Stochastic estimation of biogeochemical parameters from Globcolour ocean color satellite data in a North Atlantic 3D ocean coupled physical-biogeochemical model

Maéva Doron[a], Pierre Brasseur[a,*], Jean-Michel Brankart[a], Svetlana N. Losa[b], Angélique Melet[c]

[a] *CNRS/UJF Grenoble 1, LGGE UMR5183, Grenoble, F-38041, France*
[b] *Alfred Wegener Institute for Polar and Marine Research, P.O. Box 120161, 27515 Bremerhaven, Germany*
[c] *Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey, USA*

Abstract

Biogeochemical parameters remain a major source of uncertainty in coupled physical-biogeochemical models of the ocean. In a previous study (Doron et al., 2011), a stochastic estimation method was developed to estimate a subset of biogeochemical model parameters from surface phytoplankton observations. The concept was tested in the context of idealized twin experiments performed with a 1/4° resolution model of the North Atlantic ocean. The method was based on ensemble simulations describing the model response to parameter uncertainty. The statistical estimation process relies on nonlinear transformations of the estimated space to cope with the non-Gaussian behaviour of the resulting joint probability distribution of the model state variables and parameters. In the present study, the same method is applied to real ocean colour observations, as delivered by the sensors SeaWiFS, MERIS and MODIS embarked on the satellites OrbView-2, Envisat and Aqua respectively. The main outcome of the present experiments is a set of regionalised biogeochemical parameters. The benefit is quantitatively assessed with an objective norm of the misfits, which automatically adapts to the different ecological regions. The chlorophyll concentration simulated by the model with this set of optimally derived parameters is closer to the observations than the reference simulation using uniform values of the parameters. In addition, the interannual and seasonal robustness of the estimated parameters is tested by repeating the same analysis using ocean colour observations from several months and several years. The results show the overall consistency of the ensemble of estimated parameters, which are also compared to the results of an independent study.

Keywords: Coupled physical-biogeochemical ocean model; North Atlantic; Parameter estimation; Stochastic method; Kalman filter; Anamorphosis; Satellite ocean colour data; Globcolor; SeaWiFS.

1. Introduction

Coupled physical-biogeochemical models (CPBM) of the ocean have been the subject of major developments during the past years with the goal of delivering hindcasts of the biogeochemical state of the ocean (e.g. Brasseur et al., 2009, Ford et al., 2012), short-term forecasts or

[*] Corresponding author: pierre.brasseur@legi.grenoble-inp.fr

longer-term simulations in relation with climate change (IPCC, 2007). However, the capacity of these CPBMs to realistically represent the observed variability of biogeochemical marine properties is still limited by the various sources of errors that occur in the simulations as a result of imperfect descriptions of the physical environment that drives the biology and empirical parameterizations of the biogeochemical interactions. From the seminal work of Fasham et al. (1990) to present day models, (e.g. Aumont et al., 2003), a common questioning about the development of biogeochemical models is the choice of state variables (each pool representing an organic or inorganic quantity) and the deterministic relationships that govern the fluxes between these pools. There is still debate about the minimal number of variables required to represent a given process and the optimal level of complexity of the model chosen to realistically simulate basic features such as primary production (Friedrichs et al. 2007 or Kriest et al. 2010).

A number of studies performed with CPBM involve the comparison with real-world observations. An special issue of the *Journal of Marine Systems* (Volume 76: *Skill assessment for coupled biological/physical models of marine systems*, 2009) was devoted to this question, describing a number of normalized tests to objectively assess the results of realistic simulations. Related to this question is the availability and quantity of data which can be used to validate model outputs. Ocean biogeochemistry can be characterized by different quantities, such as the concentration of chlorophyll *a*, micro and macro-nutrients, or particulate/dissolved organic/inorganic carbon/nitrogen. The main assets of *in situ* measurements are their accuracy and the possibility to sample the sub-surface water column at appropriate depths. However, such measurements are very sparse in space and time. Worldwide, very few sites in the open ocean are routinely monitored (e.g., on a monthly basis), whereas automatic sampling performed from floats or gliders is still a long-term prospective for biogeochemical quantities (Claustre et al., 2010). On the other hand, achieving a global coverage, repetitivity on a few days cycle and availability of the archive (more than a decade) are the main advantages of ocean colour satellite data. The main drawback is that the measurements from space are representative of surface quantities only.

The optimal combination of model results and observations is the field of data assimilation. A comprehensive review of data assimilation in CPBM is provided by Gregg (2008). A collection of studies have been performed quite recently using ocean colour data, which demonstrate the value of the assimilation concepts inherited from optimal control or estimation theory (Hemmings et al. 2003, 2004, Losa et al. 2004, Friedrichs 2002, Garcia-Gorriz et al. 2003, Ishizaka 1990, Armstrong and Sarmiento 1995, Natvik and Evensen 2003a,b, Gregg 2008, Nerger and Gregg 2007, 2008, Fontana et al. 2009, 2010, 2012, Ford et al. 2012, Hu et al. 2012, Mattern et al. 2012, Korres et al. 2012, Roy et al. 2012). Very often, variational data assimilation methods are implemented with the purpose of parameter estimation. Model simulations performed after optimisation are expected to yield better agreement with the measured data. Other studies that rely on sequential methods such as the Kalman filter, aim at a better control of the model trajectory and/or biogeochemical parameters.

In a recent paper, Doron et al. (2011) proposed an ensemble approach to objectively estimate a subset of biogeochemical parameters using ocean colour observations. In that paper, three key biogeochemical model parameters were identified to have the strongest impact on the chlorophyll *a* concentrations. Assuming independent parameter uncertainties for the different ecological provinces of the North Atlantic basin, Monte Carlo experiments were carried out to explore the model sensitivity to the uncertain model parameters. More precisely, an ensemble of

200 one-month simulations were performed during the spring bloom, to describe the joint probability distribution of the model state and model parameters. This probability distribution was observed to non-Gaussian, so that the linear observational update formulas of the Kalman filter were not directly applicable. For this reason, a nonlinear transformation (anamorphosis) was introduced to restore the Gaussianity of the marginal probability distribution for every model state variable and every uncertain parameter. With this upgraded description of the prior probability distribution, parameter corrections were computed from synthetic observations of the phytoplankton concentration (in the context of twin experiments). The nonlinear transformation has proved to be a key ingredient to obtain reliable estimates of the uncertain parameters in the various ecological provinces of the North Atlantic (each one with a specific non-Gaussian behaviour).

The twin experiments of Doron et al. (2011) provided an idealized framework where the only error source lies in the parameter values while all other elements are perfectly controlled. However, the error sources in real-world simulations may have different origins: the model initial conditions, the probability distribution function of the parameters, the physical forcings, or even observation errors being possibly the largest. Therefore, in a realistic case where real world observations are assimilated, the assumptions that were necessary to develop the method are not strictly valid anymore. For all these reasons, a real challenge in using a biogeochemical parameter estimation method with real ocean colour observations is to reduce the uncertainty in the biogeochemical parameters of CPBM and to objectively quantify the spread of the posterior distribution.

In this context, the objective of the present paper is to investigate to what extent the estimation method described in Doron et al. (2011) can be implemented to obtain realistic corrections of a few biogeochemical parameters using real ocean colour observations.The present paper is structured as follows: section 2 presents the coupled model, the ocean colour data and the parameter estimation method. Section 3 describes the results of the experiments and provides an assessment of the estimated parameters. In particular, the optimised parameters are compared to a regional set of biogeochemical parameters obtained in the frame of an independent approach (Losa et al., 2004). A summary, conclusions and future research lines are proposed in Section 4.

2. Model, data and estimation method

2.1. The coupled physical-biogeochemical model

The physical component of the coupled model is the primitive equation ocean circulation model OPA-NEMO (Madec et al. (1998), http://www.nemo-ocean.eu/), implemented in the North Atlantic basin between 20°S and 80°N, and between 98°W and 23°E with a horizontal resolution of 1/4° (i.e., the Drakkar NATL4 model configuration as described in Barnier et al. 2006). The vertical coordinate is geopotential with 46 prescribed levels enabling a resolution from 6 to 15 meters in the upper 100 meters. The physical model computes the advection and diffusion terms of the biogeochemical variables included in the model LOBSTER model (Lévy et al. (2005)).

LOBSTER describes the dynamics of the first trophic level of the food chain using the following six variables: nitrate (NO_3), ammonium (NH_4), phytoplankton (P), zooplankton (Z), dissolved organic material (DOM) and particulate detritus (D). The evolution of the prognostic variables at each point of the 3D model grid is based on their respective concentrations expressed in nitrogen unit (in $mmolN\ m^{-3}$), while the fluxes between the pools are governed by biogeochemical parameters and limiting environmental factors such as light and temperature.

The physical and biogeochemical models are coupled on-line. Eq. 1 is the advective-diffusive equation of the coupled model written for P, where $\mathbf{u}$ is the velocity, k_z is the vertical diffusivity and D_{lat} is the lateral diffusion. S_{P} designates the source minus sink term which is specific to the LOBSTER model. When it is written for the phytoplankton (see Eq. 2), the three terms of the right hand side correspond respectively to the growth of phytoplankton, the grazing of phytoplankton by zooplankton and the mortality of phytoplankton. These three terms correspond to the sources and sinks of phytoplankton in the biogeochemical model and each term is governed by one rate parameter: μ is the maximal growth rate for phytoplankton, m is the mortality rate for phytoplankton and g is the maximal grazing rate for zooplankton (all in s^{-1} or day^{-1}). The λ_μ and λ_g are non-dimensional functions modulating the growth and grazing rate according to the concentrations of $\mathrm{NO_3}$, $\mathrm{NH_4}$, Z, D and I, the light intensity. There is no explicit dependence on temperature in the present version of the model.

$$\frac{\partial \mathrm{P}}{\partial t} + \nabla \cdot (u\mathrm{P}) = \frac{\partial}{\partial \mathrm{z}}(\mathrm{k_z}\frac{\partial \mathrm{P}}{\partial \mathrm{z}}) + \mathrm{D_{lat}} + \mathrm{S_P} \tag{1}$$

$$S_{\mathrm{P}} = \mu \mathrm{P} \lambda_\mu(\mathrm{NO_3}, \mathrm{NH_4}, \mathrm{I}) - \mathrm{gP}\lambda_{\mathrm{g}}(\mathrm{P}, \mathrm{Z}, \mathrm{D}) - \mathrm{mP} \tag{2}$$

The chlorophyll a concentration ([chl$a$] in $mg\ chl\ a\ m^{-3}$) is a diagnostic variable computed from P, a prescribed chlorophyll a-to-nitrogen ratio and taking into account the Photosynthetically Available Radiation (PAR, in $W\ m^{-2}$) as described here below :

$$[chl\,a] = R_{Chl:N}P \tag{3}$$

$$R_{Chl:N} = max(R^{min}_{Chl:N}, R^{max}_{Chl:N}(1 - \frac{R^{max}_{Chl:N}}{R^{min}_{Chl:N}}(\frac{\overline{PAR}}{PAR_{max}}))) \tag{4}$$

with $R^{min}_{Chl:N}$= 1 $mg\ chla/mmolN$, $R^{max}_{Chl:N}$= 2.62 $mg\ chla/mmolN$ and PAR_{max}= 5 $W\ m^{-2}$.

In Ourmières et al. (2009), the parameterization of the biogeochemical model was tuned in the NATL4 configuration as described in Lévy et al. (2005), and the model was further validated using uniform values for the triplet μ (0.60 day^{-1}), m (0.05 day^{-1}) and g (0.80 day^{-1}).

The CPBM described here above has been considered as a good benchmark in a number of earlier studies, such as Ourmières et al. (2009), Béal et al. (2010), Doron et al. (2011) and Fontana et al. (2012), making its choice fully consistent with our present goal of parameter uncertainty reduction. These papers addressed different aspects of data assimilation relatively to the CPBM: assimilation of physical quantities and of nutrients; sensitivity study to the physical forcing and the corresponding model response; stochastic estimation of biogeochemical parameters and data assimilation of ocean colour satellite observations to update the multivariate model state. The present work extends the methodological exploration initiated by Doron et al. (2011) using the ocean colour data set described here after.

