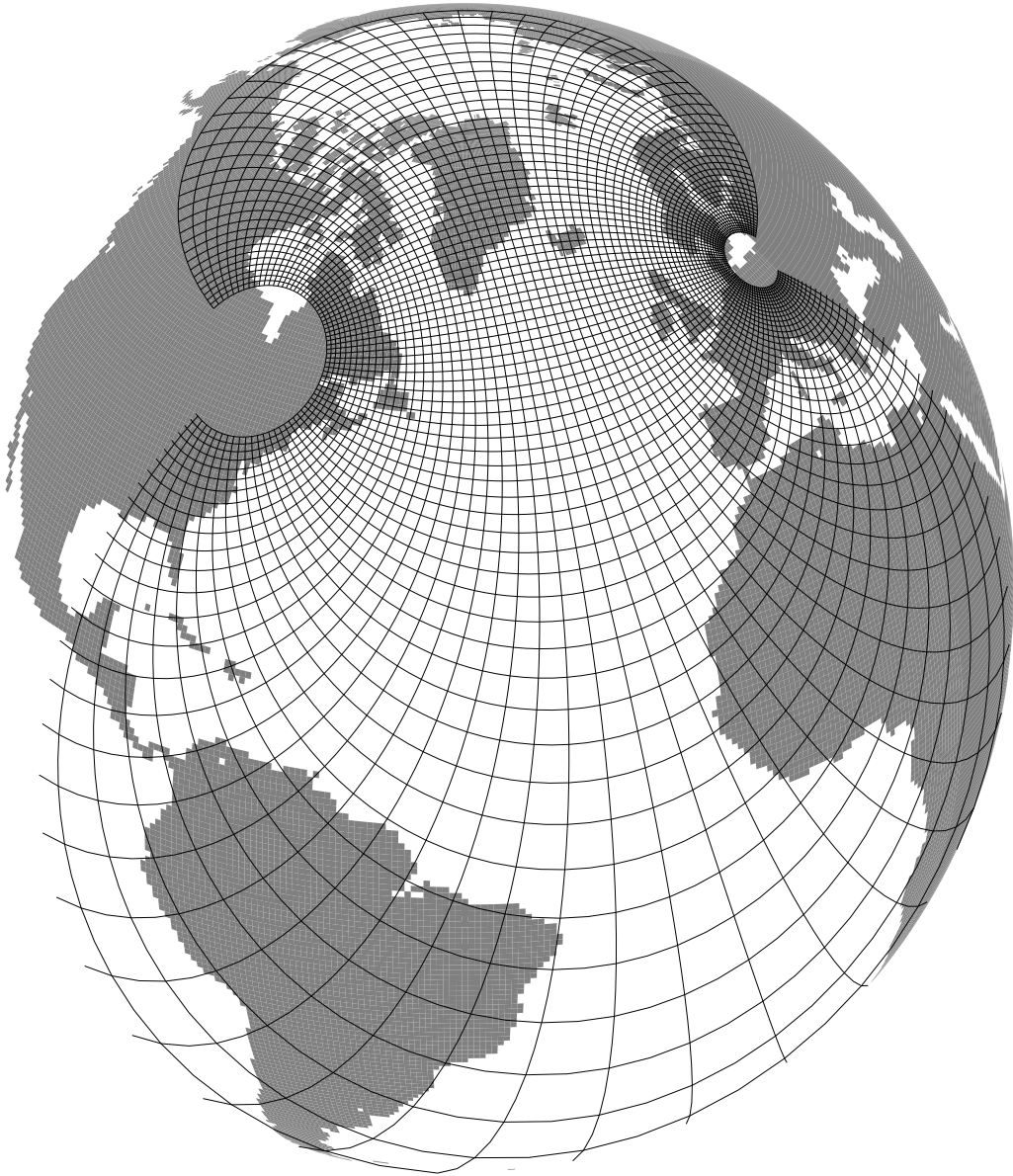


Fig. 1. Orthogonal curvilinear grid of the DIADDEM North Atlantic model.

(Drange and Simonsen, 1996). For real data assimilation experiments that will be conducted later, high frequency data sets (e.g., four times a day) will be chosen rather than climatological averages to synchronize the temporal response of the model with atmospheric events, and to respect in this way the synopticity between the observations and the model forcings. In order to avoid a long-term drift of the model simulations, the sea surface salinity is relaxed toward the Levitus climatological fields in the mixed layer. The model has been spun-up from rest during a five year's period before coupling with the biological model, using a baroclinic time step of 1600 s. A five

year's spin-up is a good compromise between the dynamical adjustment of the circulation and the limitation of the climatological drift from initial conditions.

The seasonal cycle of the upper ocean circulation is the dominant constraint imposed to the functioning of the marine ecosystem, and the ventilation of sub-surface layers plays a particular role in the cycling of organic matter that one seeks to simulate. Fig. 2 illustrates the seasonal variability in the vertical distribution of isopycnic layers along a vertical section crossing the North Atlantic at 57 N. During winter, the signature of the mixed layer extends from east to west at a mean depth of about 400 m, with a series of outcropping layers below. From spring to summer, detrainment takes place, the stratification is restored and the mixed layer depth reaches a minimum value.

2.2. The marine ecosystem model

The ecosystem model is originally based on the one dimensional, seven nitrogen-compartments, ecosystem model developed by Fasham et al. (1990) and modified by Fasham (1993). In this work, we follow the approach of Sarmiento et al. (1993), and split the ecosystem into two parts: one part that describes the exchange of nitrogen between the compartments in the euphotic zone, as illustrated in Fig. 3, and one part that parameterizes the decay of particulate and dissolved organic nitrogen below the euphotic zone. It is therefore, only the ecosystem in the euphotic zone that is explicitly modeled; below the euphotic zone the organic matter is gradually turned over to ammonium, and then to nitrate.

The model has seven nitrogen compartments, i.e., the phytoplankton P , the zooplankton Z , the bacteria B , nitrate NO_3^- , ammonium NH_4^+ , dissolved organic nitrogen (DON) and particulate organic nitrogen (PON) (the concentrations of the nitrogen species are all expressed in mmol-N m^{-3} units). In order to simulate the carbon cycling in the ocean, four additional variables have been coupled to the flow of nitrogen, yielding a total of 11 biogeochemical compartments: dissolved organic carbon (DOC), particulate organic carbon (POC), dissolved inorganic carbon (DIC) and total alkalinity (ALK). The concentration of the carbon species are expressed in mmol-C m^{-3} , and total ALK is expressed in $\text{mmol equiv-N m}^{-3}$.

The transport equations that govern the flow of the components of the ecosystem all take the form

$$0 = T(C_i) + h\mathcal{S}\mathcal{M}\mathcal{S}(C_i), \quad i = 1, \dots, 11 \quad (1)$$

with

$$T(C_i) = -\frac{\partial h C_i}{\partial t} - \nabla \cdot (\mathbf{u}h C_i) + \nabla \cdot (K_h \nabla h C_i) + \frac{\partial}{\partial z} \left(K_v \frac{\partial h C_i}{\partial z} \right) + \left(\frac{\partial h C_i}{\partial t} \right)_{\text{diap}} \quad (2)$$

and where the layer thickness h (in meters) for the examined layer is deduced from the mass continuity equation

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u}h) - \nabla \cdot (K_h \nabla h) - \frac{\partial}{\partial z} \left(K_v \frac{\partial h}{\partial z} \right) = \left(\frac{\partial h}{\partial t} \right)_{\text{diap}}. \quad (3)$$

Those equations are solved in the same isopycnic layers as the circulation model, with the exception of the uppermost layer which is split into two sub-layers to better resolve the attenuation

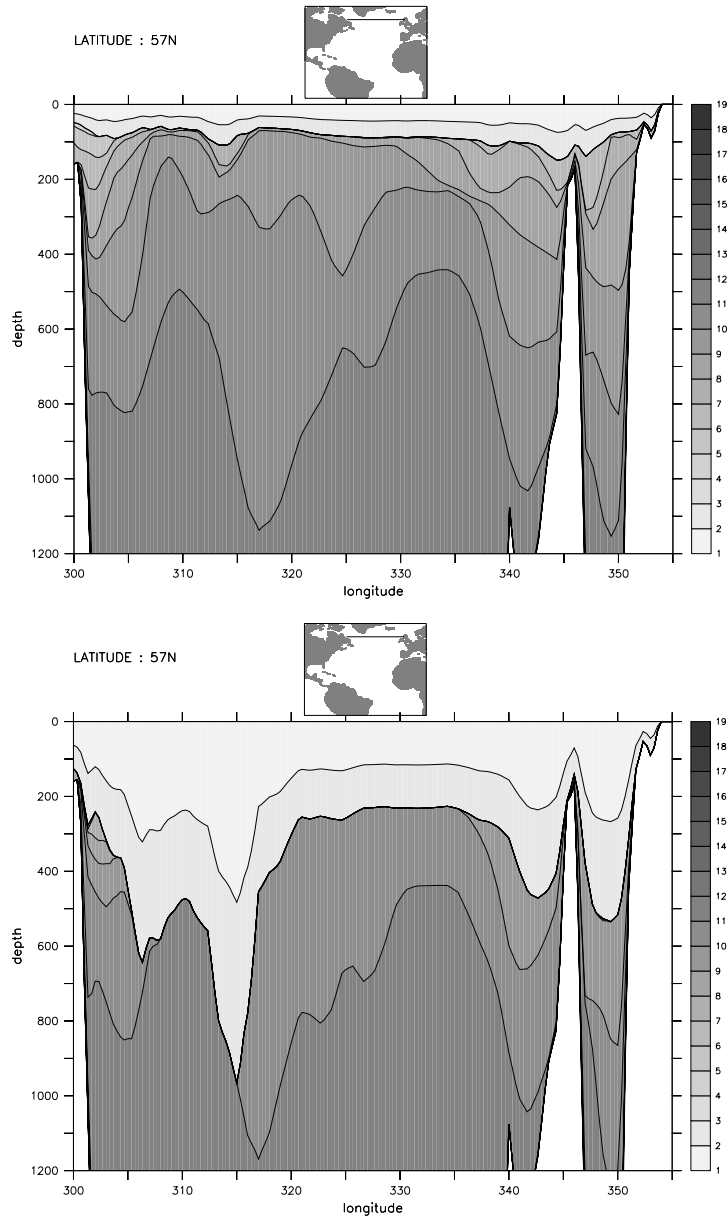


Fig. 2. Vertical distribution of isopycnic and biological layers along a vertical section crossing the North Atlantic at 57 N: winter (upper) and summer (lower) situations. Solid lines denote the interface between layers. The two uppermost layers (in white) correspond to the physical mixed layer which includes two biological sub-layers.

of light with depth in the euphotic zone. Thus, the vertical discretization in our implementation is achieved with a total of 18 biological layers (Fig. 2).

