Ocean Science

Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data

C. Fontana[1], P. Brasseur[2], and J.-M. Brankart[2]

[1]Université Pierre Marie Curie, Laboratoire d'Océanographie de Villefranche – UMR 7093, Villefranche sur mer, France
[2]CNRS, Laboratoire de Glaciologie et Géophysique de l'Environnement – UMR 5183, Grenoble, France

Correspondence to: C. Fontana (clement.fontana@obs-vlfr.fr)

Abstract. Today, the routine assimilation of satellite data into operational models of ocean circulation is mature enough to enable the production of global reanalyses describing the ocean circulation variability during the past decades. The expansion of the "reanalysis" concept from ocean physics to biogeochemistry is a timely challenge that motivates the present study. The objective of this paper is to investigate the potential benefits of assimilating satellite-estimated chlorophyll data into a basin-scale three-dimensional coupled physical–biogeochemical model of the North Atlantic. The aim is on the one hand to improve forecasts of ocean biogeochemical properties and on the other hand to define a methodology for producing data-driven climatologies based on coupled physical–biogeochemical modeling. A simplified variant of the Kalman filter is used to assimilate ocean color data during a 9-year period. In this frame, two experiments are carried out, with and without anamorphic transformations of the state vector variables. Data assimilation efficiency is assessed with respect to the assimilated data set, nitrate of the World Ocean Atlas database and a derived climatology. Along the simulation period, the non-linear assimilation scheme clearly improves the surface analysis and forecast chlorophyll concentrations, especially in the North Atlantic bloom region. Nitrate concentration forecasts are also improved thanks to the assimilation of ocean color data while this improvement is limited to the upper layer of the water column, in agreement with recent related literature. This feature is explained by the weak correlation taken into account by the assimilation between surface phytoplankton and nitrate concentrations deeper than 50

meters. The assessment of the non-linear assimilation experiments indicates that the proposed methodology provides the skeleton of an assimilative system suitable for reanalyzing the ocean biogeochemistry based on ocean color data.

1 Introduction

Monitoring the evolution of the marine biogeochemistry with relevant accuracy and resolution is a key requirement to better understand the ocean response to accelerating global climate change and the consequent effects on the carbon cycle and living resources. Unfortunately, the signature of the oceanic biogeochemical functioning, such as the regional patterns, vertical extension and timing of primary production at basin-scale, is still poorly known as a result of sparse historical data (Garcia et al., 2010) and the incomplete deployment of dedicated observing systems (Claustre et al., 2010a, b).

While it is conceivable to characterize the biogeochemical properties of a limited zone in the coastal domain through field measurements only (oceanographic cruises, autonomous sensors, etc.), it seems unrealistic to obtain spatially and temporally synoptic descriptions of vast ocean basins using similar approaches in the foreseeable future. Spatial ocean color sensors are the main source of global biogeochemical data available today. These sensors enable the observation of optical properties of the upper ocean such as the water leaving radiance in the visible spectrum, which can be related to the sea surface chlorophyll concentration.

Today, more than one decade of global ocean color data have been collected (Wilson, 2010), starting with the "proof-of-concept" Coastal Zone Color Scanner mission and more recently with several missions such as MEdium Resolution Imaging Spectrometer (MERIS), Moderate Resolution Imaging Spectrometer (MODIS), and Sea-viewing Wide Field-of-View (SeaWiFS). In spite of the invaluable merit of ocean color data, these sensors do not measure (directly or indirectly) other biogeochemical components such as nutrients (e.g., nitrates, ammonium) or trophic species. In addition, measurements are limited to the ocean surface, while the only source for deep observations is through the deployment of in situ sensors (Johnson et al., 2009).