2.2. Ocean colour observations

Among the quantities that can be retrieved from space is the ocean colour or Sea Surface Reflectance, which is obtained from passive measurements in the visible and near infrared wavebands. The signal can be inverted to obtain several parameters including the chlorophyll *a* concentration (noted [chl *a*] here after), but also the backscattering coefficient, the absorption coefficient, the dissolved organic matter (Maritorena and Siegel 2005, IOCCG 2006) or the ocean transparency.

The SeaWiFS sensor was launched in 1997 by NASA as part of the Orbview mission, followed in 2002 by the MODIS sensor (launched by NASA onboard the Aqua spacecraft) and MERIS (launched by ESA onboard the Envisat platform). During the sensors lifetime, a series of data processing standards were applied to follow the upgrade of the sensor characterization and the improvement in the algorithms used to transform the radiance measurements into [chl *a*]. The Globcolour project (www.globcolour.info) aimed at providing a consistent merging of ocean colour data from these three main ocean colour sensors. The observations considered in the present study are the Globcolour data at 25 km resolution, which is roughly the resolution of the CPBM. In the open ocean, the quantity [chl *a*] is obtained with the so-called Garver-Siegel-Maritorena (GSM) algorithm, based on the results of Garver and Siegel (1997), Maritorena et al. (2002) and Maritorena and Siegel (2005). The [chl *a*] product is derived from the SeaWiFS observations alone between 1998 and 2002 and from the three sensors afterwards.

2.3. Parameter estimation

Doron et al. (2011) proposed a stochastic approach to perform parameter estimation using ocean colour data together with ensemble simulations. The method is derived from the Kalman filter and includes an augmented state vector approach to perform a joint analysis of the state variables and model parameters. The analysis step is performed after anamorphic transformation of the prior ensemble into a Gaussian distribution to maximize the information extracted from the data (Brankart et al. (2012)). This methodology was developed by Doron et al. (2011) in an idealised framework based on twin experiments. The objective was to investigate whether surface phytoplankton observations could be exploited to reduce the uncertainty on biogeochemical parameters and provide a posterior probability distribution that objectively reflects the gain in confidence. The various implementation steps of the method are briefly recalled hereafter.

2.3.1. The augmented vector

In the biogeochemical model, the state vector includes the following variables : $\mathbf{x}$=[CHL, P, Z, NO_3, NH_4, D, DOM], where CHL is the discretized [chl *a*] concentration and the six other state variables are the discretized LOBSTER variables. The three biogeochemical parameters μ, m and g are considered uncertain in the present study and the main source of error for the model. They are written under the form of the vector of parameters $\alpha = [\mu, m, g]$. The model response depends not only on the physical forcing, but also of α, which can be written as $\mathbf{x}$=$\mathbf{x}(\alpha)$. The augmented state vector, including both variables and parameters can be written as follows: $\hat{\mathbf{x}} = [\mu, m, g$, CHL, P, Z, NO_3, NH_4, D, DOM$] = [\alpha, \mathbf{x}(\alpha)]$.

2.3.2. The biogeochemical parameters as a source of uncertainty

The LOBSTER model is a relatively simple model as it only includes a single phytoplankton group. Other models including more complex biogeochemical processes are potentially required to faithfully describe a variety of ecosystem functioning, as effectively observed at the scale of large oceanic areas. This issue was objectively tested and discussed for a dozen models by Friedrichs et al. (2007). Our interest in having a relatively simple model is its implementation cost. The approach proposed by Doron et al. (2011) to represent different ecosystem regimes with a simple model is to allow spatial variations of biological model parameters. Regional values are thus specified for a number of key parameters, increasing thereby the number of degrees of freedom of the model solutions. The regional values of the parameters are considered as fundamentally uncertain, and our goal is to reduce this uncertainty by performing stochastic estimations based on ocean colour data. The three key rates here are governing the growth, grazing and decay of phytoplankton. The ecological provinces proposed by Longhurst (1995, 2007) are the basis for the definition of the regions considered in Doron et al. (2011) and in this study. We assume that all regions have independent values of α. Thus, the parameters are uniform over each of the M= 13 regions (see Table 1), instead of being uniform on the NATL4 domain, as in previous standard LOBSTER implementations.

A sensitivity study to test the model response to the regionalisation of the biological model parameters is described by Doron et al. (2011) with an ensemble of N=200 simulations. For each of the N simulations, a map of parameters (actually 39 parameters) is sampled from a predefined distribution. The statistics of these parameters is not well known, but a strong constraint on the values is their positivity. A Gamma distribution $\Gamma(k, \theta)$ depending on two parameters (the shape parameter k and the scale parameter θ), is chosen so that the 95% percentile is equal to 2.1 times the reference value. For every parameter $\lambda = \mu$, m or g in every province, $p(\lambda/\bar{\lambda}) = \Gamma(k, \theta)$ with $k = 4.236$ and $\theta = 0.309$, where $\bar{\lambda}$ is the mode of the distribution. Overall, there are M times 3 parameters = 39 degrees of freedom.

The initial state of the model is the same for all N simulations and corresponds to the model state of 16^{th} April 1998 in the free run described in Ourmières et al. (2009). The only difference between the ensemble members is thus α. The N simulations were integrated for a duration of 30 days. By the end of the simulation at the date of the 16^{th} of May 1998, the model results are extracted to analyse the surface phytoplankton (P) concentrations in relation to the parameter perturbations. The model response is strongly nonlinear and region-dependent, as described in Doron et al. (2011). For instance, in some regions, P is strongly correlated with μ, while it is closely related to m or g in other regions. Weak correlations also happen between P and one or two parameters. The amplitude in the variations of P is sometimes very different from one region to another. In addition, the distributions of P are generally non-Gaussian (For more details about the ensemble and the model response to the parameter perturbation, the reader is refered to Doron et al. (2011)). The ensemble used here after and in Doron et al. (2011) are the same, since the motivation of the two studies is the same, i.e. to perform stochastic parameter estimation from surface phytoplankton data, either in the context of twin experiments or real-world data.

2.3.3. Anamorphic transformation of the augmented state vector

The Kalman filter with an augmented state vector performs sequential updates of the state variables and of the parameters using a linear combination of the model forecast and the observations. The linear observational update is optimal in the case of a linear model and Gaussian error statistics. In the present case, the nonlinear model response to the uncertainty in the biological parameters and the resulting non-Gaussian PDFs has clearly been demonstrated. As a result, the optimality conditions of the Kalman filter in the sense of maximum likelihood esti-

Province	Short name	Long name
1	BPLR	Boreal Polar
2	ARCT	Atlantic Arctic
3	SARC	Atlantic Subarctic
4	NADR	North Atlantic Drift
5	GFST	Gulf Stream
6	NASW	North Atlantic Subtropical Gyral
7	NATR	North Atlantic Tropical Gyral
8	WTRA	Western Tropical Atlantic
9	ETRA	Eastern Tropical Atlantic
10	SATL	South Atlantic Gyral
15	NWCS	North West Atlantic Shelves
17	CARB	Caribbean
18	NASE	North Atlantic Subtropical Gyral

Table 1: List of the Longhurst-defined ecological provinces located in the North Atlantic. The number of the provinces, their short and long names are those proposed by Longhurst (1995).

mator are obviously not satisfied. To overcome this issue, a so-called anamorphic transformation of the augmented state variables is implemented to perform the analysis step in a space where the assumption of Gaussian PDFs becomes valid.

The idea of anamorphosis was first applied to data assimilation problems by Bertino et al. (2003). Subsequent applications of anamorphosis have been proposed by Simon and Bertino (2009), Simon and Bertino (2012) or Béal et al. (2010). The present implementation is essentially based on the scheme developed by Béal et al. (2010).

Doron et al. (2011) proposed a method based on univariate transformations of the model variables and uncertain parameters performed independently at each grid point. The random variable, noted as x, is either μ, m, g or CHL. For μ, m and g, the marginal distribution is a Gamma distribution, while for CHL a sample of the marginal distribution is taken from the ensemble of 200 simulations as described is Section 2.3.2. The transformed variable y is derived from x as $y = f(x)$, in such a way that y has a marginal distribution close to $\mathcal{N}(0,1)$. The function f is the anamorphosis transformation and is built from the ensemble itself (see Eq. 5):

$$f(x) = \begin{cases} z_1 & \text{for } x < x_1 \\ z_k + \frac{z_{k+1}-z_k}{x_{k+1}-x_k}(x - x_k) & \text{for } x_k \leq x \leq x_{k+1} \\ z_p & \text{for } x > x_p \end{cases} \tag{5}$$

where $x_k, k = 1, \ldots, p$ are the p percentile of x such that $p(x < x_k) = r_k$, for the set of values $r_k = 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.92, 0.94, 0.96, 0.98$ and z_k are the corresponding percentiles of $\mathcal{N}(0,1)$. This makes a significant difference compared to Bertino et al. (2003) who prescribe the transformation as a pre-defined analytical function which doesn't depend on the ensemble distribution.

The function f is only uniequivocal on the range $[x_1, x_p]$ so that the reciprocal function g is only defined on the range $[z_1, z_p]$ (see Eq. 6). With the anamorphosed variables, all values of CHL and parameters are between -3 and 3, because the physical quantities have been projected on the Gaussian percentiles.

$$g(z) = \begin{cases} x_1 & \text{for } z < z_1 \\ x_k + \frac{x_{k+1}-x_k}{z_{k+1}-z_k}(z - z_k) & \text{for } z_k \leq z \leq z_{k+1} \\ x_p & \text{for } z > z_p \end{cases} \tag{6}$$

Brankart et al. (2012) evaluated the effect of the present local anamorphic transformations on spatial correlations for various kinds of ocean uncertainties. Their results indicate that this transformation is accurate enough to faithfully preserve the correlation structure if the distribution is already close to Gaussian; however the transformation has the general tendency of increasing the correlation radius as soon as the dependence between random variables becomes nonlinear.

2.3.4. The estimation algorithm

In the present study, the Kalman filter analysis equations are combined with the augmented state vector approach $\hat{\mathbf{x}}$ and the nonlinear transformations of all quantities prior to the calculations of the observational update. Equations 7, 8 and 9 are the basis for our calculations.

$$\hat{\mathbf{x}}^a_{\mathbf{an}} = \hat{\mathbf{x}}^f_{\mathbf{an}} + \hat{\mathbf{K}}_{\mathbf{an}}(\mathbf{y}_{\mathbf{an}} - \hat{\mathbf{H}}\hat{\mathbf{x}}^f_{\mathbf{an}}) \tag{7}$$

$$\hat{\mathbf{K}}_{\mathbf{an}} = (\hat{\mathbf{H}}\hat{\mathbf{P}}^f_{\mathbf{an}})^{\mathbf{T}}(\hat{\mathbf{H}}\hat{\mathbf{P}}^f_{\mathbf{an}}\hat{\mathbf{H}}^{\mathrm{T}} + \mathbf{R}_{\mathbf{an}})^{-\mathbf{1}} \tag{8}$$

$$\hat{\mathbf{P}}^a_{\mathbf{an}} = (\mathbf{I} - \hat{\mathbf{K}}_{\mathbf{an}}\hat{\mathbf{H}})\hat{\mathbf{P}}^f_{\mathbf{an}} \tag{9}$$

In these equations, $\hat{\mathbf{x}}^f_{\mathbf{an}}$ is the model forecast (augmented with the parameters) and $\mathbf{y}_{\mathbf{an}}$ is the observation vector, both after anamorphosis transformation. $\hat{\mathbf{H}}$, the observation operator, is determined by the availability of the observations. The matrix $\hat{\mathbf{K}}_{\mathbf{an}}$ is the Kalman gain for the augmented vector with the anamorphosed variables.

The link between the model output and the available [chl a] products is made with the observation operator $\hat{\mathbf{H}}$. A first issue with real ocean colour products is the cloud coverage, which randomly results into unobserved grid points especially in Northern regions during the winter season. An additional difficulty with biogeochemical variables is that the observed quantities do not exactly correspond to the prognostic biogeochemical variables. Here, the observations correspond to [chl a], which is linked through Eqs. 3 and 4 to the modelled P. Another issue with the use of real ocean colour data is the specification of the observation error matrix $\mathbf{R}$, which is discussed for instance in Ford et al. (2012). The different options considered to tackle these issues in the present study are discussed in Section 3.