In practice, C_i , the concentration in mmol m^{-3} of nitrogen or carbon of the ecosystem component in question, is calculated by dividing hC_i from Eq. (2) by h from Eq. (3). SMI symbolizes

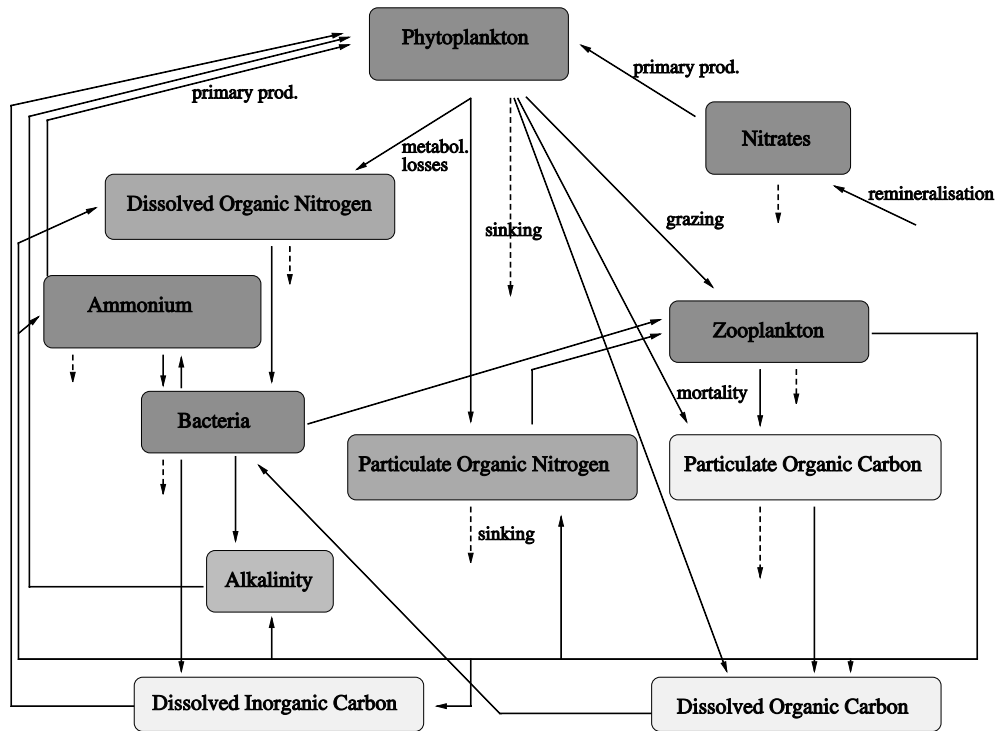


Fig. 3. Schematic diagram of the 11-compartment marine ecosystem model in the euphotic zone.

the source minus sink terms, thus $\mathcal{PMS}(C_i)$ represents the biochemical interaction terms within the compartment i . $\mathbf{u} = (u, v)$ is the horizontal velocity obtained from the physical model. Both ∇ and \mathbf{u} are defined along the isopycnal surfaces. The subscript diap refers to exchanges across the isopycnal surfaces. K_h and K_v are the isopycnal and diapycnal diffusivities. K_h is constant in the interior of the model domain. K_v depends on the local buoyancy frequency given by the physical model, and takes two forms: one in the mixed physical layer, another in the interior ocean (Drange, 1994). Instead of a perfect mixing as assumed in a Krauss–Turner model, the vertical mixing between the upper two biological sub-layers is computed explicitly based on a vertical diffusion coefficient of $0.05 \text{ m}^2/\text{s}$. This value is consistent with earlier modeling studies of the mixed layer (Gaspar et al., 1990).

In this simplified implementation, the partitioning between the FDM and Sarmiento models takes place at 180 m depth, a large enough value to include the entire euphotic zone in normal oceanic conditions. In general, the prescribed isobaric surface does not coincide with an isopycnal surface. At each time step, a check is performed in the code to determine which biological layers have an upper interface less deep than 180 m, and the FDM model then computes the SMS terms in these upper layers. In the deeper layers, the regeneration model computes the corresponding SMS terms of the biological equations to simulate the recycling of detritic material. The daily averaged surface irradiances are determined as a function of space and time using standard astronomical parameters. The extinction of light with depth is then computed explicitly in the upper layers for use in the FDM model. In particular, the phytoplankton light-limited growth rate

is evaluated within each layer by computing the effective irradiance with self-shading effects included (Fasham et al., 1990).

Fig. 3 illustrates the various *P.M.S* exchange processes between the compartments in the FDM model. The phytoplankton, which is the only carbon fixing compartment, grows if the water is illuminated and the water contains ammonium and/or nitrate; its concentration decreases due to grazing by zooplankton and mortality losses. The zooplankton is included in the model formulation partly because it acts as a dynamic switch on the phytoplankton concentration, and partly because it produces heavy particulate material that rapidly sinks out of the euphotic zone. Zooplankton grazes on phytoplankton, bacteria and particulate organic matter (or detritus), excretes dissolved organic matter, and a fraction of the pool dies or is grazed by higher predators. The grazing functions are defined in such a way that the zooplankton compartment grazes on the most abundant of the phytoplankton, bacteria and particulate organic matter according to specified preferences.

Detritus is the result of dead phytoplankton and zooplankton grazing (or sloppy feeding) on phytoplankton, bacteria and particulate organic matter. A fraction of the detritus pool is converted to dissolved organic matter, and a fraction sinks vertically in the water column. Bacteria participate to the decomposition of dead organic matter and liberate inorganic nutrients, which can then be recycled back to the primary producers. The bacteria model compartment obtains its carbon from dissolved organic matter (which is exuded by phytoplankton and zooplankton, and formed through breakdown of particulate organic matter), and takes up ammonium to obtain sufficient nitrogen to synthesize cell protein.

Dissolved inorganic carbon is removed from seawater by formation of soft tissue and calcium carbonate shells, and is produced through bacterial and zooplankton excretion. In addition, as CO_2 concentration changes due to evaporation, precipitation, river input, freezing/melting of ice and gas exchange across the air–sea interface, the compartment of dissolved organic carbon is adjusted in the two uppermost biochemical layers assuming that the composition of the dissolved salts in seawater does not change with salinity.

In the layers below 180 m depth, we follow the parameterization used by Drange (1996), and let the biogenic compartments *P*, *Z*, *B* and DON decay to ammonium, and then to nitrate, with a constant rate. In a similar way, DOC decays to DIC with the same rate.

2.3. *The coupled system*

The partial differential equations of the ecosystem model are integrated on the same grid as described before, but using a time step of 3 h. The advective velocities, layer thicknesses and thermodynamic variables are computed from a prior simulation of the ocean circulation model forced with climatological atmospheric conditions. As there is no feedback from the biology to the physics, the ecosystem can be driven off-line without particular restriction. No assimilation has been applied yet to the ocean circulation model, but it is planned to consider an assimilated solution at a later stage of the *DIADEM* project in order to improve the representation of the hydrodynamic constraints on the ecosystem dynamics.

Initial conditions are similar to those used by Drange (1996). The initial distributions of nitrate are derived from the climatology of Levitus interpolated onto the model grid, while total inorganic carbon and total alkalinity are initialized from data sets obtained during several field experiments.

For the biological compartments, constant initial values are specified as follows: $P = 0.14 \text{ mmol-N m}^{-3}$, $Z = B = 0.014 \text{ mmol-N m}^{-3}$, $N_r = \text{DON} = 0.1 \text{ mmol-N m}^{-3}$ in the upper biochemical layer, with exponential decrease at depth with a scale length of 100 m. The variables DOC and POC are set to 0.7 mmol m^{-3} of carbon in the upper layer. The concentration of DOC also decreases exponentially with depth, but the concentration of POC variable is set to zero below the first layer; in a same way, the concentration of PON is set to $0.1 \text{ mmol-N m}^{-3}$ in the upper layer and to zero below.

Fig. 4 illustrates the phytoplankton distribution in the mixed layer during spring, obtained after a spin-up of three years of the coupled system. At this period of the year, the spring bloom occurs, first in the mid-latitudes and then later in the sub-polar regions, showing a general agreement between the model and historical sets of CZCS data. A detailed analysis of the run, however, reveals unrealistic features in specific areas such as in coastal zones.