An alternative approach to obtain depictions of the biogeochemical oceanic state is to use large-scale coupled physical–biogeochemical models (CPBM). The concept behind these models is to advect and diffuse biogeochemical tracers consistently with the ocean circulation as simulated by a numerical model solving the Navier–Stokes equations. A variety of biological formulations (either empirical or mechanistic) are used to update biogeochemical concentrations in the coupled model. One of the key assets of CPBMs is their ability to provide information on the coupled system with a high temporal and spatial resolution in three dimensions. A recognized weakness of CPBMs, however, is the approximate modeling of processes governing exchanges between the biogeochemical compartments. These processes are mostly dependent on the level of complexity of the model formulation, while in reality these interactions are time- and space-dependent as a lot of local factors may interfere in it (Doney et al., 1999). This is an obvious source of errors in the parameterization of the biogeochemical model and resulting model simulations.

The present study aims to combine ocean color satellite measurements with a CPBM to improve the representation of the biogeochemistry and its variability, extracting the best features from the model and the observations while reducing their respective weaknesses. Since satellite data are thought to describe the near-surface biogeochemistry with some faith, it is assumed here that the CPBM has enough skill to extrapolate the surface observations onto non-observed biogeochemical properties (especially at depth), in agreement with the underlying ocean physics and the biogeochemical principles of the model. Essentially, this is achieved by assimilating satellite chlorophyll data into a CPBM to increase the realism of the biogeochemical state variables. This approach has many similarities with the philosophy of altimeter data assimilation into dynamical models, which aims at inverting the signature of the surface dynamic topography into estimates of its internal dynamics (e.g., Fukumori et al., 1995; Brasseur et al., 1999).

Today, the routine assimilation of satellite data (e.g., altimetry, sea surface temperature) into operational forecasting models of ocean physics is mature enough to provide relevant information on non-observed parameters such as salinity, temperature and velocity fields (Cummings et al., 2009).

This capacity has been demonstrated by the production of global reanalyses of ocean physics to reconstruct the variability of its circulation during the past decades (Stammer et al., 2010).

The expansion of the "reanalysis" concept from physics to biogeochemistry is a timely challenge that motivates the present study. The sequential assimilation of a biogeochemical data set into CPBMs has, however, not yet reached the same level of maturity as for the physics, in spite of a number of successful studies on the subject (Carmillet et al., 2001; Natvik et al., 2003; Ford et al., 2012). A comprehensive review of biological data assimilation experiments, both sequential and variational, can be found in Gregg et al. (2009).

Several specific difficulties appear when considering the assimilation problem into CPBMs. Firstly, the measurement of top-of-atmosphere water-leaving radiance usually exhibits large differences with above-sea-surface equivalent values, as a result of strong interactions between visible light and the atmosphere (Lavender et al., 2005); this issue partly explains why most of the pioneer studies dealing with ocean color data assimilation were first carried out using pseudo-data (extracted from a model) rather than real data (Carmillet et al., 2001; Natvik et al., 2001). Secondly, it is often difficult to use the ocean color information to control the effect of errors in the ocean physics that cascade onto the biogeochemistry (Béal et al., 2010). Thirdly, in general the response of three-dimensional biogeochemical models to external forcings and parameterizations is highly non-linear, making the traditional assimilation framework inappropriate to develop these applications (Bertino et al., 2003; Doron et al., 2011). In the context of multivariate state estimation, where not only the observed variables are impacted by the assimilation process, these non-linearities can lead to failure of the method where corrections applied to the non-observed variables are unrealistic (Nerger and Gregg, 2007; Gregg 2008). Finally, global ocean circulation models require important numerical resources and are generally designed to be run on the most powerful computers dedicated to oceanographic research or operational systems. The coupling of global and biogeochemical models requires the advection and diffusion of supplementary state variables which increase the numerical needs such that the CPBM becomes hard to handle from a practical point of view. This is especially true for ensemble-based methods where the model needs to be run $\sim$O(100) times to obtain statistically consistent ensembles of simulations.