The augmented forecast error covariance matrix $\hat{\mathbf{P}}^f_{\mathbf{an}}$ is decomposed in reduced-rank form as $\hat{\mathbf{P}}^f_{\mathbf{an}}= \hat{\mathbf{S}}^f_{\mathbf{an}}\hat{\mathbf{S}}^f_{\mathbf{an}}{}^{\mathrm{T}}$ and is evaluated using the 200-member ensemble described as follows. Multivariate EOFs are calculated based on the 200 simulations, by considering all biogeochemical variables and the three regional parameters after anamorphosis transformation. The 30 dominant eigenmodes compose the matrix $\hat{\mathbf{S}}^f_{\mathbf{an}}$. The observation error $\mathbf{R}_{\mathbf{an}}$ is written in the form $\mathbf{R}_{\mathbf{an}} = rI$, where r is dimensionless. Its value is chosen to be very low ($r = 0.0001$ in anamorphosed space), to give the maximum confidence in the transformed observations. The augmented analysed error covariance matrix $\hat{\mathbf{P}}^f_{\mathbf{an}}$ is thus calculated with anamorphosed quantities.

The quantity $\hat{\mathbf{x}}^a_{\mathbf{an}}$ is calculated according to the analysis equations. Being an augmented vector, it includes not only the six prognostic variables of the model (NO_3, NH_4, DOM, P and

Z), but also the diagnostic variable CHL and the three values for the parameters μ, m and g. Since it was obtained with anamorphosed quantities, it needs to be transformed back from its anamorphosed values to its physical values before being used in the posterior model.

3. Results

In this section are shown the results obtained with the implementation of the stochastic parameter estimation using real satellite ocean colour data. The results are organised in the following order: first, the analysis results, both for the parameters and [chl *a*], are considered when the observation data is from May 1998. Second, the method is extended to ocean colour observations captured for other time periods and the impact on parameters is assessed. Finally, the estimates of the three key biogeochemical parameters obtained with our method is compared to a totally independent parameter dataset.

3.1. Analysis for May 1998

The surface chlorophyll concentrations simulated by the model for the 16^{th} of May 1998 using reference parameters (i.e. the so-called reference simulation) is shown in Fig. 1 (top left panel). At the basin level, there is a large range of [chl *a*] variations, from less than 0.01 $mg\ chl\ a\ m^{-3}$ to more than 10 $mg\ chl\ a\ m^{-3}$, i.e. over more than 3 orders of magnitude. The anamorphosis as described in Section 2.3.3 is used here to define the anamorphosed variables. The top right panel of Fig. 1 shows the anamorphosed values of [chl *a*], in which all quantities are between -3 and 3, by construction. However, almost all [chl *a*] model values are in the range [-0.2, 0.4]. Basically, it means that the simulation with reference parameters is quite close to the median of the 200 simulations. The [chl *a*] in coastal areas should not be considered here, since the coastal areas are not included in our sensitivity study on biogeochemical parameters. The bottom left panel of Fig. 1 shows observed [chl *a*] obtained from ocean colour sensors, averaged for the month of May 1998. The same colour scale is applied for both observations and model results. In oligotrophic regions, the oligotrophy is more pronounced in the model than in observations (e.g. in subtropical gyres), whereas the simulated bloom leads to larger concentrations than observed (e.g. in subpolar gyres). The anamorphosis of the observations is based on the same ensemble and set of equations as the model (bottom right panel of Fig. 1). At mid-latitudes, anamorphosed [chl *a*] is in the range [-1.2, 1.8]. In oligotrophic regions, anamorphosed [chl *a*] hits the top of the range of [chl *a*] values. An interesting feature of the anamorphosed maps is that [chl *a*] concentrations become more easily comparable. One can see for instance, that although modelled [chl *a*] and observed [chl *a*] appear quite different in the left column (with their physical values), the same maps with the anamorphosed values show more details by ranking the concentrations at their respective position in the ensemble.

The analysis with the anamorphosed variables $\hat{\mathbf{x}}^a_{\mathbf{an}}$ is computed using Eqs. 7, 8 and 9 and the monthly mean [chl *a*] observations for May 1998. Figure 2 shows the portion of $\hat{\mathbf{x}}^a_{\mathbf{an}}$ that contains the parameters. In five regions, the analysed parameter is larger than the reference value (regions in red) whereas eight regions have analysed μ lower than the uniform value (regions in blue). The anamorphosed m (middle right panel) is lower than the reference values for all regions but two (Gulf Stream and Atlantic Arctic). As for g, the analysis provides values either larger (8 regions) and lower (7 regions) than the initial value. After being transformed back to their physical values, these three rates are positive by construction, and the corresponding maps are shown in the left column of Fig. 2. The figures show that the corrections occur in both directions, towards larger and smaller values depending on the region. One can identify clusters of similar behaviour for the regional parameters. For instance, analysed values of μ for the Arctic are larger than the reference (regions 1, 2 and 3). The mortality rate m has analysed values lower than the reference in all regions but two (regions 2 and 5). During the analysis process, it happens that very low values of anamorphosed m are obtained (for instance, in regions 8 and 15, see middle right panel). One practical positive aspect of the anamorphosis is the fact that it prevents the estimation of irrealistic negative values of physical (here biological) quantities. Indeed, this may occur with the Kalman filter equations through the linear combination of model estimates and observations. Here, even very low anamorphosed variables are transformed back to physical variables with realistic, positive values. It is included by construction in the equation of the inverse function (see Eq. 6). The lowest value of the physical variables for the analysed parameter is indeed the lowest value for the region sorted in the ensemble. These results show that, starting with a triplet of parameters that are constant at basin scale, it is possible to objectively estimate the spatial distribution of the 3 parameters by combining ensemble model outputs and ocean colour data.

Figure 3 shows the analysed [chl a] computed according to Eq. 7. It corresponds to the CHL portion of $\hat{\mathbf{x}}^a_{\mathbf{an}}$. The top right panel is the analysis in the anamorphosed space in which all calculations are performed. In the top left panel, [chl a] is transformed back into the physical space, illustrating how the analysis is constrained by the observations displayed in the bottom line of Fig. 1. At high latitudes, [chl a] is strongly reduced with respect to the reference model and exhibits distribution patterns that are similar to ocean colour data. The tongue of low concentrations between the oligotrophic pool and Togo has been corrected. The anamorphosed map (Fig. 3, top right) better agrees with the anamorphosed observations (Fig. 1, bottom right) then the reference model simulation did.

To evaluate the reduction of model errors due to the optimization of the biogeochemical model key parameters, a new simulation is performed. This free simulation uses the set of regionalised parameters α from the analysis $\hat{\mathbf{x}}^a_{\mathbf{an}}$, is initialized in the same way as the previous simulations on 16^{th} of April 1998, and lasts until 16^{th} of May 1998. The map of [chl a] provided by this new simulation can be compared to the analysis, the data and the model results obtained with uniform parameters. The simulated [chl a] is similar to the analysis one at latitudes lower than 40°N. Similar patterns of oligotrophy and equatorial concentrations are observed in both physical and anamorphosed space. However, a region right at the equator exhibits patterns of too large concentrations. For latitudes larger than 40°N, there is a significant deviation between the two maps. The new run output (Fig. 3, bottom left) shows very large concentrations, particularly in the North-Eastern part of the basin where the concentrations from the new run are in the upper part of the ensemble, whereas the analysed values are in the lower part of the ensemble. Compared to the reference model run with uniform values of α, the new run yields increased concentrations of [chl a] for latitudes larger than 40°N. For lower latitudes, the new run reproduces an oligotrophic region with too low values, but the latitudinal and longitudinal extensions of this zone is closer to the oligotrophic zone in the observations than in the forecast. For latitudes lower than 10°N, the new run provides patterns that are similar to those observed although there is a spot of relatively large [chl a] around 0°N, 30°W. If the observations and the new run are considered in the anamorphosed space, the new run and the observations are fairly consistent except for high latitudes and the Western side of the basin.

A quantification of the reduction of the model error in terms of [chl a] can be performed by comparing the RMS of [chl a] in the anamorphosed space for the reference simulation, the analysed state, and the simulation using the optimal set of parameters (Fig. 4). The calculation of this norm in the anamorphosed space has been first introduced by Doron et al. (2011) for the following reasons. In the framework of data assimilation, the root mean square error (usually noted RMS) is the quantity to be minimized using a least square method. For Gaussian errors, this criterion is optimal. In the present case where the errors of the physical variables are not Gaussian, this criterion is no more optimal and the RMS is no more a good criterion of the accuracy of the estimation. With the anamorphosed variables, the errors are Gaussian and the RMS calculated on the anamorphosed variables is relevant and can be used to define an objective norm. The RMS with the anamorphosed values is calculated as follows: $RMS_{ana}(x,k) = \sqrt{\frac{1}{N_k}\sum_{i,j,\text{in region k}}(y_{i,j}^{obs} - y_{i,j}^{model})^2}$ where y is the anamorphosed value for x. Since it is calculated with the anamorphosed variables, it is dimensionless. Using this norm, all regions exhibit relatively similar RMS values. This would not be the case if [chl a] were expressed into physical values (in $mg\ chl\ a\ m^{-3}$) since [chl a] spans at least 3 orders of magnitude.

For every regions, the [chl a] inferred from the analysis step is closer to observations than the [chl a] before the analysis, i.e. in the reference simulation (comparing the values of the misfits in red and in green in Fig. 4). This objectively shows that the analysis performs well and improves the [chl a]. The regionalised set of parameters obtained in the analysis clearly improves the simulated [chl a] relative to the reference simulation (comparing red and blue misfits in Fig. 4). Indeed, misfits of [chl a] from the observations are systematically lower than the initial distance, except for region 3 where it is twice larger. In regions 2, 4 and 6, the RMS is the same between the observations and the forecast, as between the observations and the new run, which means that for these regions, the regionalised parameters are not able to improve the estimation of [chl a] after one month. With the present approach, the optimised parameters might just compensate for other possible model deficiencies such as uncertainties in internal parameterisations either physical or biogeochemical processes on their own. Thus, the combined parameter and state estimation to allow for the model errors distinct from the parameter uncertainties, as done in Losa et al (2004) and mentioned in the conclusion, seems faithful future research direction. From this figure, one can claim that the estimated parameters perform successfully since they provide, for most regions, a simulation closer to observations than the forecast with uniform parameters.

In Doron et al. (2011), it was possible to calculate the true error committed in the parameter estimation, since in the twin experiments framework the reference values of the parameters to be retrieved by data assimilation were known. In general, the three parameters were successfully retrieved, using the nonlinear stochastic estimation. However for a few cases, some parameters are more difficult to estimate than others (see their Figure 8). This is the case for μ in regions 5 and 6, and m in regions 5 and 15. A possible explanation is that regions 5 and 15 have a small area and are influenced by advection from other surrounding areas. Nevertheless in the present parameter estimation, there is a clear improvement in the [chl a] after one month for regions 5, 6 and 15, as can be seen from Fig. 4.

3.2. Analysis performed with different monthly datasets

The previous section showed the results obtained for the regionalisation of parameters when the dataset is the monthly mean of [chl a] observations for May 1998. This is the first year where the spring bloom was observed with modern era ocean colour satellites, but the measurements actually span a longer timeframe: from 1998 to present time. To take advantage of this extended observation period, we applied the stochastic parameter estimation method for different observation dates. However, the ensemble is kept identical for all these analyses because of the huge computer time needed to generate the 200 ensemble members. Keeping the same ensemble statistics to perform different analyses using different datasets is a quite common approach in data assimilation. For instance, the sub-optimal EnOI simplification of the Ensemble Kalman filter relies on a similar concept. In the present case, it also reflects the difficulty to generate robust ensembles that would faithfully describe the variability of error statistics at seasonal and interannual timescales.

To illustrate the temporal evolution of [chl a] as observed with ocean colour satellite sensors, Fig. 5 shows the monthly maps for March, April and June 1998 both in physical and anamorphosed space. One can see the evolution of the spring bloom: faintly appearing around 40°N in March (first line, left), then its displacement in the Gulf Stream drift and south of Iceland in April. In May, the bloom is general for latitudes north of 50°N, with a greater norther extension in June (Fig. 5, bottom left panel). Conversely, the Northern oligotrophic region becomes more oligotrophic from March to June with decreasing [chl a] in its inner part. When observing this temporal evolution with the anamorphosed variables, one can see the fluctuations of the observed values in relation with the local range spanned by the ensemble. This can be seen almost everywhere in the model domain, except in the oligotrophic regions (where the observations are systematically larger than the model outputs) and for the equatorial band (in the Eastern part, where the observed values are usually lower than the model outputs).