3. The SEEK data assimilation scheme

The method used to assimilate ocean color data in the coupled model is the SEEK filter, which is a non-linear, reduced-order assimilation scheme described in several earlier publications (e.g., Pham et al., 1998; Verron et al., 1999; Brasseur et al., 1999). The state vector of the problem is composed of all biological variables represented on the model grid. We will come back later on the possible choices concerning the definition of the state vector of this particular assimilation problem.

In essence, the conventional Kalman filter consists of sequential corrections of the model trajectory, taking into account the balance between the confidence in the model prediction and the accuracy of the observed quantities (Kalman and Bucy, 1961; Gelb, 1974). However, the conventional Kalman filter cannot be implemented by brute force because of the excessive amount of computations required by the error propagation equation, and also because of the lack of knowledge in the prescribed error statistics.

The SEEK filter introduced by Pham et al. (1998) is a sequential estimation method of reduced-order, based on some relevant decomposition of the error covariance matrix. The initial error covariance is specified in a way which subsequently simplifies the computation of the error propagation: a finite number of directions are selected in the state space to intermittently reset the consistence between the model and the observations. In the optimal case, the dynamical propagation of the error covariance from one analysis step to the next is performed according to the KF equations, using either the non-linear model operator or its tangent linearization (Verron et al., 1999). In the numerical experiments described in Section 5, however, the error sub-space will not be updated using the model dynamics (as it is done in the generic SEEK filter) in order to avoid expensive calculations and dedicate more attention to the statistical analysis updates.

In order to preserve the rank of the error covariance matrix, the model error is parameterized by means of a so-called compensation technique that amplifies the pre-existing error modes during the forecast. As the dimension of the sub-space is much smaller than the dimension of the state vector, the computational burden of the filter is substantially reduced.

In practice, the first-guess error covariance matrix is initialized with a limited number of three-dimensional, multivariate, empirical orthogonal functions (EOFs) describing the dominant modes

of the system variability and defining by this way the structure of the error sub-space at the initial time. A key-point of the method is in the three-dimensional, multi-variate nature of the analysis, which is essential to spread the information from the observed quantities to the whole state vector

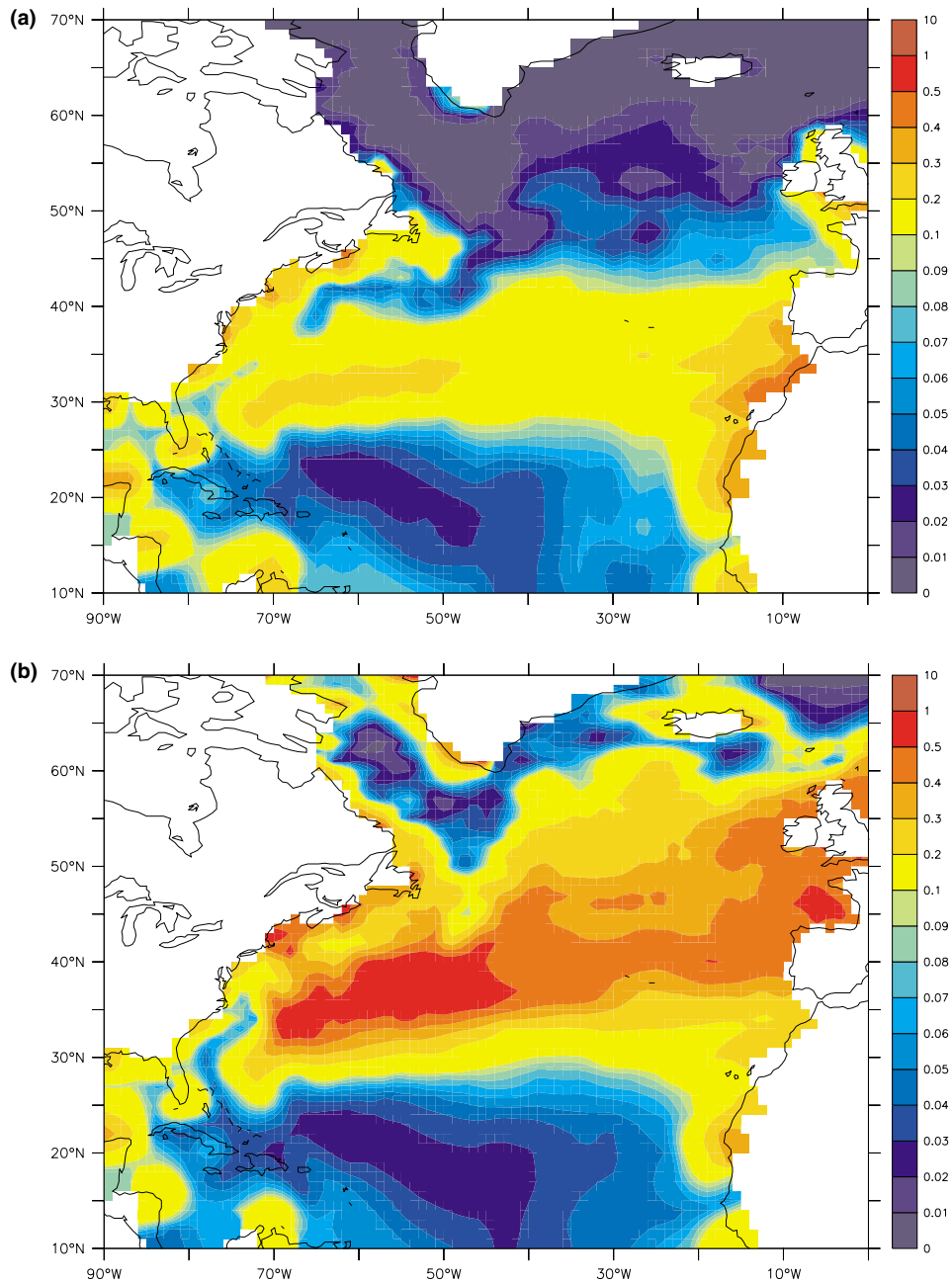


Fig. 4. Distribution of phytoplankton in the mixed layer (in mmol-N m^{-3} units): (a) February, (b) April, (c) June, (d) August (third year spin-up of the coupled model before assimilation).

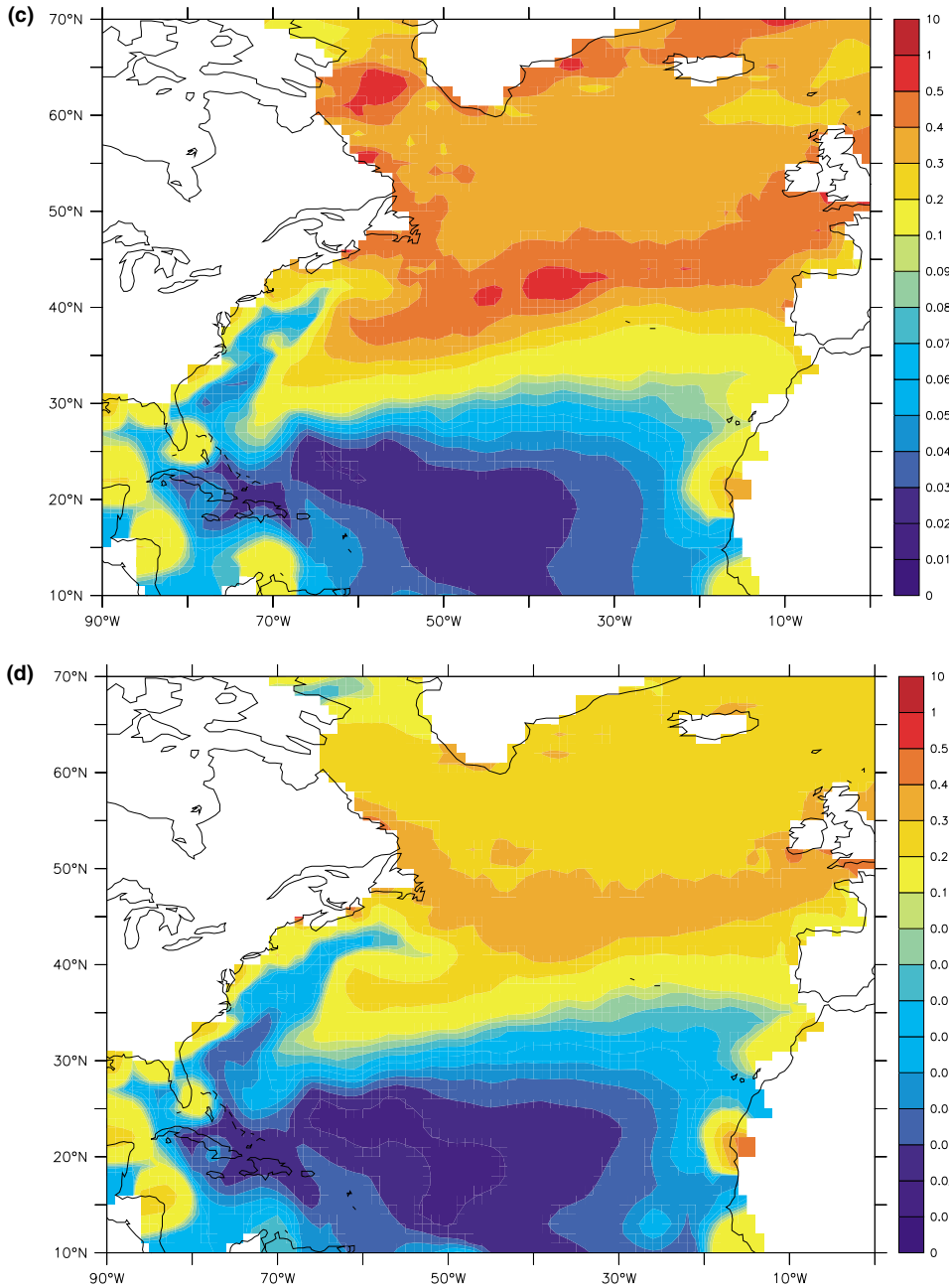


Fig. 4. (Continued).