Considering recent advances made in the field of non-linear data assimilation (e.g., Bertino et al., 2003; Simon and Bertino, 2009; Bocquet et al., 2010; Brankart et al., 2012), and the need to develop the next generation of operational ocean monitoring systems within the framework of the MyOcean project (http://www.myocean.eu.org/), the aims of the present study are (i) to implement a multivariate, ocean color assimilative system based on state-of-the-art methods and to assess its performance in a pre-operational configuration,

(ii) to produce a multi-year reanalysis of the North Atlantic biogeochemistry using SeaWiFS satellite chlorophyll data from the period 1998–2006, which could eventually supersede the biogeochemical climatologies available today, and (iii) to investigate more specifically how information on superficial chlorophyll concentration can be projected onto non-observed variables. Here, non-observed variables mean both chlorophyll where no satellite data are available and other unmeasured biogeochemical components (e.g., nutrients).

The strategy adopted for this work relies on assimilation of the SeaWiFS data set because this mission consists of the longest time series of ocean color data to date (September 1997 to December 2010). The model domain is the North Atlantic, which exhibits highly contrasted seasonal and spatial biogeochemical behavior as well as many ocean circulation features found in other ocean basins. In order to make the assimilation tractable from a computational point of view, a simplified version of the Singular Evolutive Extended Kalman (SEEK) filter (Pham et al., 1998) has been chosen while anamorphic transformations as developed by Béal et al. (2010) are used to deal with the non-linear and non-Gaussian behavior of the CPBMs. Validation of the reanalysis is performed using independent data gathered from the World Ocean Atlas 2009 (WOA09) nitrate data set (Garcia et al., 2010).

The paper is organized as follows: Sect. 2 describes the model setup, observations and the assimilation method implemented in the assimilative system; Sect. 3 presents the experimental setup of the reanalysis system; and Sect. 4 discusses the results of the 1998–2006 reanalysis, showing the impact of the assimilation on a selection of observed and non-observed variables. Finally, an assessment of the results is presented in Sect. 5 before drawing conclusions.

2 Data, models and assimilation method

The coupled physical–biogeochemical model and assimilation framework considered in this paper is inherited from previous modeling studies of the North Atlantic biogeochemistry (Berline et al., 2007; Ourmières et al., 2009) and related assimilation developments (Béal et al., 2010; Doron et al., 2011). In the next section the main features of the modeling system developed for ocean color assimilation are briefly described.

2.1 The coupled physical–biogeochemical model and associated modeling errors

The physical component of the coupled model is simulated using the Nucleus for European Modelling of the Ocean (NEMO) code (Barnier et al., 2006) implemented in the North Atlantic basin at 1/4° horizontal resolution, which is considered as "eddy-permitting" in the mid-latitudes. NEMO
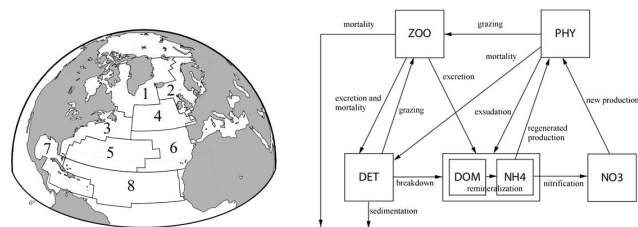

Fig. 1. Model domain and biogeochemical regions (left); LOBSTER biogeochemical model components and fluxes (right).

is a primitive equation model based on the free surface formulation. The prognostic variables are the three-dimensional velocity fields, temperature and salinity. The model domain covers the North Atlantic basin from 20° S to 80° N and from 98° W to 23° E (Fig. 1, left). Buffer zones are specified at the southern, northern and eastern (Mediterranean) boundaries (Treguier et al., 2001). Vertical discretization involves 45 geopotential levels, with grid spacing that increases from 6 m at the surface to 250 m at the bottom. The model is forced by ERA-INTERIM atmospheric fields (Dee et al., 2011) from the European Centre for Medium-Range Weather Forecasts (ECMWF), using bulk formulations as proposed by Large and Yeager (2004).