A number of analysis is then performed as described by Eq. 7, Eq. 8 and Eq. 9, using a few dozens of observations extracted from the Globcolour monthly averages for 4 spring months (March, April, May and June) and for 13 years (1998 to 2010). Consistently with the use of the same ensemble statistics for all analyses, the background state is also the same as in the May 1998 experiment (Section 3.1). While the initial state is best suited for the May 1998 dataset, it obviously represents a less accurate information for the other datasets, but it is still good enough to perform meaningful statistical estimations. In the case where the model simulates a too early spring bloom, observations for June are relevant and in places where the spring bloom is modelled too late, the observations of April are relevant. Thus, we consider the ocean colour observations from March to June, to span these different situations.

Using the same protocol as in the previous experiment, all quantities are converted into anamorphosed variables before computing the joint state/parameter analysis step. To comment on the biological rates in their physical units, they are transformed back in their physical values. The resulting values for μ, m and g parameters estimated for the different ecological regions in March, April, May and June of the 13 years are shown in Figs. 6, 7 and 8 respectively. While the different observations used to estimate the parameters exhibit seasonal and interannual variations, the values of the estimated μ fall into a reduced range for most regions (Fig. 6). Values differ from regions to regions, for instance, in the Arctic regions (1, 2 and 3) they are larger than the initial value. In region 1, μ first increases from March to May, then decreases in June. In regions 6 and 7, μ decreases from March-April to May-June. All other regions exhibit μ values lower than the initial value, with minor changes depending on the month. Figure 6 clearly shows a strong consistency between the months and years, with a slight temporal evolution of the parameter values between March and June. The values of the parameters are stable throughout the years, suggesting that a perpetual yearly cycle of parameter values could be adopted in future experiments.

Figure 7 shows similar results for the m parameter. For m, the most significant feature of the analysis is the consistency throughout the years, similarly to what was observed for μ. The largest scatter are observed for regions 2 and 5. For regions 7 to 18, the scatter is very small and the obtained values are very close to the minimal possible value. Regions 1 to 6 show larger variations either for a given month or for the seasonal trend. For instance in region 1, m increases from March to May, then decreases. For region 2, m is larger in March than for other months. For region 5, m steadily increases during the four months. Regions 3, 4 and 6 shows less clear variations.

Figure 8 shows the results of the 52 analysis for the grazing rate g, again sorted by month (March to June), thus showing the seasonal variations of g. The interannual variations appear through the scatter (13 analyses per a given month). Regional variations of g are noticeable. Some regions have g values systematically lower than the initial value, such as regions 5, 8, 15 and 17. Other regions experience g values systematically larger than the initial values, such as regions 1, 6, 7 and 9. The other regions experience larger variations in g, from values larger and lower than the reference. In the Arctic part of the domain (regions 1, 2 and 3), there is a contrasted evolution of the grazing rate g during the four months. Analysed g is quite large for region 1 (from March to June), it decreases steadily for region 2 (from March to June) and decreases slowly for region 3 during the same period. Region 4 experiences a slight decrease in g from March to June, as region 10. On the opposite, region 18 has growing values for g from March to June. These figures show a contrasted temporal parameter evolution for different trophic regions, while keeping the consistency among the different years.

In a nutshell, we applied the previous method to obtain regional parameter values with different observations (four months times 13 years). Regional variations of the parameter values are clearly observed for the three parameters μ, m and g. Seasonal variations between March and June are obvious for most regions, while other regions exhibit less marked seasonality of the parameter values. An interannual variability can be seen through the scatter obtained on the 13 years of observations. But, the overall consistency between the different years for a given region and a given parameter is an important aspect, which paves the way to the implementation of a regionalised set of parameters, possibly including their annual cycles. It shows that the interannual variability is relatively small and could be neglected as a first approximation in the prospect of future implementations.

3.3. Comparison with independent estimates

In order to consolidate the results obtained from the previous expriments based on real ocean colour data, it is interesting to compare our approach with another totally independent study also dedicated to parameter estimation. The goal of this additional investigation is not to identify more precisely what would be the most likely value of the parameters, but rather to check the overall consistency of our estimates with another similar modelling exercise. Losa et al. (2004) investigate the estimation of a few parameters (actually six) in a biogeochemical model of the North Atlantic. Their estimation relies on a weak constraint assimilation method implemented with a simple NPZD biological model similar to LOBSTER. The assimilated data are the monthly averaged ocean colour from the Coastal Zone colour Scanner (CZCS), which was an ocean colour satellite sensor in activity between 1979 and 1986. The geographical area under consideration is the North Atlantic between 30°N and 60°N. The domain is separated into independent 5° by 5° boxes. Their six optimised parameters were the maximum phytoplankton growth rate, the initial slope of the production versus light intensity (P-I) curve, the maximum specific rate for grazing, the half-saturation constant for grazing, the maximum specific mortality rate for phytoplankton, and the maximum specific mortality rate for zooplankton. The result of this study consisted in regional values of the six parameters for the basin with a spatial resolution of 5°. The obtained parameter values are representative of the whole year, since the assimilated dataset covers the whole annual cycle from January to December.

Although the set of biological equations in Losa et al. (2004) differs from our set of LOBSTER equations, the two models basically simulate the same primary production process, i.e. the growth and decay of the first trophic level through photosynthesis. Three of their parameters are thus closely related to the three parameters studied in the present paper: their maximum specific rate for grazing corresponds to our grazing rate g, their maximum specific mortality rate for phytoplankton corresponds to our mortality rate for phytoplankton m and their maximum phytoplankton growth rate constant corresponds to our μ. It is thus possible to compare their spatial variations for these three parameters with our regionalised values, although the two processes of optimisation are quite different. More specifically, the two estimation methods are different: stochastic vs. variational. Our model is 3D and covers the latitudes from 20°S to 80°N, while their model is 0D and is restricted between 30°N and 60°N. Our spatial resolution is 1/4° with 13 Longhurst regions and theirs is 5°, with each square having its own set of 6 parameters. Our ocean colour observations are extracted from the Globcolour dataset, with three satellites providing the data between 1998 and 2010, while their observations are the Nimbus-CZCS data from 1979 to 1985. We estimate regional values of three parameters for four months and 13 years, while they optimise six parameters with a better spatial resolution but a single temporal value. For all these reasons, the two methods are independent and the results of both estimations should not be compared on a one-by-one basis, but we believe that meaningful comparisons could be done regarding for instance the overall spatial variability of the parameters.

Figure 9 shows the map of parameters obtained by Losa et al. (2004), with the same colour scale than our plot in Fig. 2 for the physical values. The first map corresponds to our growth rate μ, the second map to our mortality rate m and the third map to our grazing rate g. Their initial values for these three parameters are very close to ours: 0.6 day^{-1} for the growth rate for both studies, 0.05 day^{-1} for the mortality rate for both studies. Their first guess value for the grazing rate is 0.73 day^{-1} while our initial value was 0.8 day^{-1}. In Fig. 9, one can observe significant spatial variations of the three parameters at the basin scale. Some gradients in the values can be seen, for instance, the values of the growth rate are smaller in the North-Western part of the basin (to the South of our region 2) than for lower latitudes (to the North of our region 6). For the mortality rate, the values in the South of the basin are larger than in the northern part of the basin, thus showing a general North-South gradient. The grazing rate is on a reverse pattern with larger values in the North of the basin and smaller values in the South of the basin, although with some patchiness. Losa et al. (2004) found their results to be in qualitative agreement with other estimates. However, the differences in the phytoplankton mortality rates estimated in both studies could be explained partly by the fact that, in the LOBSTER model, the phytoplankton mortality parameterisation is linear while, in Losa et al. (2004) model, the process is presented in a saturated quadratic form. Given identical fluxes and phytoplankton concentration, the value of m in the linear formulation is always less than those in the alternative parameterisation. The less the phytoplankton concentration is, the bigger are the differences between the m estimates. Further, in the primary production parameterisation of the Losa et al. (2004) model, there are two main parameters: the maximum phytoplankton growth rate and the P-I curve initial slope, while in the LOBSTER parameterisation considered in the current study everything is controlled just by the maximum phytoplankton growth rate.

To go beyond this qualitative comparison with the Losa et al. (2004) parameters, we extracted histograms of the parameter values (see Fig. 10). Since the two studies are different, the two types of histograms have different meanings, but both represent uncertainty in the parameter values and illustrate their scatter. The histogram from our study for each region and each parameter is based on the 52 analyses (4 months times 13 years) and the results are shown in solid lines. The histogram from Losa et al. (2004) study is based on the values of the parameters extracted on the Longhurst regions. Each time a box from the grid 5° by 5° is partly included in a given region, the parameter value is considered for the histogram (the results are shown in dotted lines). The uncertainty illustrated with these histograms is either a temporal uncertainty (parameters optimised in the present study) or a spatial uncertainty (Losa et al. (2004) annual parameters but with a better spatial refinement). Fig. 10 shows the results and the two histograms for each parameter and for three regions. The histograms were calculated for three regions, 4 (NADR), 6 (NASW) and 18 (NASE), which are the main regions covered by the Losa et al. (2004) study. There are 26 values of the parameters estimated by Losa et al. (2004) for region 4, 22 for region 6 and 21 for region 18. The vertical black line is our initial value, which are the same than the initial value in Losa et al. (2004), except for the grazing rate (0.8 in our study versus 0.73 day^{-1} in their study). The vertical dotted line is the average of Losa et al. (2004) parameters for the region. The results in Fig. 10 are organised as follows: results for μ are in the first column, for m in the second column and g in the third column. The three parameters for region 4 are displayed on the first line, for region 6 on the second line and for region 18 on the third line.

In the top left plot, which represents the output for μ in the region 4, one can see that our analyses produced values systematically lower than our initial value. The Losa et al. (2004) optimisation produces growth rate values both larger and lower than their initial value as previously described. Some of their values are also represented in our study, but their histogram is more flat. The average value of Losa et al. (2004) growth rate in region 4 is smaller than our initial value and inside the range of our values. The same type of conclusion holds for μ in region 6. This time, the analysed values with the two methods are larger and smaller than the initial value, but there is consistency between the two histograms. The average of Losa et al. (2004) values is too close to the initial value (0.605 day^{-1}) to distinguish the two lines. For μ in region 18, our histogram is tilted towards values smaller than the reference whereas the Losa et al. (2004) average is larger than the reference. To sum up, the regional values of μ obtained with the two methods are overall in fairly good agreement.

For the parameter m, the comparison shows less consistency between the resulting two datasets. For the three regions, the histogram based on our values and the histogram based on Losa et al. (2004) values have only a few values in common, if any. The average Losa et al. (2004) value is outside our histogram. In the study of Doron et al. (2011), it was already noticed that the parameter m was the most difficult to retrieve. It can be seen from their Fig.8 that the RMS on m is quite large for regions 3, 4, 5, 8 and 15. This parameter m seems to be the most difficult one to estimate among the three parameters estimated here.

Results for the parameter g are closely consistent between the present study and Losa et al. (2004) study. The histograms have a rather large common range for the three regions 4, 6 and

18. The average value based on Losa et al. (2004) parameters are well inside the histograms.

Since our study and Losa et al. (2004) are different and totally independent, it was not at all expected that both estimated parameters datasets would be identical. But for the three parameters μ, m and g, the comparison between the results of both studies yield positive conclusions. The results for these three parameters are rather consistent: the scatter of the values of the parameters show some similarity for each region. The averages for μ and g based on Losa et al. (2004) results are generally within the range of our estimated parameters. This gives confidence in our results and paves the way to future estimation of regional values for coupled physical-biogeochemical parameters.