(e.g., to specify how phytoplankton observations are used to correct the other biological variables). With such an algorithm, it is always possible to represent the error covariance matrix of the background state in reduced form, as:

$$\mathbf{P} = \mathbf{S}^T \Lambda^- \mathbf{S}, \quad (4)$$

where \mathbf{S} is a simplification operator (of dimension $r \times n$) related to the r directions $\{\mathbf{S}\}_k$ of the error sub-space, and Λ^- is related to their amplitude (Cane et al., 1996; Pham et al., 1998). The reduced-rank approximation allows the analysis step of the SEEK filter to be written as:

$$x_i^a = x_i^f + \mathbf{K}[y_i - \mathbf{H}(x_i^f)], \quad (5)$$

in which x_i^f is the forecast state vector (of dimension n) obtained by model integration up to time t_i ; x_i^a is the state vector after the analysis step and y_i is the vector of observed quantities (in our case study, maps of phytoplankton concentrations) of dimension p . The observation operator \mathbf{H} is of dimension $p \times n$ and relates the data to the model variables. The gain matrix \mathbf{K} is expressed in terms of the simplification operator \mathbf{S} as:

$$\mathbf{K} = \mathbf{S}^T [\rho(\Lambda^-)^{-1} + \mathbf{S}\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}^T]^{-1} \mathbf{S}\mathbf{H}^T \mathbf{R}^{-1}, \quad (6)$$

where \mathbf{R} is the observation error covariance matrix. Following the same approach as Pham et al. (1998), the model error is globally taken into account by means of a forgetting factor ρ in (6), expressing that the main contribution of an imperfect model is to amplify the already existing modes of the background error.

Combining (5) and (6), the correction of the forecast can further be expressed as a weighted combination of the r error modes:

$$x_i^a = x_i^f + \mathbf{S}^T \mathbf{w} \quad (7)$$

with

$$\mathbf{w} = \Lambda^+ \mathbf{S}\mathbf{H}^T \mathbf{R}^{-1} [y_i - \mathbf{H}(x_i^f)] \quad (8)$$

and

$$\Lambda^+ = [\rho(\Lambda^-)^{-1} + \mathbf{S}\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{S}^T]^{-1}. \quad (9)$$

The error of the analysis state vector is now equal to:

$$\mathbf{P}^a = \mathbf{S}^T \Lambda^+ \mathbf{S}. \quad (10)$$

Eq. (10) expresses the error reduction due to the addition of new information into the system: the error decreases as a function of the accuracy (i.e., \mathbf{R}^{-1}) and quantity (i.e., $\dim \mathbf{H}$) of observations available at each analysis cycle. A requirement for optimal weighting of the data vs. model solution when cycling several forecast/analysis steps is to properly update the error covariance according to (10). If this operation is neglected, a too strong weight is eventually given to the data, and the estimation process becomes sub-optimal. In Section 5, we will examine the importance of carefully considering the error update in the assimilation sequence.

Another determining element of the SEEK algorithm is r , the dimension of the error sub-space. As the error structure is expressed in terms of statistical data, a series of modes are needed to approximate the dominant components of the error covariance. The number of modes needed in this particular problem will be assessed on the basis of sensitivity experiments with respect to the truncation of the EOF analysis of the model variability.

4. The twin experiment strategy

The main objective of this work is the validation of the assimilation system with the coupled model based on a twin experiment approach. The assimilation of pseudo-measurements of a state variable such as surface phytoplankton, allows the assessment of the multi-variate properties of the scheme, examining the assimilation impact on non-observed variables (e.g., zooplankton) sampled from a previous model run and considered as the truth.

The twin experiments are conducted in a similar way as described by Brasseur et al. (1999). First, the coupled model is integrated during two years after the spin-up of the physical model in order to adjust the seasonal cycling of the ecosystem. The last annual cycle is selected as the reference to compute empirical orthogonal functions of the biological model variability during that period. In addition to the layer thicknesses, the biological state vector includes all discretized variables of the 11-compartment ecosystem distributed in the 18 biological layers. As the numerical grid has 70 points in the zonal direction, and 65 points in the meridional direction, the dimension of the estimation space is equal to $n = 11 \times 18 \times 70 \times 65 = 900900$.

The model history is then sampled every 10 days, so that 36 biological model states are available to build a sample covariance matrix representative of the variability during the reference annual cycle. The eigenvectors associated with the positive eigenvalues of this matrix define the possible directions of the initial error sub-space of our assimilation problem. These multivariate, empirical orthogonal functions are ordered according to the magnitude of their eigenvalues, and the series is truncated to the first dominant modes in order to initialize the assimilation scheme with a reduced-rank error covariance matrix such as (4). Sensitivity experiments have been conducted with 3, 9 or 11 modes to investigate the impact of the truncation on the assimilation results.

The pseudo-measurements considered for the twin experiments consist of basin-scale snapshots of surface phytoplankton generated every 10 days by the model simulation from June to September. One important difference with real data lies in the absence of clouds masking the ocean surface (especially at high latitudes), and also in the fact that synoptic pictures of the ocean color taken from satellites only cover small regions at a time. Another problem arises when using real ocean color data, as the measured quantities (i.e., irradiances) must be converted into chlorophyll and then to phytoplankton concentrations. The observation operator corresponding to real data is therefore a highly complex function of the model variables. With a twin-experiment approach, we skip the uncertainties in the specification of this observation operator, and so, we dedicate more attention to the assimilation mechanisms.

The length of an assimilation cycle corresponds to 10 days, i.e. 60 time steps of the ecosystem model or 540 baroclinic time step of the circulation model. The observation error specified in (6) is assumed to be homogeneous and uncorrelated in space, and is accordingly parameterized as a diagonal matrix of the form $\sigma^2 \mathbf{I}$. In order to determine the observation error variance σ^2 , the standard deviation of phytoplankton around its mean distribution has been calculated from the 36 samples of the reference run, assuming that observation errors represent a small fraction (10%) of the natural variability. A value of $0.007 \text{ mmol-N m}^{-3}$ has eventually been prescribed for σ . An observational error of 10% is highly optimistic compared to the ultimate error budget expected for SeaWiFS data. However, our goal here with twin experiments is primarily to examine the multivariate properties of the scheme and the relevant estimation space. A small value for the observation error is thus preferable in this context.

A final element to be prescribed in the assimilation scheme is the model error, i.e. the error generated by the imperfect model between two successive analysis steps. With a coupled ecosystem model, a significant part of the error arises from inadequate process parameters of the biological equations. As these parameters are kept constant in space and time, it is unlikely that their values adjusted to particular biological species found in specific areas, be adequate for the whole model domain. Errors in the forcings (e.g. solar radiation), circulation model (e.g., mixed layer dynamics), numerical discretization (e.g., limited number of sub-layers in the euphotic zone), and non-linear interactions should also be considered in the error budget, but their statistical characterization remains basically unknown. After a series of empirical trials, the forgetting factor used to parameterize the model error was set to 0.5, which corresponds to doubling the error covariance from one analysis step to the next.

In order to initialize the twin experiments, the model state of a previous run corresponding to 1st January is selected as a first guess. The choice of a winter situation is made with the intention to introduce an explicit mismatch between the biological state and the summer forcing conditions, such as solar radiation and mixed layer properties, encountered at the beginning of the assimilation period. Thus, the misfit between the first guess and the reference run is large at the initial time, and is expected to decrease afterwards. The efficiency of the assimilation system will be evaluated in terms of RMS misfit between the reference and the assimilated solutions.

5. Assimilation experiments

In this section, we discuss the results of the twin experiments performed during seven assimilation cycles to validate the assimilation system into the coupled model and get practical experience before starting the assimilation of real data. The first series of experiments has been set up to investigate the possibility of using a state vector limited to the mixed layer for the analysis step, instead of the full state vector defined in Section 4. In the second set of experiments, the effect of considering a truncated EOF decomposition to initialize the error covariance matrix is examined, with the aim to determine the dimension of the error sub-space needed to control the state variables efficiently. Finally, the impact of carefully computing the statistical evolution of the error covariance in the reduced-space as expressed by (9) is evaluated, by comparing the assimilation performances with the previous experiments in which the background error covariance is kept unchanged from one analysis to the next.

Due to the prospective nature of this work, the statistics illustrated in the figures will be primarily focused on the phytoplankton, which is the observed quantity, and also on the other biological variables directly related to the evolution of the phytoplankton in the food web, i.e., zooplankton, bacteria, nitrate and ammonium.

5.1. *Reduced-state vector in the mixed layer*

In essence, the Kalman filter is a multivariate assimilation method which statistically propagates the information from the observed quantities to the rest of the state vector. This propagation takes place at every analysis step according to the background error covariance used to

compute the gain matrix. In a reduced-order Kalman filter like SEEK, the multivariate correction depends critically on the representativeness of the modes defining the error sub-space, as shown by Eq. (7). In practical assimilation problems, however, our capacity to characterize the error structure with a sufficient degree of robustness is often limited by the lack of statistical data.