The biogeochemical component of the coupled model is the LOBSTER model (Levy et al., 2005) in the North Atlantic setup described by Ourmières et al. (2009). The LOBSTER formulation is nitrogen-based and contains six prognostic variables: nitrate, ammonium, phytoplankton, zooplankton, detritus and semilabile dissolved organic matter (Fig. 1, right). All biogeochemical variables are advected and diffused in three-dimensional space following the oceanic circulation computed by the physical model. The LOBSTER model considers closed boundaries at model grid frontiers. The chlorophyll concentration is a diagnostic variable, computed according to the phytoplankton concentration through a space- and time-dependent chlorophyll-to-nitrogen (Chl/N) ratio. The LOBSTER model updates biogeochemical concentrations with the same time step as the circulation model, i.e., every 40 minutes.

As in every simulation, the numerical modeling system has many imperfections so that the simulations exhibit a variety of errors compared to the "true" system evolution. This may be due to the atmospheric input data which are themselves derived from a model and therefore cascade into physical and biogeochemical modeling errors. Another possible source of errors lies in the parameterizations used to represent the effect of physical and biological processes that are not resolved explicitly by the model. This includes, among others, the effects of the sub-grid scale physical processes as well as the overly simplified structure of the ecosystem model, which is unable to represent the behavior of the actual ecosystem in nature. Furthermore, spatial and temporal discretizations used in numerical algorithms are another source of errors

that necessarily impact the model solutions. Finally, the initialization of physical and biogeochemical oceanic fields is generally done using approximate values (climatology or homogenous assumptions), and in spite of a necessary spin-up period the resulting simulation always keeps the signature of the chosen initialization method including its unrealistic features. The main goal of data assimilation is to reduce the adverse impacts of the modeling errors on the representation of the biogeochemistry by combining the model information with observations.

2.2 The ocean color data set and associated errors

The satellite chlorophyll concentrations that are assimilated in the coupled model are derived from SeaWiFS using the OC4 algorithm applied to the remotely-sensed water leaving radiance (Feldman and McCain, 2010). To simplify the assimilation process, satellite products are systematically remapped onto the model grid prior to their assimilation. The use of a composite data set (i.e., obtained by merging scenes from different sensors such as MODIS, MERIS and SeaWiFS) is not essential within the framework of the present study since the number of sensors available during the period of interest varies between 1 and 3 in one given location, and composite data would therefore make the diagnostics of the assimilation experiments rather complex.

Limited accuracy of ocean color products is another source of errors that must be taken into account in the assimilation process. The fraction of water leaving radiance measured at top-of-atmosphere by satellite sensors such as SeaWiFS is typically only 5–20 % of the total measured signal (Lavender et al., 2005). The remainder of the signal is due to interactions between light and the Earth atmosphere (e.g., Rayleigh reflectance, aerosol reflectance). Thus, in order to obtain reliable satellite-derived data in the visible spectrum, the application of atmospheric corrections is required to estimate the original above-sea-surface water leaving radiance. These corrections, however, are a major source of errors in retrieving the meaningful physical parameters from the water leaving radiance. More specifically, the OC4 algorithm used here is an empirical function conveying uncertainties in the chosen parameterizations. Other biogeochemical features (e.g., different assemblages of phytoplankton communities, presence of chromophoric dissolved organic matter, etc.) are also responsible for imperfections in the retrieval process of the chlorophyll concentration. All things considered, it will be admitted in our experiments that the error of the SeaWiFS chlorophyll products is on average 30 % of the measured signal amplitude in open oceans (Gregg and Casey, 2004).

2.3 The assimilation method

The assimilation method chosen to develop the reanalysis system is a sequential algorithm derived from optimal estimation theory: the model state trajectory is corrected intermittently by computing statistical updates of the state vector using a combination of available data and model predictions weighted according to their respective uncertainties.