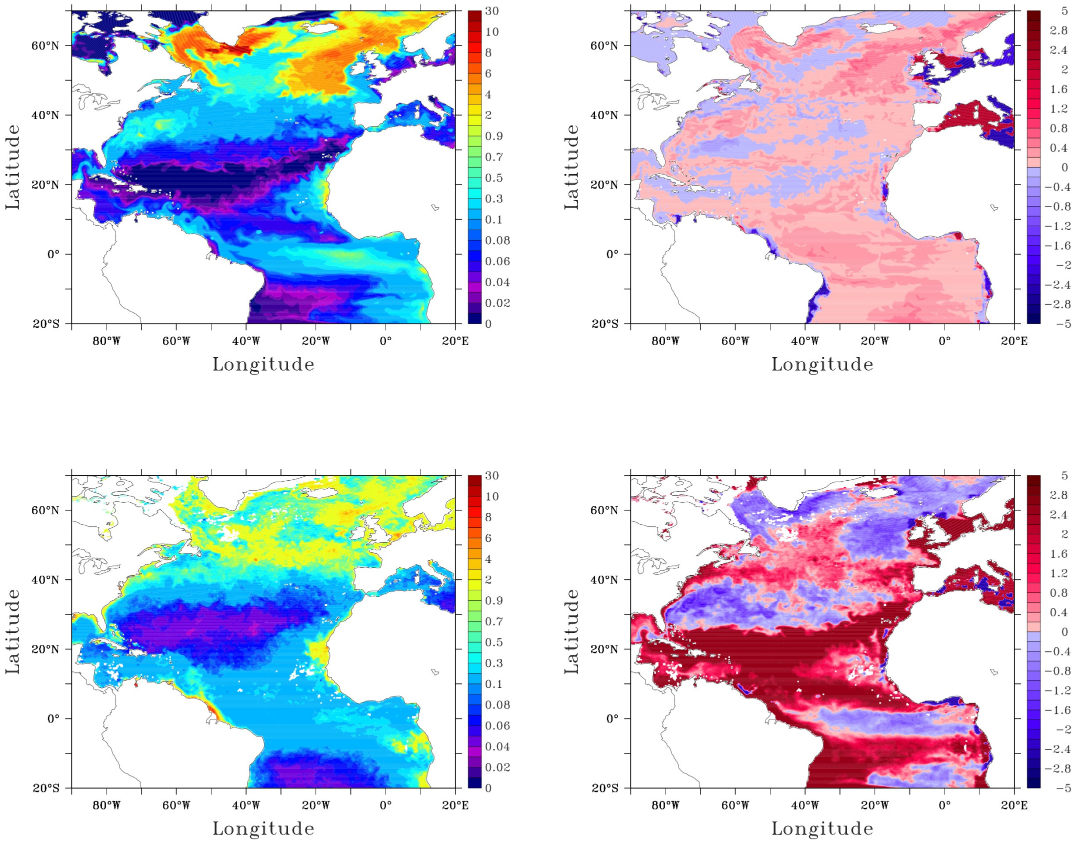

Figure 1: First line: maps of chl a concentrations simulated by the model for the day 16^{th} May 1998 with the physical variables (left) and with the anamorphosed variables (right). Second line: maps of chl a concentrations observed with ocean color during May 1998 with the physical variables (left) and with the anamorphosed variables (right).

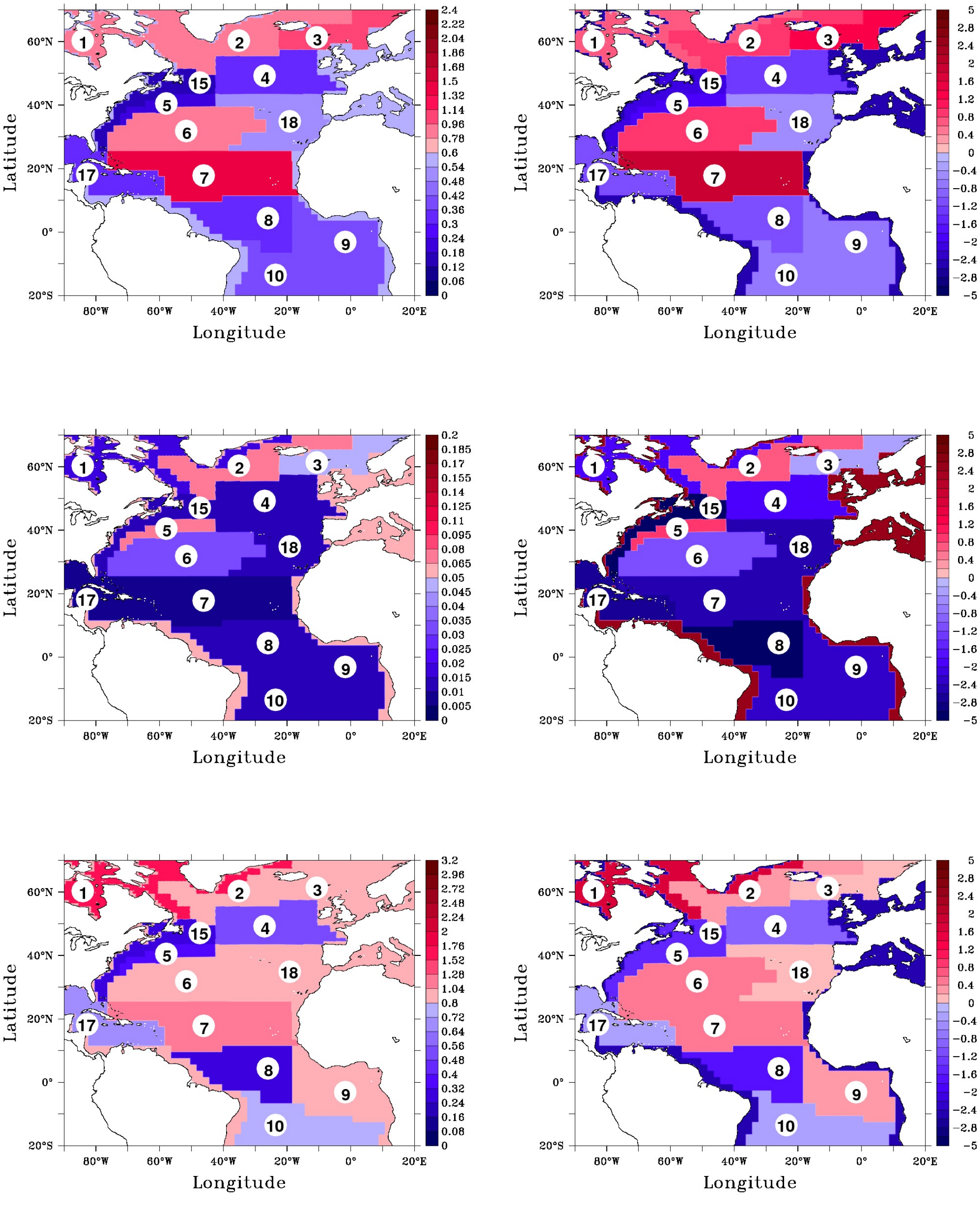

Figure 2: Results provided by the analysis step for the parameter values with the ocean color observations of May 1998. First line: maps of the regional growth rate μ with the physical variables (left) and with the anamorphosed variables (right). Second line: maps of the regional mortality rate m with the physical variables (left) and with the anamorphosed variables (right). Third line: maps of the regional grazing rate g with the physical variables (left) and with the anamorphosed variables (right). Units are day^{-1} for the physical variables and dimensionless for the anamorphosed variables.

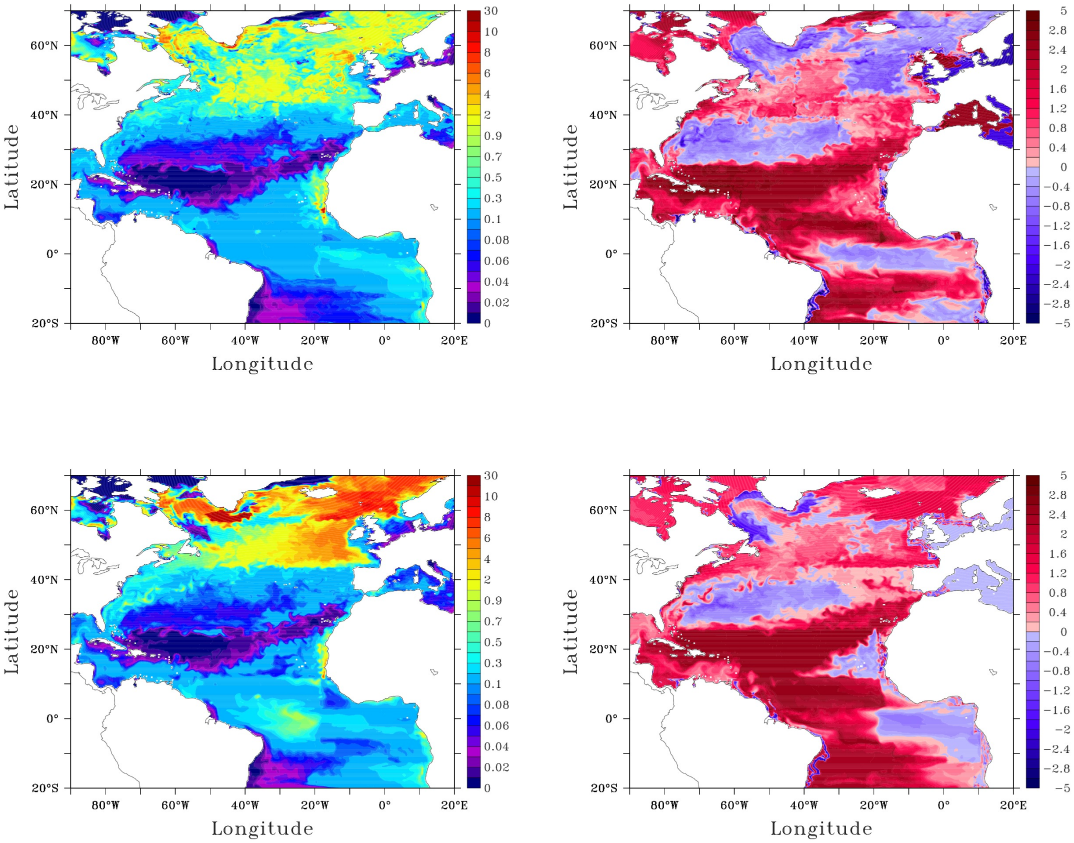

Figure 3: First line: maps of chl *a* concentrations after analysis (day 16^{th} May 1998) with the physical variables (left) and with the anamorphosed variables (right). Second line: maps of chl *a* concentrations simulated with the model including corrected regional parameter values with the physical variables (left) and with the anamorphosed variables (right).

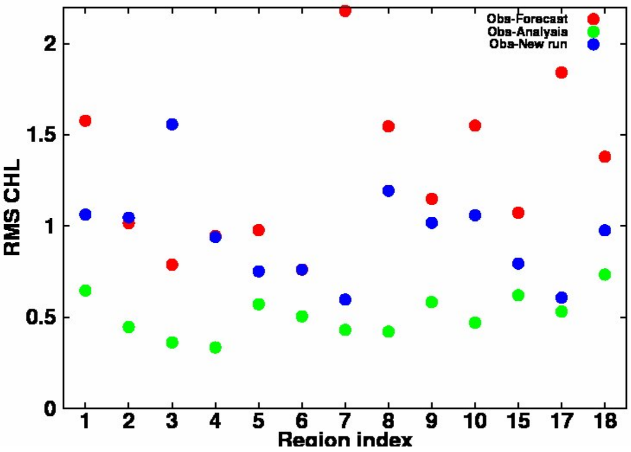

Figure 4: RMS of the [chl *a*], which has been calculated with the anamorphosed variables. The observation is the monthly map for May 1998. The forecast is with the uniform parameters, the analysis is the update with the observed [chl *a*] for May 1998 and the new run is the run at the date 16^{th} of May 1998 when the analysed parameters have been imposed at the date 16^{th} of April 1998.

4. Summary and conclusions

In the present study, regional values of key parameters of a 3D biogeochemical ocean model are estimated through the assimilation of satellite measurements of ocean colour distributed by the Globcolour project. The method implemented in this study was previously developed in the framework of twin experiments (Doron et al. (2011)). Using the augmented state vector formalism, the stochastic estimation method relies on multivariate correlations between biogeochemical parameters and model variables, such as the surface phytoplankton concentration considered here as a proxy of ocean colour. To prescribe these correlations, a Monte Carlo experiment involving 200 simulations is performed in which all simulations differ only by the specification of the set of key parameters. A nonlinear transformation (named anamorphosis) derived from the ensemble of simulations, is implemented to better describe the multivariate correlations and ensure that the analysis is optimally performed with Guassian transformed variables.