In the ecosystem model examined in this work, one can assume that the covariances between surface properties will be more robust than the covariances between surface and deep variables. This is especially the case with an “hybrid” ecosystem model which consists of different dynamical equations in the euphotic zone and in the deep ocean. Instead of considering the full biological state vector defined in the 18 biological layers, we have tested the idea to reduce the state vector to the two uppermost levels of the mixed layer, and to use this reduced-state vector to statistically correct only the surface biological variables at the analysis step. While the lower layers are left unchanged by the Kalman gain, the biological variables in the sub-surface layers are then adjusted dynamically by the model forecast up to the next analysis. In order to validate this idea, we need to verify if the physical and ecosystem dynamics which couple the mixed layer to the other layers, are efficient enough to transmit the observed information at the analysis step to the rest of the water column.

Fig. 5 compares the results in the mixed layer of two assimilation experiments which differ by the use of a full vs. reduced-state vector as described above. The solid curves of Fig. 5 show the RMS misfit, expressed in mmol-N m^{-3} , in the mixed layer between the true states and the analyses obtained when the biological variables are updated in the mixed layer only; the dotted curves identify the same quantity in the assimilation experiment with the full state vector. The simulation without assimilation (dashed line), initialized from the same conditions as the assimilation experiments, exhibits large model/data misfits which remain of the same order of magnitude during the 70 days of the experiment. By contrast, the error of the first assimilated state is strongly reduced with respect to the initial misfit. As expected, the first analysis in the mixed layer is strictly the same in the two experiments, while the subsequent analyses differ slightly. The control on the phytoplankton is fairly good, with a misfit below $0.1 \text{ mmol-N m}^{-3}$ during the seven assimilation cycles, while the control on the non-observed variables (e.g., zooplankton or ammonium) suffers from some lack of stability.

Below the mixed layer, the two experiments differ strongly (Fig. 6). The RMS error for the first day is left unchanged in the reduced-state vector experiment as there is no statistical correction in layer 3 (and below). However, the misfit 10 days later has decreased significantly to the same level as in the full state vector experiment, illustrating that some dynamical adjustment has occurred with the analyzed variables in the mixed layer. The following assimilation cycles amplify the difference between the two assimilation methods for all biological variables except for bacteria and phytoplankton. The pure simulation always exhibits larger errors than in the assimilation, with the exception of the final ammonium analysis in the experiment where all layers are updated statistically.

These results indicate that, relying on the model adjustment to correct sub-surface layers is preferential to making statistical updates based on poorly known surface to sub-surface correlations. The truncation of the error sub-space to nine modes (in these experiments) has obviously a more severe impact on the vertical than on the horizontal direction. We have therefore adopted the reduced-state approach for the remaining experiments.

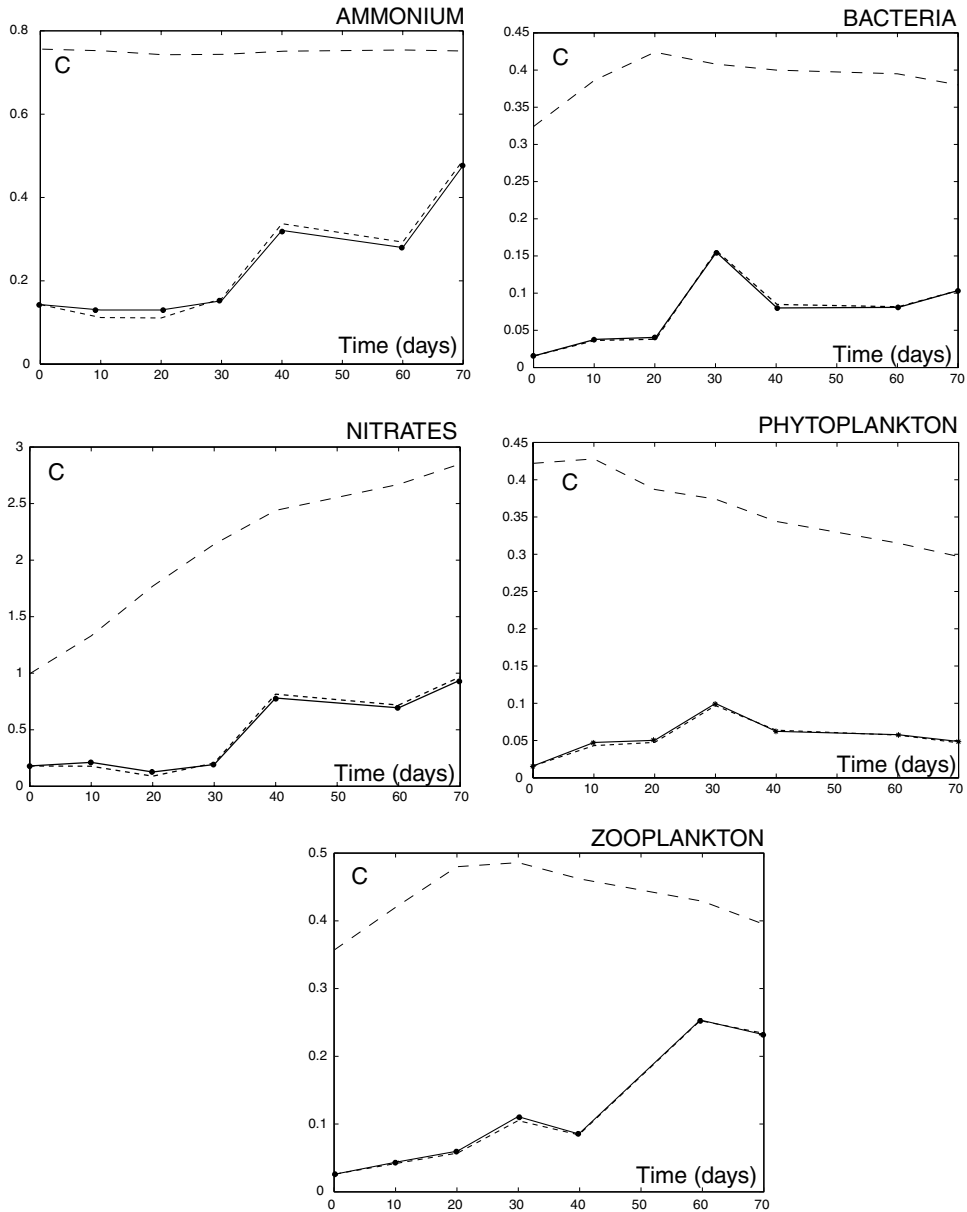


Fig. 5. Reduced-state vector experiments: RMS misfit in the mixed layer (expressed in mmol-N m^{-3}) between the true states and the analyses with the full state vector (dotted line), and with the biological variables updated in the mixed layer only (solid line). The dashed line represents the free simulation, without assimilation.

5.2. Sensitivity to the dimension of the error sub-space

Sensitivity experiments are performed afterwards to determine the most appropriate rank for the error sub-space of the SEEK filter. The initial EOF decomposition of the model variability