The ocean color data are assimilated using a Singular Evolutive Extended Kalman (SEEK) filter (Pham et al., 1998) implemented in the coupled model using the System of Sequential Assimilation Modules (SESAM) assimilation platform (Brankart et al., 2012). This tool is used to perform all matrix operations required by the assimilation scheme, such as the computation of empirical orthogonal functions (EOF) of the reduced-order filter, the innovation vector and the analysis update. A reduced-rank Kalman filter with static error sub-space (Brasseur and Verron, 2006) is chosen here because its comparatively low computational burden enables making extended experiments vs. ensemble-based methods (e.g., ensemble Kalman filter) which require the explicit computation of ensemble evolution. Nevertheless, the upgrade of the assimilation scheme toward a fully explicit ensemble scheme will be straightforward in forthcoming applications.

The state vector entering the assimilation procedure is composed of all prognostic biogeochemical state variables of the three-dimensional model grid. This means that a multivariate analysis update is computed, where all observed and non-observed components of the biogeochemical model are modified. In addition, the SEEK filter is implemented using two different versions of the analysis step: in the first version, the analysis is performed using the original model state variables; in the second version, anamorphosis transformations are applied to each separate variable of the state vector prior to the analysis step, and the corresponding inverse transformation is applied after analysis to restart the model integration in the original model space. This aspect is identified as a key ingredient of the assimilation scheme that contributes to the efficiency of the procedure. When the analysis includes anamorphic transformations, the marginal probability density functions (PDFs) of the forecast variables are transformed into PDFs that are close to Gaussian. We will not enter into the mathematical details here, as this aspect is already fully documented elsewhere (e.g., Bertino et al., 2003; Béal et al., 2010; Brankart et al., 2012; Simon and Bertino, 2012). The parameterization of the anamorphic transformation is equivalent to the one used in Doron et al. (2011). The anamorphosis transformation presents several advantages that are expected to improve the assimilation efficiency. Firstly, as the final marginal PDFs are more Gaussian than the original ones, the assimilation process should be more respectful of the assumptions underlying the Kalman filter optimality. This ensures a better description of the correlations between observed and non-observed variables, as discussed in Brankart et al. (2012). Secondly, it is possible to parameterize the error statistics (more specifically the tails of the marginal probability distributions) in such a way that any "extrapolation" outside the range of values described by the ensemble is avoided. In essence, it means that it becomes

impossible to obtain negative values for the concentration variables of the state vector after the analysis update. In the linear case, negative concentrations obtained after an assimilation step were systematically set to a concentration value of $1\,e^{-6}\,\mathrm{mmol\,m^{-3}}$.

3 Experimental setup of the reanalysis system

3.1 Initialization of the coupled model and simulation strategy

The simulations described in this paper were performed on a 9-year period from 1 January 1998 to 1 January 2007. Physical variables were initialized following a 16-year spin-up period starting from the Levitus monthly climatology (Levitus et al., 1998) for temperature and salinity fields, while velocity fields were null. Biogeochemical variables were initialized following a 2-year spin-up period. This short spin-up was initialized using nitrate concentrations from the World Ocean Atlas climatology (Garcia et al., 2010), while homogeneous values were assigned to the other biogeochemical variables (see Ourmières et al., 2009).

The coupled model is subsequently run without data assimilation; this simulation will be referred hereafter as the "free" run. Two simulations using the data assimilation system are performed in parallel. These two reanalysis runs assimilate a set of temporally-binned satellite chlorophyll maps every 8 days. The assimilated maps are a binning of satellite data for the 8 days preceding the assimilation date, allowing to stay in an operational framework where future observations are not known.