The reference model simulation and the ensemble runs have been performed for the month of May 1998, and the first analysis is done for the observations of May 1998. The results are explored in terms of both anamorphosed and physical values, demonstrating the interest of using the anamorphic transformations. The set of optimally estimated parameters exhibits values both larger and smaller than the initial value, showing an effective impact of the assimilated data to refine the parameterization of the model regionally. A new run using the optimal set of regional parameters enables to reduce the RMS between the model output [chl *a*] and the observations (both with their anamorphosed values), compared to the initial simulation with uniform parameters. A second objective of this study was to investigate the sensitivity of the regionalisation of the parameters to the assimilated data. To do so, we performed a series of analysis (52 for different months and years), with observations captured for the months of March to June, between 1998 and 2010. The analysed parameters showed a strong consistency among the different years with little interannual variability , while the month-to-month variability was more clearly evidenced. This suggests that the temporal variability of the parameters might be

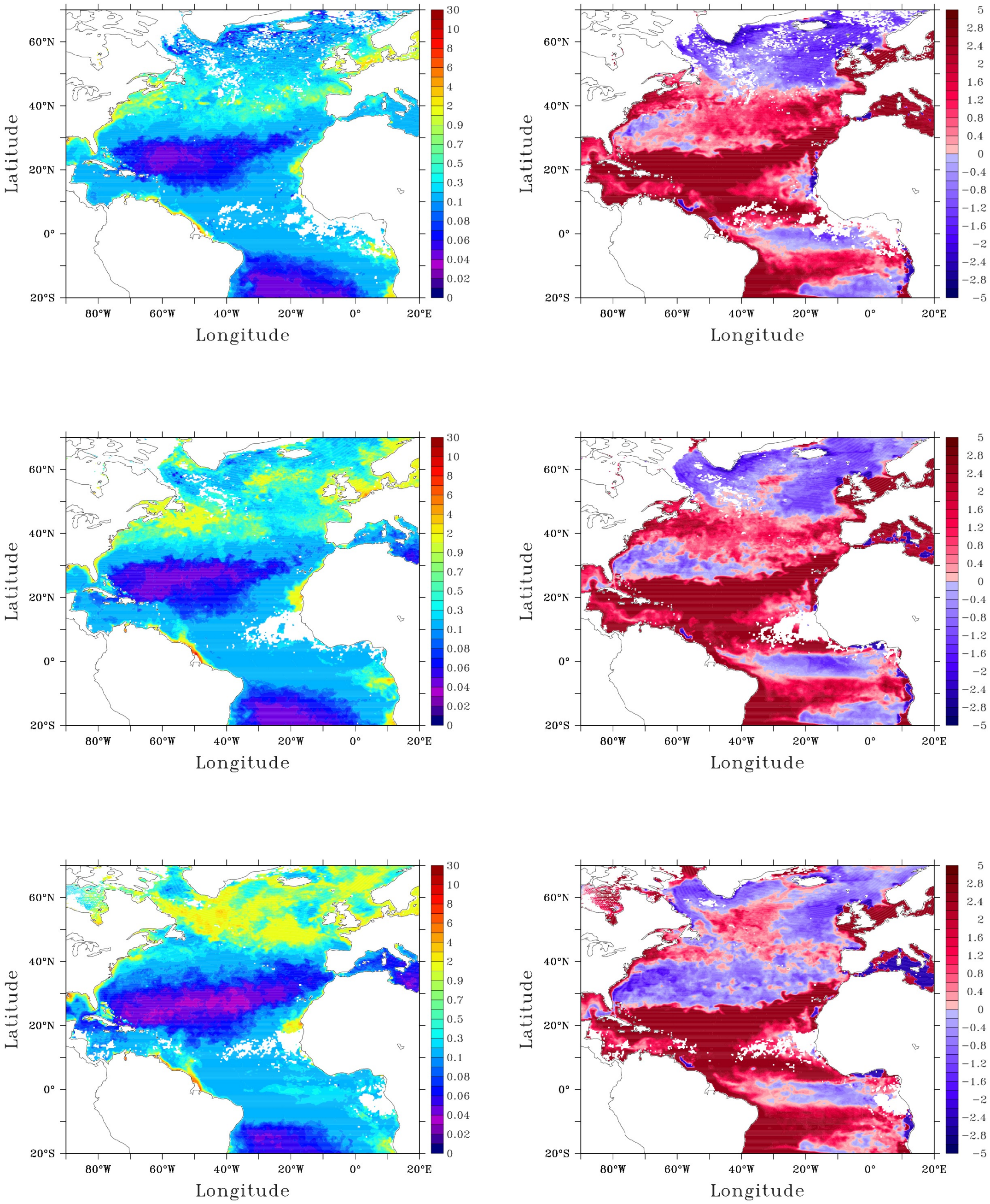

Figure 5: Maps of surface chl *a* concentrations observed with ocean color - First line: monthly average of March 1998, in $mg\ chl\ a\ m^{-3}$ (left) and the corresponding anamorphosed values (right). Second line: monthly average of April 1998, in $mg\ chl\ a\ m^{-3}$ (left) and the corresponding anamorphosed values (right). Third line: monthly average of June 1998, in $mg\ chl\ a\ m^{-3}$ (left) and the corresponding anamorphosed values (right).

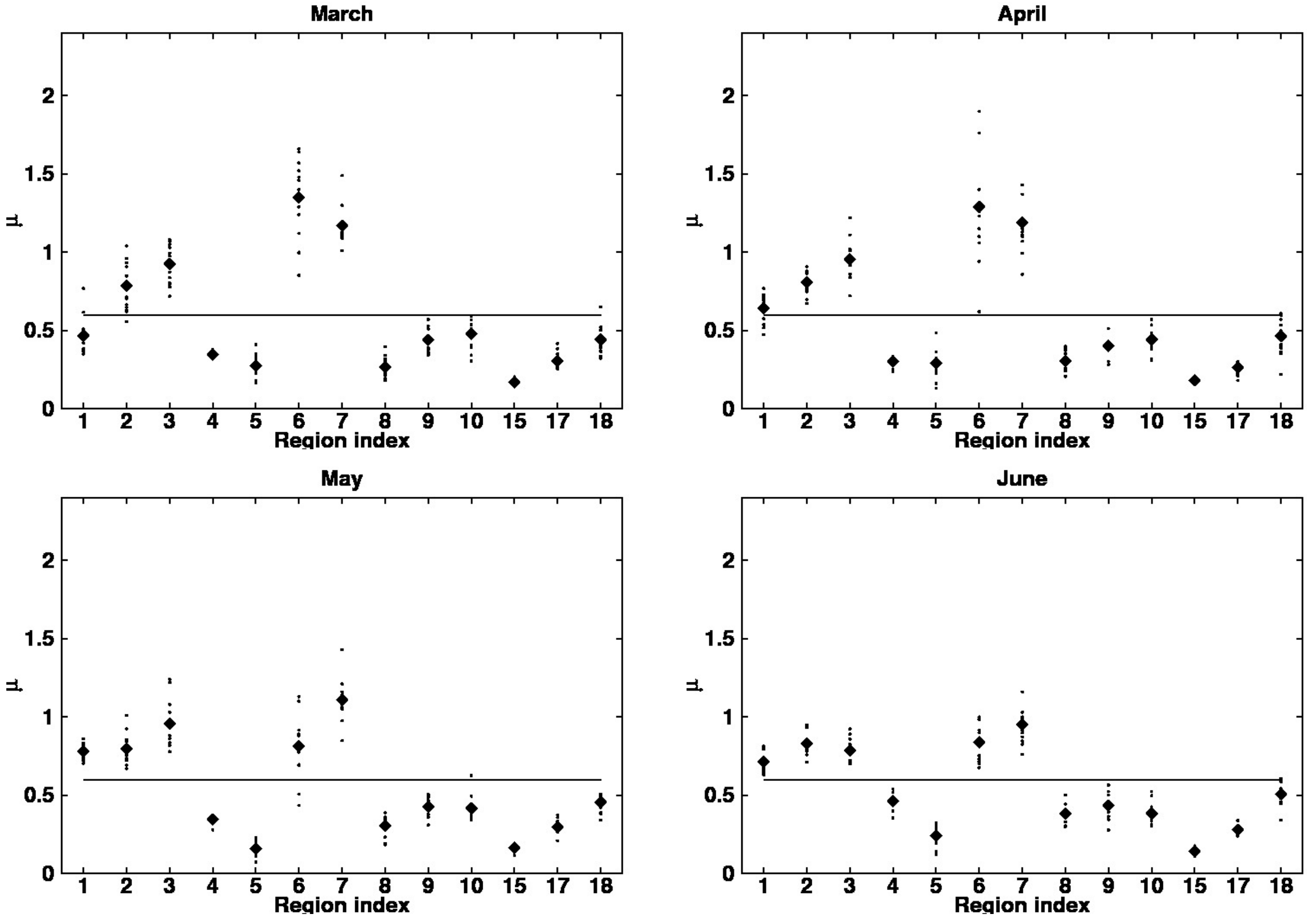

Figure 6: Average regional values of μ obtained with the analysis of different Globcolour observations for the month of March, April, May and June (from 1998 to 2010). Small symbols are for individual years, whereas the large symbols stand for the average of the 13 years. The uniform value (as used in the reference simulation) is identified by the horizontal line.

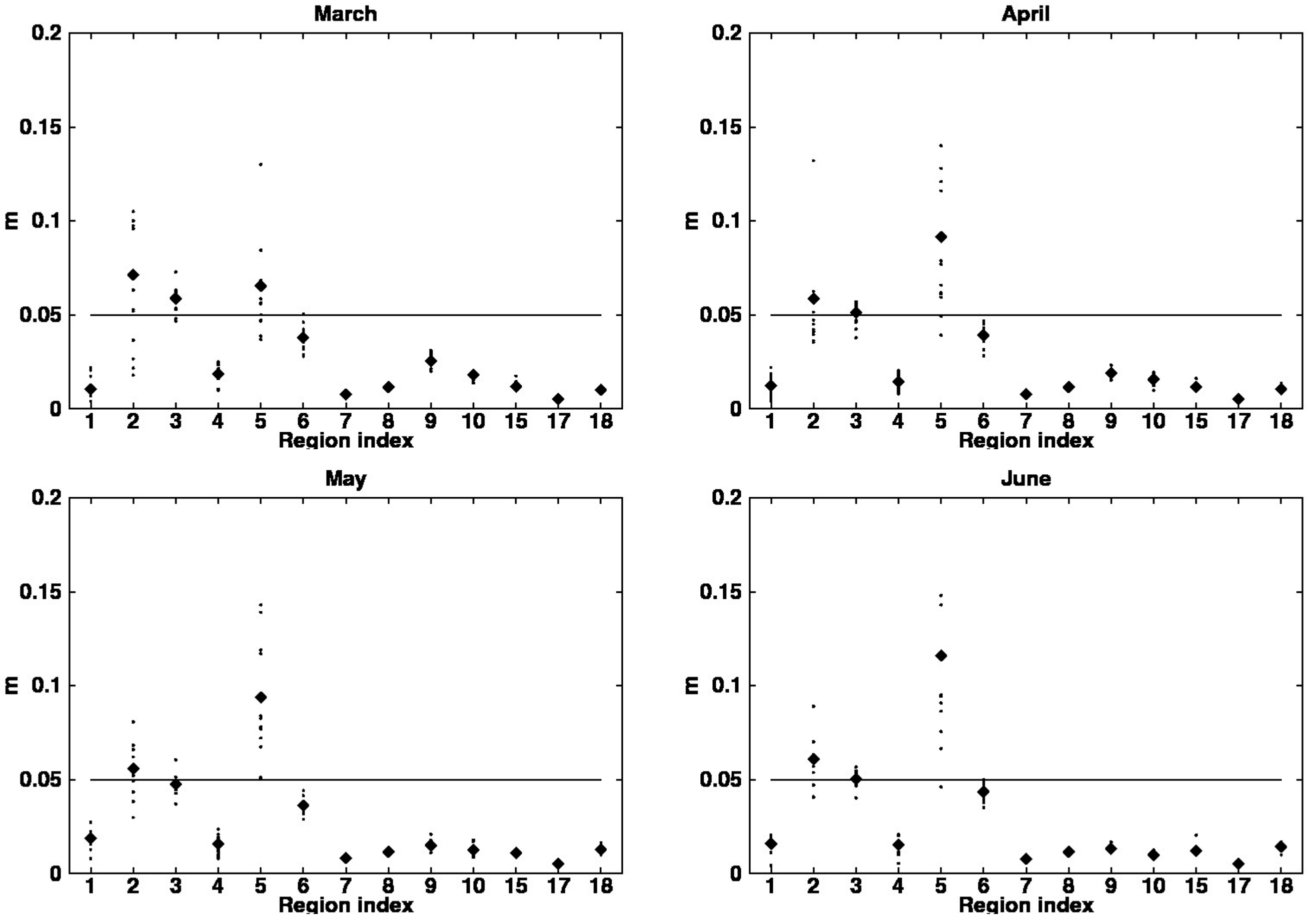

Figure 7: Average values of m obtained with the analysis of different Globcolour observations for the month of March, April, May and June (from 1998 to 2010). Small symbols are for individual years, whereas the large symbols stand for the average of the 13 years. The uniform value (as used in the reference simulation) is identified by the horizontal line.