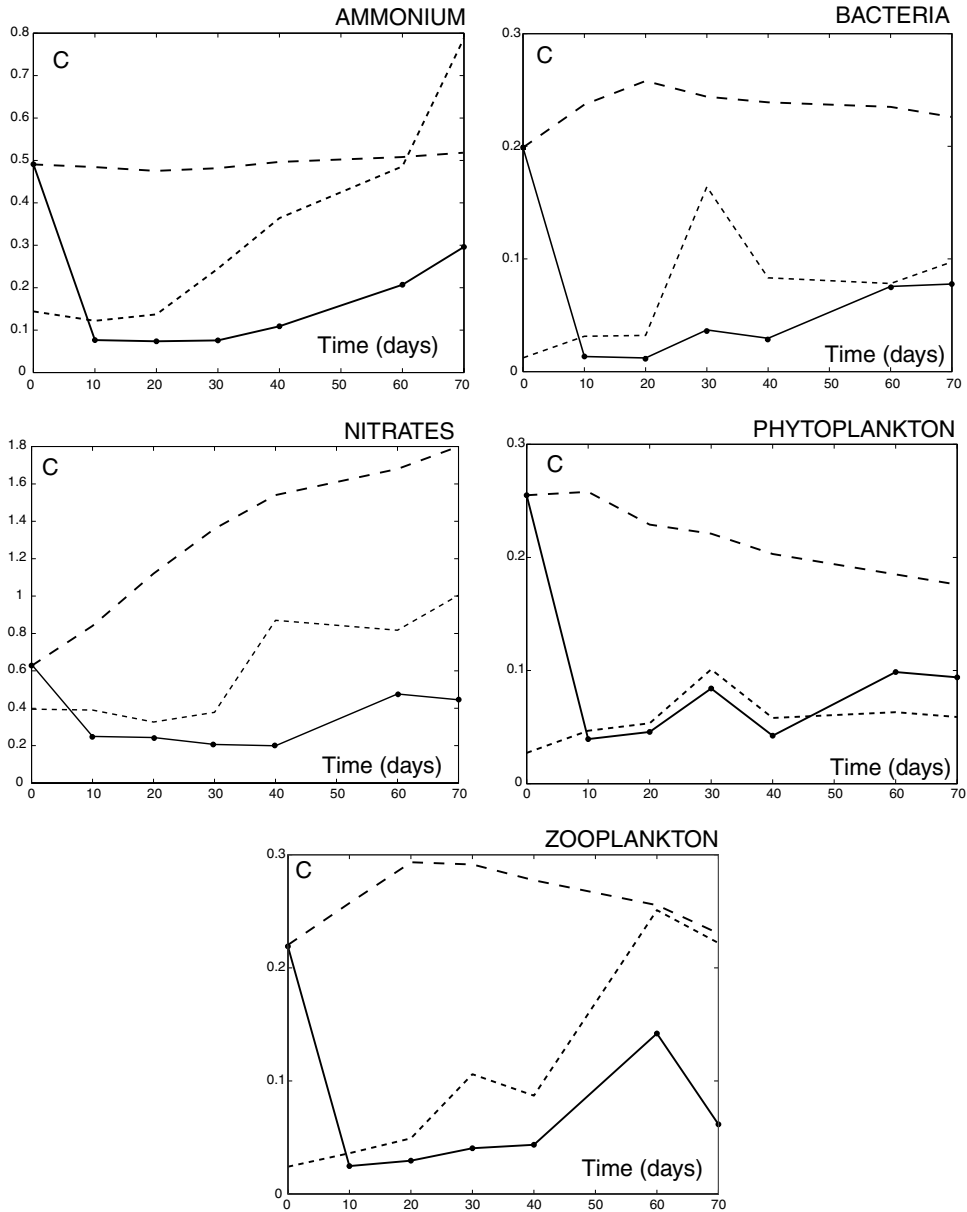


Fig. 6. Reduced-state vector experiments: RMS misfit below the mixed layer (expressed in mmol-N m^{-3}) between the true states and the analyses with the full state vector (dotted line), and with the biological variables updated in the mixed layer only (solid line). The dashed line represents the free simulation, without assimilation.

during one annual cycle includes 35 modes. The three dominant modes explain more than 80% of the model variance, while six additional modes are needed to explain 99% of the total variance.

Additional experiments have thus been conducted to assess the relevance of an EOF decomposition truncated to nine or three modes to initialize the background error sub-space.

In spite of the simplified assimilation context, it is important indeed to determine if the model trajectory can be controlled with a few error directions only as the computational burden of the method increases with the dimension of the error sub-space. It is worth noting, however, that these twin experiments will not address the question of the representativeness error arising from the difference between the model and the actual variability of the ecosystem.

The RMS error of the free run, the assimilation run with three modes, and the assimilation run with nine modes referenced to the true states are illustrated on Figs. 7 and 8. One can first notice that the model without assimilation conserves its initial error level over a 70-day period or even exhibits some drift (e.g., for the nitrate). In the mixed layer (Fig. 7), the assimilation run with three modes efficiently reduces the error on all variables at the first analysis step, but the control on all but phytoplankton variables is progressively lost in the subsequent analyses. The assimilation run becomes even worse than the free run for ammonium. The residual error left on the missing modes introduces inconsistencies which tend to amplify as a result of the complex, non-linear interactions between the ecosystem compartments. By contrast, the assimilation with nine modes improves the analyses quite well. The error on the phytoplankton is stabilized below $0.1 \text{ mmol-N m}^{-3}$ during the whole assimilation experiment, while the drift of the error on the other variables has been reduced interestingly. However, some signs of instability still persist for some variables such as zooplankton.

Below the mixed layer (Fig. 8), the same general behavior is observed. The impact of six additional modes is mostly visible on nitrate and ammonium, which are better adjusted by the model dynamics than with three modes only. However, the stability of the filter is not achieved, as all curves exhibit an increase of the error after the initial convergence.

5.3. Statistical evolution of the background error

The experiments discussed so far have been all conducted with a constant background error covariance expressed in terms of the original EOFs. The last experiment demonstrates the relevance of updating the error covariance after each analysis step according to (9). It is worth remembering that, like in the previous experiments, the simplification operator does not evolve with the model dynamics. A final assimilation experiment is thus achieved, using nine modes to initialize the error covariance matrix.

The RMS error statistics are superimposed to those of the previous experiments to make the comparison easier. Figs. 7 and 8 demonstrate that the convergence and stability of the assimilation is now achieved for all variables, both in the mixed layer and below. An interesting element to note is the similar misfit obtained for a given analysis and the forecast 10 days later (not shown in the figures), suggesting that the new information introduced by the assimilation process is well accepted by the model. The quality of the estimates is in agreement with the standard deviation of the measurement error.

In order to give an idea of the geographic distribution of the correction made at a particular analysis step, we finally compare maps of the absolute misfit of the data vs. the free run in Fig. 9, and the data vs. the assimilated run on Fig. 10 over the Atlantic domain. These maps show that the work of the assimilation dominates at mid- and high-latitudes, i.e., where the ecosystem activity is mostly significant.

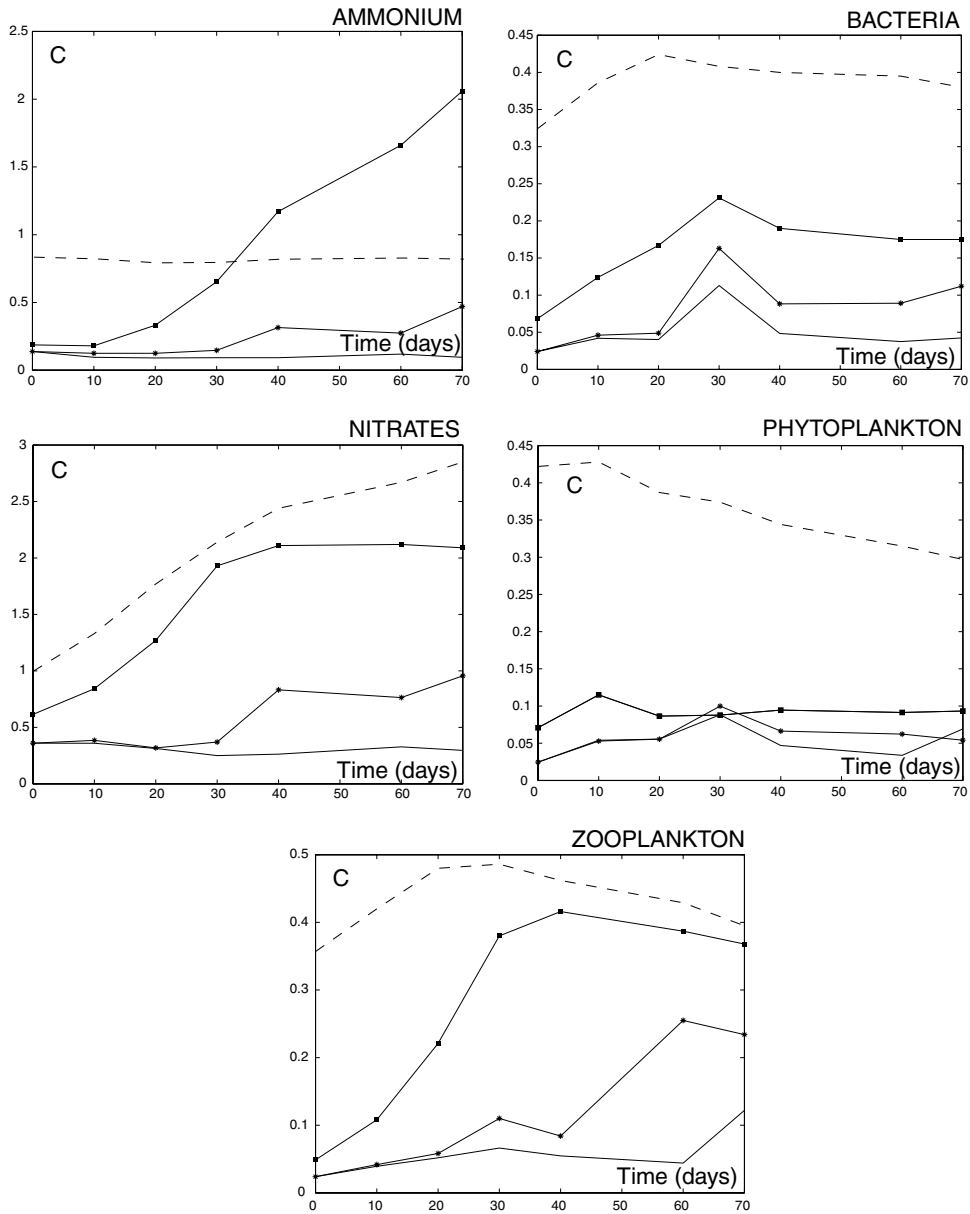


Fig. 7. Rank reduction experiments: RMS misfit in the mixed layer (expressed in mmol-N m^{-3}) between the true states and the free model run (dashed line), the analyses with three error modes (square-solid line) and the analyses with nine error modes (circle-solid line); the solid line is for the analyses with statistical update of the error sub-space.

6. Conclusions

A data assimilation system for coupled ocean circulation and marine ecosystem model has been implemented for the North Atlantic and the Nordic Seas. The originality of the work discussed in

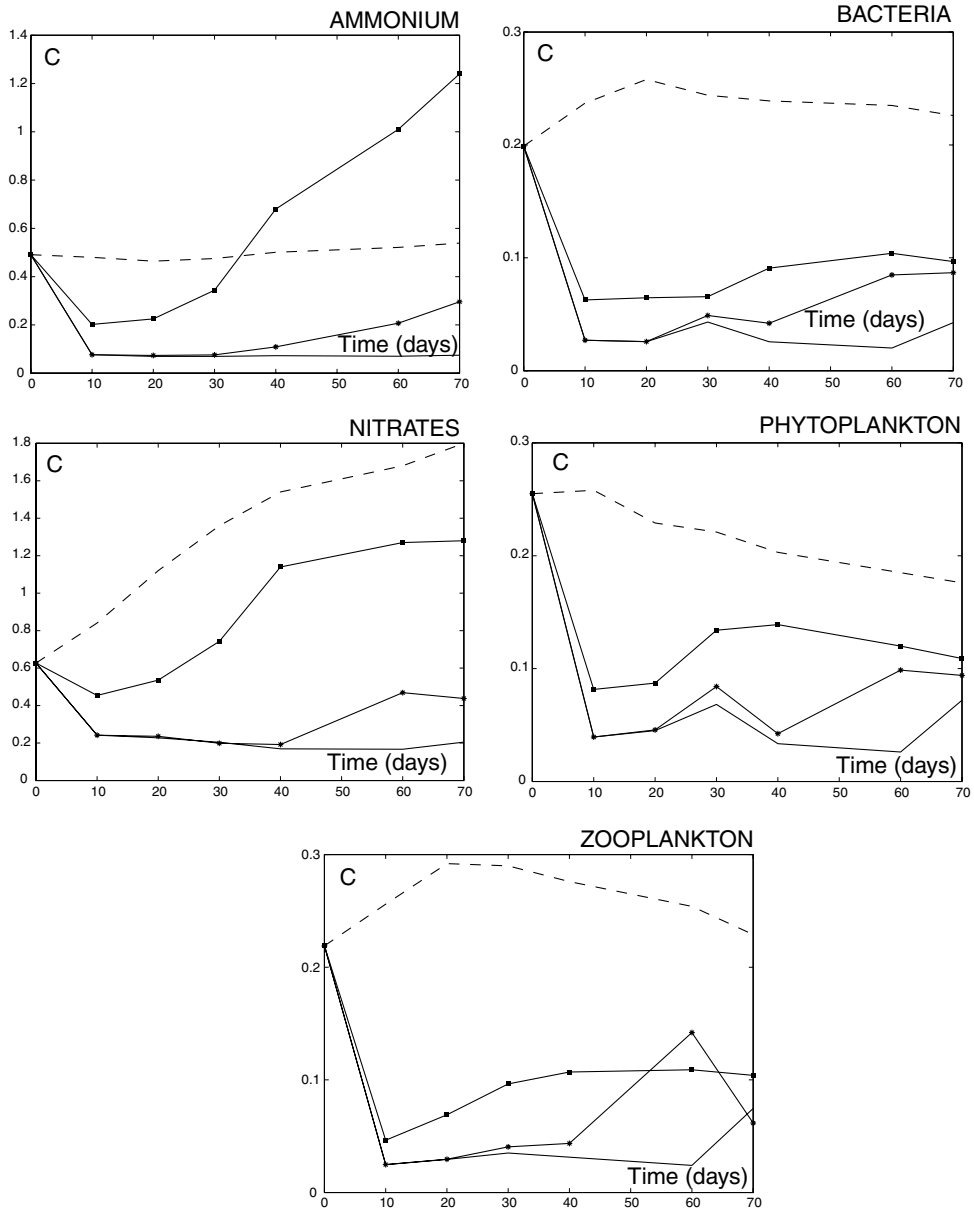


Fig. 8. Rank reduction experiments: RMS misfit below the mixed layer (expressed in mmol-N m^{-3}) between the true states and the free model run (dashed line), the assimilation with three error modes (square-solid line) and the assimilation with nine error modes (circle-solid line); the solid line is for the assimilation with statistical update of the error sub-space.

this paper lies in the sequential assimilation method – the singular evolutive extended Kalman filter – which has been used for the first time with a three-dimensional model of primary production. Our strategy significantly differs from the more conventional approaches based on the

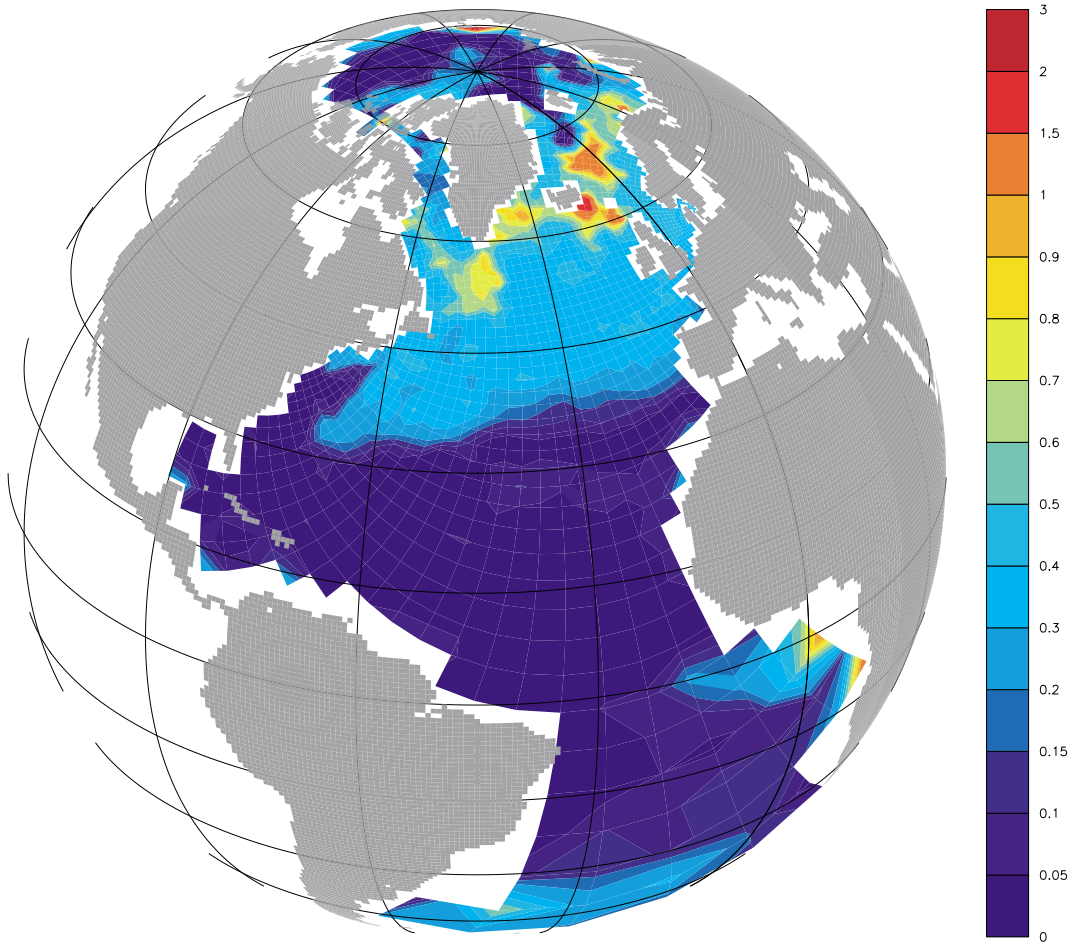


Fig. 9. Absolute difference of phytoplankton concentration in the mixed layer (in mmol-N m^{-3} units) between the free run and the observations at the beginning of June.

optimization of poorly known biological model parameters, which is most often found in the current practice.

Twin experiments have been performed to validate the assimilation system into the coupled model and get practical experience before starting the assimilation of real data. These numerical results demonstrate that a sequential updating scheme can efficiently correct the multiple components of a complex ecosystem with measurements of one single biological variable only. This is due to the multivariate character of the error sub-space which contains the dominant modes of the system variability, and therefore permits the transfer of information from observed to non-observed quantities in the euphotic zone.

However, our experiments also indicate the need to discriminate the behavior of the assimilation scheme between the upper layers in the euphotic zone where the FDM model is integrated, and the deeper layers where simple regeneration equations are applied. The diagnostics suggest that a reduced-order initial error covariance matrix is primarily damaging