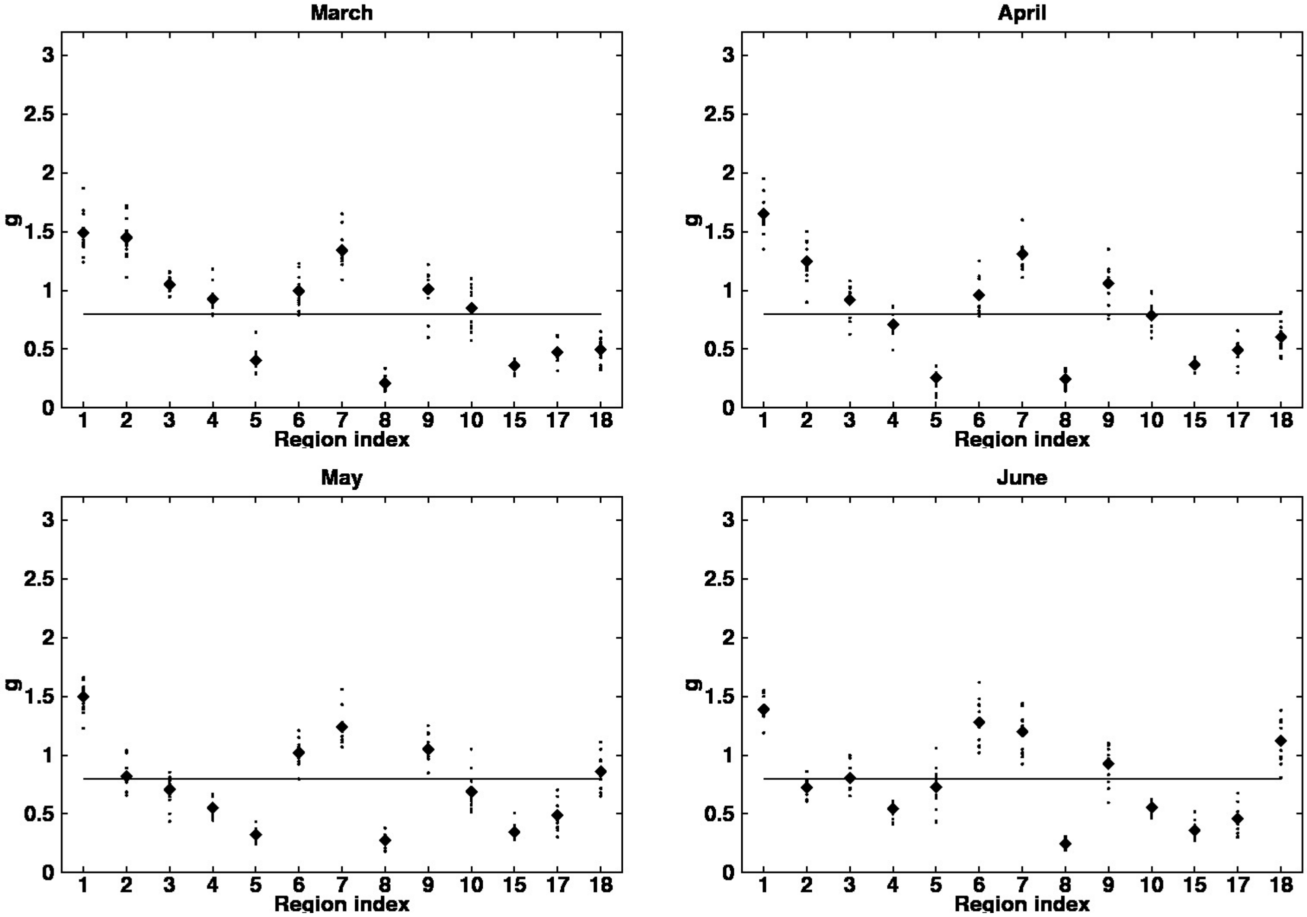

Figure 8: Average values of g obtained with the analysis of different Globcolour observations for the month of March, April, May and June (from 1998 to 2010). Small symbols are for individual years, whereas the large symbols stand for the average of the 13 years. The uniform value (as used in the reference simulation) is identified by the horizontal line.

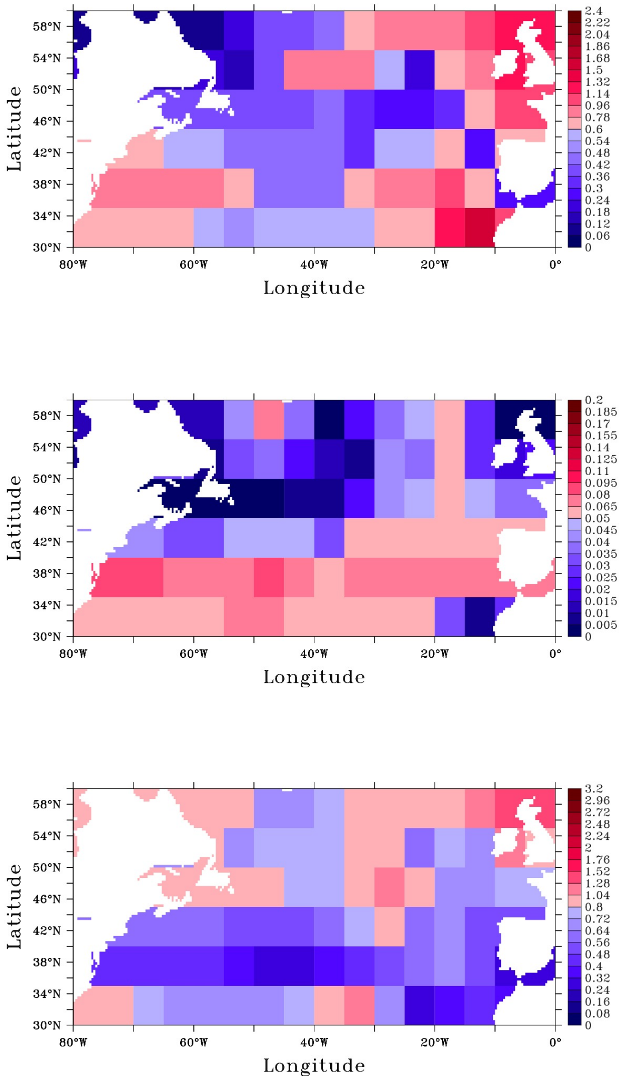

Figure 9: Data obtained by Losa et al. (2004): from top to bottom, the phytoplankton maximum growth rate (with the reference value 0.6), the phytoplankton maximum specific mortality rate (with the reference value 0.05) and the maximum specific grazing rate for zooplankton (with the reference value 0.73). These three parameters are optimised in their study with a weak constraint data assimilation method, using ocean color data. In their study, six parameters are optimised, but these three biological parameters corresponds quite directly with the parameters we consider in our study (although the underlying set of equations is different). See text in Section 3.3 for more details.

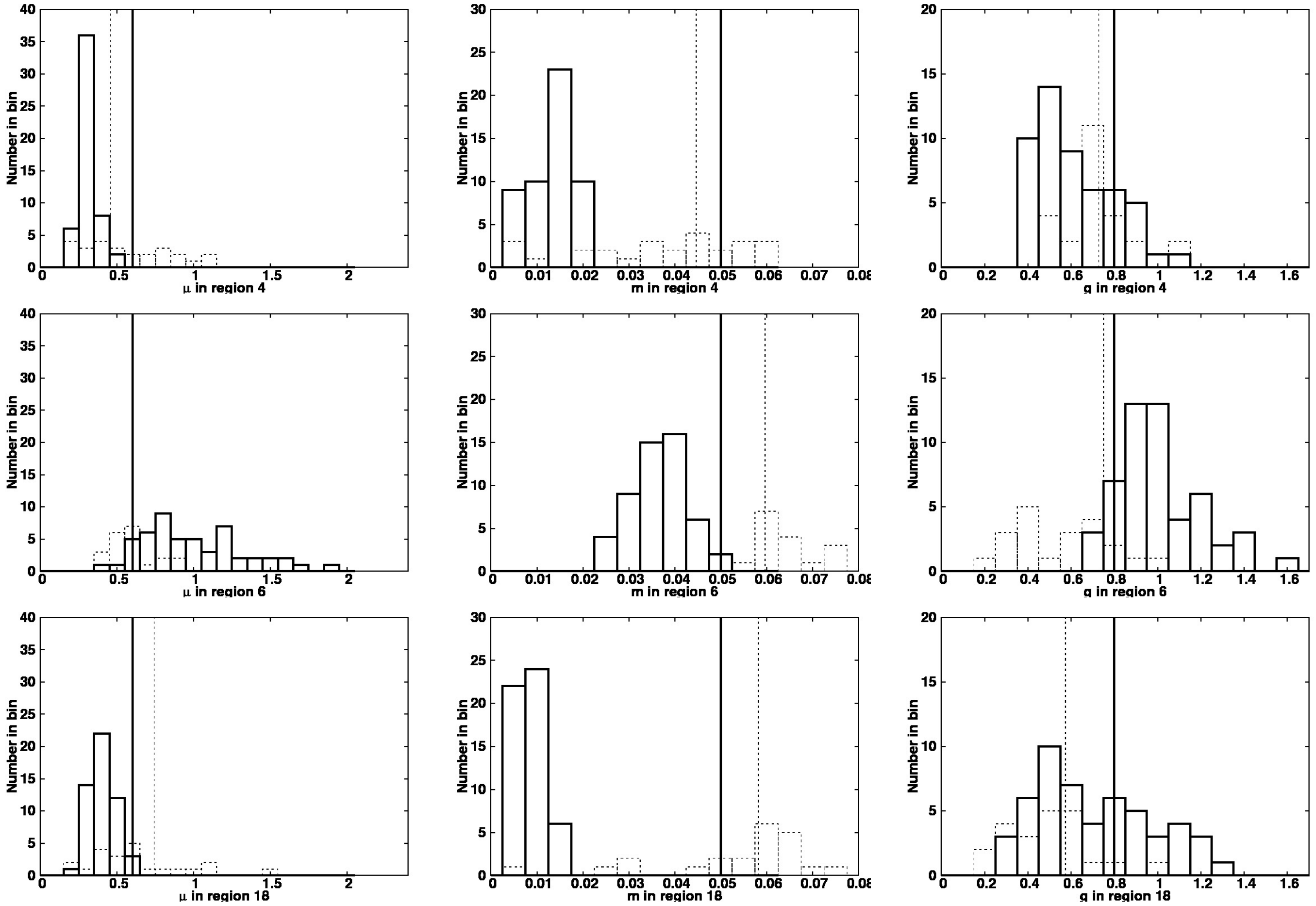

Figure 10: Histograms of analysed parameters for the three different Longhurst regions included in the Losa et al. (2004) study. The thick black histograms are based on the 52 analysed values (13 years times 4 months of observations) from this study. Our reference values are shown as a vertical thick black line (respectively 0.6, 0.05 and 0.8 respectively for μ, m and g), they are very close from Losa et al. (2004) reference values (respectively 0.6, 0.05 and 0.73). The average value of Losa et al. (2004) parameters is calculated on the three regions (4, 6 and 18) for μ, m and g. It is shown as a vertical dotted black line. Note that for region 6 and for μ, the average is 0.605 day^{-1}, hence, the dotted line is not visible. The dotted histograms are based on the spatial variable values of Losa et al. (2004) study in each region.

a future direction of research by exploring the variability during a full annual cycle.