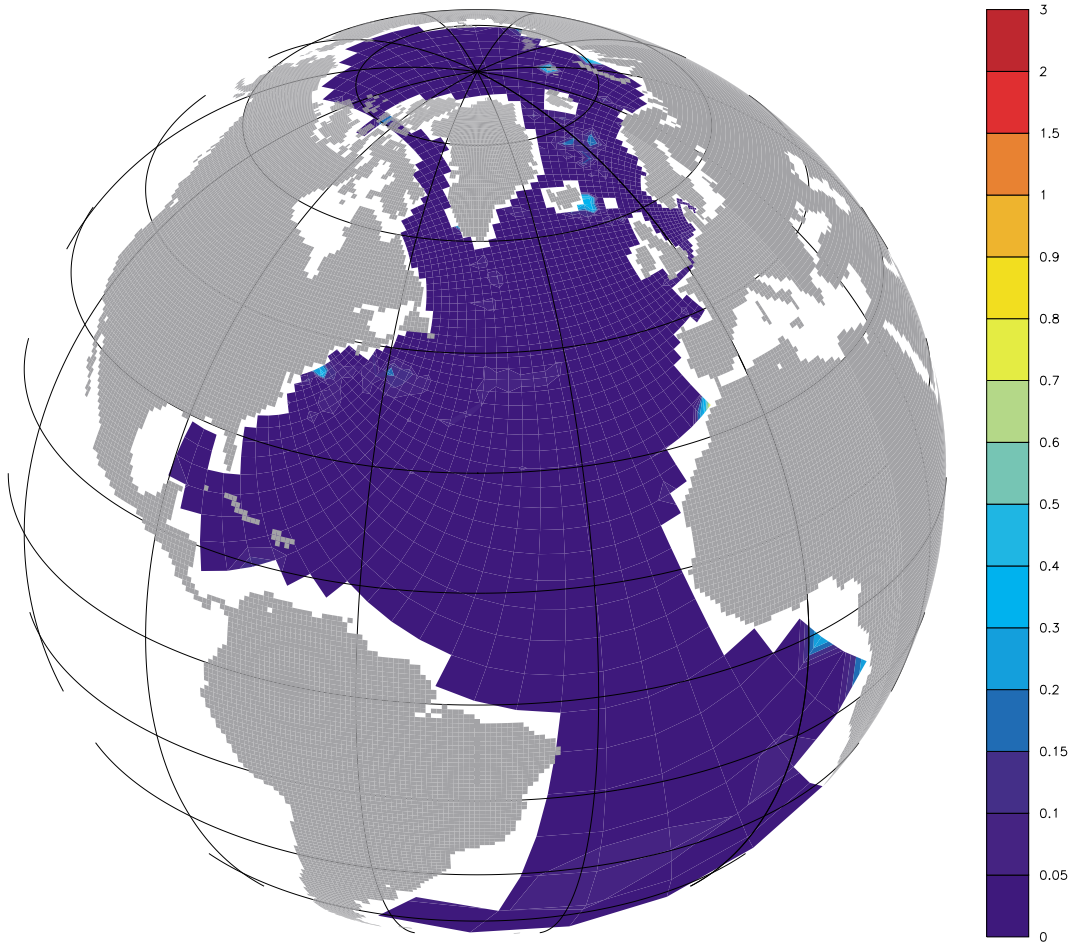


Fig. 10. Absolute difference of phytoplankton concentration in the mixed layer (in mmol-N m^{-3} units) between the assimilation run and the observations, beginning of June.

the covariances on the vertical direction, at least for specific variables such as nitrate, zooplankton and ammonium. It may therefore be preferential to explicitly analyze the ecosystem in the upper layers only and let the model adjust the deeper distributions dynamically, rather than to correct the three-dimensional system statistically. Of course, some care must be given to these conclusions which may be biased by the simplified context of the twin experiments.

The stability of the filter after the first analyses has been examined in a final experiment, demonstrating the need to carefully update the error covariance of the analysis according to the statistical equations of the Kalman filter. Optimal interpolation schemes may thus be inappropriate to this kind of assimilation problems, and there is also an issue to consider a dynamical SEEK filter that would explicitly update the error sub-space according to the model dynamics. This question is also related to the necessity to further examine the dynamical consistency of the sequential updates in the system trajectory.

In order to develop the system further and consider it for more realistic experiments, a number of dedicated studies will be needed, especially to better represent the model error and the background error statistics, to account for the error structure associated with real ocean color data measured from space, and also to make intrinsic improvements in the ocean dynamics and thermodynamics as the physical environment of the marine ecosystems.

Acknowledgements

This work has been supported by the EC-MAST-III DIADEM project under contract MAS3-CT-980167.

References

- Armstrong, R.A., Sarmiento, J.L., Slater, R.D., 1995. Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models. In: Powell, Steele (Eds.), *Ecological Time Series*. Chapman & Hall, London, pp. 371–390.
- Bentsen, M., Evensen, G., Drange, H., Jenkins, A.D., 1999. Coordinate transformation on a sphere using conformal mapping. *Mon. Weather Rev.* 127, 2733–2740.
- Bleck, R., Boudra, D., 1986. Wind-driven spin-up in eddy resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Geophys. Res.* 91, 7611–7621.
- Bleck, R., Hanson, H.P., Hu, D., Kraus, R.B., 1989. Mixed layer-thermocline interaction in a three-dimensional isopycnic coordinate model. *J. Phys. Oceanogr.* 19, 1417–1439.
- Bleck, R., Hanson, H.P., Hu, D., Smith, L.T., 1992. Salinity-driven thermohaline transients in a wind- and thermohaline-forced isopycnic coordinate model. *J. Phys. Oceanogr.* 22, 1486–1515.
- Brasseur, P., Ballabrera, J., Verron, J., 1999. Assimilation of altimetric observations in a primitive equation model of the gulf stream using the steady-state limit of a singular evolutive extended Kalman filter. *J. Mar. Syst.* 22 (4), 269–294.
- Burgers, G., van Leeuwen, P.J., Evensen, G., 1998. On the analysis scheme in the ensemble Kalman filter. *Month. Weather Rev.* 126, 1719–1724.
- Cane, M.A., Kaplan, A., Miller, R.N., Tang, B., Hackert, E.C., Busalacchi, A.J., 1996. Mapping tropical pacific sea level: data assimilation via a reduced-state Kalman filter. *J. Geophys. Res.* 101 (C10), 22599–22617.
- Drange, H., 1994. An isopycnic coordinate carbon cycle model for the North Atlantic, and the possibility of disposing of fossil fuel in the ocean. Thesis of Bergen University, Norway.
- Drange, H., 1996. A 3-dimensional isopycnic coordinate model of the seasonal cycling of carbon and nitrogen in the Atlantic Ocean. *Phys. Chem. Earth* 21, 503–509.
- Drange, H., Simonsen, K., 1996. Formulation of Air–Sea Fluxes in the ESOP2 Version of MICOM, Technical Report 125. Nansen Environmental and Remote Sensing Centre, Norway.
- Evensen, G., 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.* 99 (C5), 10143–10162.
- Evensen, G., vanLeeuwen, P.J., 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* 128, 1852–1867.
- Fasham, M.J.R., 1993. Modelling the marine biota. In: *The Global Carbon Cycle*, vol. 1, 15, Springer, Berlin.
- Fasham, M.J.R., Evans, G.T., 1995. The use of optimization techniques to model marine ecosystem dynamics at the JGOFS station at 47N 20W. *Phil. Trans. R. Soc. Lond. B* 348, 203–209.
- Fasham, M.J.R., Ducklow, H.W., McKelvie, S.M., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *J. Mar. Res.* 48, 591–639.
- Fennel, K., Losch, M., Schröter, J., Wenzel, M., 2000. Testing a marine ecosystem model: sensitivity analysis and parameter optimization. *J. Mar. Syst.* 28, 45–63.

- Fukumori, I., Malanotte-Rizzoli, P., 1995. An approximate Kalman filter for ocean data assimilation: an example with an idealized gulf stream model. *J. Geophys. Res.* 100 (C4), 6777–6793.
- Gaspar, P., Grégoris, Y., Lefevre, J.-M., 1990. A simple eddy kinetic model for simulations of the oceanic vertical mixing: tests at Station Papa and long-term upper ocean study site. *J. Geophys. Res.* 95 (C4), 16179–16193.
- Gelb, A., 1974. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, pp. 374.
- Gunson, J., Oschlies, A., Garcon, V., 1999. Sensitivity of ecosystem parameters to simulated satellite ocean color data using a coupled physical–biological model of the North Atlantic. *J. Mar. Res.* 57, 613–639.
- Hurtt, G.C., Armstrong, R.A., 1996. A pelagic ecosystem model calibrated with BATS data. *Deep-Sea Res.* 43, 653–684.
- Kalman, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.*, 95–108.
- Kasibhatla, P., Heimann, M., Rayner, P., Mahowald, N., Prinn, R.G., Hartley, D.E., 2000. Inverse methods in global biogeochemical cycles. In: *Geophysical Monograph Series*. AGU Publications, p. 14.
- Levitus, S., *World Ocean Atlas (1994) CD-ROM Data Set Doc.*, National Oceanic and Atmospheric Administration, Washington, DC.
- Lisæter, K.A., Johannessen, O.M., Drange, H., Evensen, G., Sandven, S., 2000. Comparison of modelled and observed sea ice concentration and thickness fields in the Arctic. *J. Geophys. Res.*, submitted.
- 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.
- Natvik, L.-J., Eknes, M., Evensen, G., 2001. A weak constraint inverse for a zero dimensional marine ecosystem model. *J. Mar. Syst.* 28, 19–44.
- Oschlies, A., Garcon, V.C., 1998. Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean. *Nature* 394, 266–269.
- Pham, D.T., Verron, J., Roubaud, M.C., 1998. A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.* 16 (3–4), 323–340.
- Prunet, P., Minster, J.F., Echevin, V., Dadou, I., 1996. Assimilation of surface data in a one-dimensional physical–biogeochemical model of the surface ocean: adjusting a simple trophic model to chlorophyll, temperature, nitrate and pCO₂ data. *Global Biogeochem. Cycles* 10, 139–158.
- Sarmiento, J.L., Slater, R.D., Fasham, M.J.R., Ducklow, H.W., Toggweiler, J.R., Evans, G.T., 1993. A seasonal three dimensionnal ecosystem model of nitrogen cycling in the North Atlantic euphotic zone. *Global Biogeochemical Cycles* 7, 417–452.
- Smith, L.T., Boudra, D.B., Bleck, R., 1990. A wind-driven isopycnic coordinate model of the North and Equatorial Atlantic Ocean. 2. The Atlantic basin experiment. *J. Geophys. Res.* 95, 13105–13128.
- Verron, J., Gourdeau, L., Pham, D.T., Murtugudde, R., Busalacchi, A.J., 1999. An extended Kalman filter to assimilate satellite altimeter data into a nonlinear numerical model of the tropical pacific ocean: method and validation. *J. Geophys. Res.* 104, 5441–5458.