Further, we compared our regionalised parameters to the regionalised parameters obtained by Losa et al. (2004) in a totally independent study. Both studies were conducted using data assimilation to estimate a few key regional biogeochemical parameters in a coupled model of the North Atlantic. The two studies differ by the models used (3D versus 0D), the resolution of the regions (Longhurst type versus 5° by 5° boxes), the estimation methods (stochastic approach with a nonlinear transformation and an augmented state vector versus weak constraint variational method), the temporal resolution (monthly versus annual), and the observations (SeaWiFS, MERIS and MODIS versus CZCS). Histograms of both types of parameters are compared on three Longhurst regions. They show a reasonably good agreement for two parameters: the phytoplankton growth rate and the zooplankton grazing rate. For the phytoplankton mortality rate, there is less consistency between the results from the two methods. This might be explained partly by the fact that, in the LOBSTER model, the phytoplankton mortality parameterisation is linear while, in Losa et al. (2004) model, the process is presented in a saturated quadratic form. More generally, it appears that the mortality rate is a delicate parameter to estimate, probably because it contributes to the closure of the biological loop and therefore incorporates many different hidden processes.

Although our results show that the application of the estimation method is feasible with real ocean colour data, possible extensions and improvements of the method are numerous. Strictly dealing with the parameter estimation method, two possibilities are to extent the time frame of application and the definition of the regions. We focused indeed in the present study on the spring bloom. One could further investigate whether the model has the same sensitivity to the parameter uncertainty for other seasons. If the small scatter for different years is confirmed, the annual cycle of parameters could be implemented in multi-year simulations. In addition, we based our regions on the work realised by Longhurst (2007), in which the boundaries between different regions are not only rectilinear but also static. However, the spatial variability of the parameters within some provinces (e.g. 4, 6, 18) supports the idea of refined definitions of the regions. Oceanic structures and circulation would probably suggest more flexible boundaries between regions. Some studies, for instance Devred et al. (2007), proposed a dynamic delineation of ecological provinces using ocean colour radiometry in the North-Western Atlantic. This dynamic delineation could be integrated in the approach of estimating biogeochemical parameters for ecological provinces, with refined definitions of the regions. Another future research direction is the combined parameter and state estimation. Recent studies, such as Mattern et al. (2012) or Roy et al. (2012), showed that implementing the possibility to have temporal variations of parameters allowed a better agreement between model outputs and observations. Such implementation should improve the phytoplankton representation through a better model.

The methodology that was assessed here is not dependent on the biogeochemical model and does not require huge inverse model developments. It could therefore be applied to any type of models as long as the ensemble simulations required to describe the model response to parameter uncertainty are computationally feasible. It paves the way to an actual way of regionalised biogeochemical models at global scale. Indeed, biogeochemical models are calibrated with in situ data, but their results could be improved if they were locally calibrated (Friedrichs et al., 2007). This is an important alternative to be considered in the current context of research where numerous model do coexist, while the computational capacity to increase the number of variables and objectively calibrate the associated parameters will be limited.

5. Acknowledgements

We are thankful to two anonymous reviewers for many helpful comments on the manuscript. This work was supported by the ESA-funded ASSOCO project (ESA/ESRIN Contract Network 22408/09/I-EC) in the framework of the Changing Earth Science Network. Partial support of the European Commission under Grant Agreement FP7-SPACE-2011-1-CT-283367-MYOCEAN2 is gratefully acknowledged. Calculations were performed using HPC resources from GENCI-IDRIS (Grant 2010-011279).

References

Armstrong, R., Sarmiento, J.L.and Slater, R., 1995. Ecological Time Series. Chapman and Hall, London, Ch. Monitoring ocean productivity by assimilating satellite chlorophyll into ecosystem models, pp. 371– 390.

Aumont, O., Maier-Reimer, E., Blain, S., Monfray, P., 2003. An ecosystem model of the global ocean including Fe, Si, P colimitations. Global Biogeochemical Cycles 17 (2).

Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Boening, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., De Cuevas, B., 2006. Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. Ocean Dynamics 56 (5-6), 543–567.

Béal, D., Brasseur, P., Brankart, J. M., Ourmières, Y., Verron, J., 2010. Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis. Ocean Science 6 (1), 247–262.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. International Statistical Review 71 (2), 223–241.

Brankart, J.-M., Testut, C.-E., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., Verron, J., 2012. Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations. Ocean Science 8 (2), 121–142.

Brasseur, P., Gruber, N., Barciela, R., Brander, K., Doron, M., El Moussaoui, A., Hobday, A. J., Huret, M., Kremeur, A.-S., Lehodey, P., Matear, R., Moulin, C., Murtugudde, R., Senina, I., Svendsen, E., SEP 2009. Integrating Biogeochemistry and Ecology Into Ocean Data Assimilation Systems. Oceanographycl 22 (3, Sp. Iss. SI), 206–215.

Claustre, H., Antoine, D., Boehme, L., Boss, E., D'Ortenzio, F., Fanton D'Andon, O., Guinet, C., Gruber, N., Handegard, N. O., Hood, M., Johnson, K., Lampitt, R., LeTraon, P.-Y., Lequéré, C., Lewis, M., Perry, M.-J., Platt, T., Roemmich, D., Testor, P., Sathyendranath, S., Send, U., Yoder, J., 2010. Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles. In: Hall, J., Harrison, D. E., Stammer, D. (Eds.), Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1). ESA Publication WPP-306, Venice, Italy, 21-25 September 2009, doi:10.5270/OceanObs09.pp.14.

Devred, E., Sathyendranath, S., Platt, T., 2007. Delineation of ecological provinces using ocean colour radiometry. Marine Ecology Progress Series 346, 1–13.

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.

Fasham, M., Ducklow, H., McKelvie, S., 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. Journal of Marine Research 48 (3), 591–639.

Fontana, C., Brasseur, P., Brankart, J.-M., 2012. Toward a multivariate reanalysis of the north atlantic ocean biogeochemistry during 1998 - 2006 based on the assimilation of seawifs chlorophyll data. Ocean Science Discussion 9, 1887–1931.

Fontana, C., Grenz, C., Pinazo, C., 2010. Sequential assimilation of a year-long time-series of seawifs chlorophyll data on the french mediterranean coast. Continental Shelf Research 30, 1761–1771.

Fontana, C., Grenz, C., Pinazo, C., Marsaleix, P., Diaz, F., 2009. Assimilation of seawifs chlorophyll data into a 3d coupled physicalbiogeochemicalmodel applied to a freshwater influenced coastal zone. Continental Shelf Research 29, 1397–1409.

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., J., D., 2012. Assimilating globcolour ocean colour data into a pre-operational physical-biogeochemical model. Ocean Sciences Discussions 9, 687–744.

Friedrichs, M., 2002. Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean. Deep-Sea Research Part II - Topical Studies in Oceanography 49 (1-3), 289–319.

Friedrichs, M. A. M., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy, Jr., D. J., Moore, J. K., Schartau, M., Spitz, Y. H., Wiggert, J. D., 2007. Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. Journal of Geophysical Research-Oceans 112 (C8).

Garcia-Gorriz, E., Hoepffner, N., Ouberdous, M., 2003. Assimilation of seawifs data in a coupled physical-biological model of the adriatic sea. Journal of Marine Systems 40-41, 233 – 252.

Garver, S., Siegel, D., 1997. Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation .1. Time series from the Sargasso Sea. Journal of Geophysical Research - Oceans 102 (C8), 18607–18625.

Gregg, W. W., 2008. Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. Journal of Marine Systems 69 (3-4), 205–225, Fall Meeting of the American-Geophysical-Union, San Francisco, CA, DEC 13-17, 2004.

Hemmings, J., Srokosz, M., Challenor, P., Fasham, M., 2003. Assimilating satellite ocean-colour observations into oceanic ecosystem models. Philosophical Transactions of the Royal Society of London. Series A - Mathematical Physical and Engineering Sciences 361 (1802), 33–39.

Hemmings, J., Srokosz, M., Challenor, P., Fasham, M., 2004. Split-domain calibration of an ecosystem model using satellite ocean colour data. Journal of Marine Systems 50 (3-4), 141–179.

Hu, J., Fennel, K., Mattern, J. P., Wilkin, J., 2012. Data assimilation with a local ensemble kalman filter applied to a three-dimensional biological model of the middle atlantic bight. Journal of Marine Systems 94, 145–156.

IOCCG, 2006. Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. IOCCG Report 5.

IPCC, 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ishizaka, J., 1990. Coupling of coastal zone color scanner data to a physicalbiological model of the southeastern united-states continental-shelf ecosystem. 3. nutrient and phytoplankton fluxes and czcs data assimilation. Journal of Geophysical Research 95, 20201– 20212.

Korres, G., Triantafyllou, G., Petihakis, G., Raitsos, D., Hoteit, I., Pollani, A., Colella, S., Tsiaras, K., 2012. A data assimilation tool for the pagasitikos gulf ecosystem dynamics: Methods and benefits. Journal of Marine Systems 94, S102–S117.

Kriest, I., Khatiwala, S., Oschlies, A., 2010. Towards an assessment of simple global marine biogeochemical models of different complexity. Progress In Oceanography 86 (3-4), 337–360.

Lévy, M., Gavart, M., Memery, L., Caniaux, G., Paci, A., 2005. A four-dimensional mesoscale map of the spring bloom in the northeast Atlantic (POMME experiment): Results of a prognostic model. Journal of Geophysical Research - Oceans 110 (C7).

Longhurst, A., 1995. Seasonal cycles of pelagic production and consumption. Progress in Oceanography 36 (2), 77–167.

Longhurst, A., 2007. Ecological geography of the sea. Academic Press, 2nd edition, San Diego, USA, 542p.

Losa, S., Kivman, G., Ryabchenko, V., 2004. Weak constraint parameter estimation for a simple ocean ecosystem model: what can we learn about the model and data? Journal of Marine Systems 45 (1-2), 1–20.

Madec, G., Delecluse, P., Imbard, M., Lévy, C., 1998. Opa 8.1 ocean general circulation model reference manual. Notes du pôle de modélisation Institut Pierre-Simon Laplace (IPSL), 91 pp.

Maritorena, S., Siegel, D., 2005. Consistent merging of satellite ocean color data sets using a bio-optical model. Remote Sensing of Environment 94 (4), 429–440.

Maritorena, S., Siegel, D., Peterson, A., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Applied Optics 41 (15), 2705–2714.

Mattern, J. P., Fennel, K., Dowd, M., 2012. Estimating time-dependent parameters for a biological ocean model using an emulator approach. Journal of Marine Systems 96-97, 32–47.

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

Natvik, L., Evensen, G., 2003b. Assimilation of ocean colour data into a biochemical model of the North Atlantic - Part 2. Statistical analysis. Journal of Marine Systems 40, 155–169.

Nerger, L., Gregg, W. W., 2007. Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. Journal of Marine Systems 68 (1-2), 237–254.

Nerger, L., Gregg, W. W., 2008. Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. Journal of Marine Systems 73 (1-2), 87–102.

Ourmières, Y., Brasseur, P., Levy, M., Brankart, J.-M., Verron, J., 2009. On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. Journal of Marine Systems 75 (1-2), 100–115.

Roy, S., Broomhead, D., Platt, T., Sathyendranath, S., S., C., 2012. Sequential variations of phytoplankton growth and mortality in an npz model: A remote-sensing-based assessment. Journal of Marine Systems 92, 16–29.

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 (4), 495–510.

Simon, E., Bertino, L., 2012. Gaussian anamorphosis extension of the denkf for combined state parameter estimation: Application to a 1d ocean ecosystem model. Journal of Marine Systems 89, 1–18.