In the first run, the analysis update is performed with no non-linear transformation of the state vector in the assimilation scheme. This experiment is referred hereafter as the "linear" run. For the second run, the anamorphosis transformation is applied to the state vector, the observation vector and the error covariance matrix. This second experiment will be referred hereafter as the "anamorphosis" or "non-linear" run.

Considering the assimilation increment procedure, the model is stopped every time an observation is available, then an analysis is computed. The model is then restarted from this analysis on the first time step and evolves freely until the next available observation.

Time-averaged data shown in this manuscript are obtained by computing the mean state of the considered data for each time step on a given period.

3.2 Specific setup of the assimilation system

The practical setup of the assimilation method is a critical step of the present reanalysis system. Numerous parameters enter the analysis computation, influencing the performance of the reanalysis experiments. However, as discussed previously, large-scale CPBMs require important computational

resources, preventing an exhaustive exploration of the sensitivity of the assimilation process on each parameter (including their mutual interplay). We describe below the strategy chosen to prescribe the key parameters of the reanalysis system.

The EOF basis entering the computation of the SEEK analysis is obtained from the free run variability. For each analysis date, a specific set of EOFs is computed using a temporal ensemble of state vectors sampled from the free model trajectory with a 2-day frequency. This "deterministic" ensemble is constructed as follows: for a given day of the year, all model states falling into the 2-month period surrounding the assimilation date are selected in the period 1999 to 2005 covered by the free simulation. Thus, every temporal ensemble contains 210 members that are used to compute the EOF basis and finally the error covariance matrix. The EOF basis is then truncated to the 20 dominant modes to perform the state vector update. The same ensemble of 210 members is also used in the non-linear run to build the anamorphosis transformation locally in space and time. More precisely, each time observations are assimilated, a specific non-linear transformation is computed for each model grid cell and for each model variable from the histogram of 210 values that are associated with this day of the year.

Concerning the parameterization of the tails of the anamorphic transformations (outside the range of the available ensemble), we make the simple assumption of a zero forecast probability in these regions of the state space. The direct consequence is that, even if an observation falls in these peripheric regions, our estimation of the observed variable cannot get close to the observation because it is bound to stay inside the range defined by the ensemble. However, in our application, this is not expected to occur very often, because the ensemble is built using the seasonal and interannual variability of the free model simulation (which is large during the bloom event), so that the dispersion of the ensemble could easily be tuned to be large enough to include most of the assimilated observations (except in some regions of the subtropical gyre). We thus preferred safety by avoiding any kind of extrapolation outside the range of values effectively explored by the model simulation. More sophisticated assumptions about the tails of the distribution (e.g., Gaussian tails) can be found in the works of Bertino et al. (2003) and Simon and Bertino (2009).

Regarding the observations, the SeaWiFS chlorophyll concentration maps are converted into phytoplankton concentration maps using the Chl/N ratio computed by the coupled model. These phytoplankton distributions are then assimilated in the coupled model and considered as representative of phytoplankton concentration in the upper first layer of the water column. The error associated with each distinct observation pixel is set to 30 % of the considered data, in agreement with the commonly used SeaWiFS error estimates for case 1 waters. In the non-linear run (with anamorphic transformations), uncertainties in the observations cannot be

specified exactly in the same way since they must be assumed Gaussian for the non-linearly transformed variables rather than for the original variables. Nevertheless, to give a similar importance to each observation in the two assimilation runs, we compute the observation error standard deviation for the transformed variable by multiplying the original observation error standard deviation (i.e., the 30 % of the observed chla concentration) by the local slope of the non-linear transformation (which we approximate in practice by a finite difference over one standard deviation).

Spatial coverage of the data is an issue that raises the question of spatial correlations of the signal in assimilation systems. As an example, some high-latitude regions may not be visible by satellite ocean color sensors during several months of the year (in winter of the corresponding hemispheres). For these conditions, performing a global analysis is an issue as it is evident that the mesoscale system state at mid-latitudes is not correlated to the system state at high latitudes. Considering this, it was decided to implement a local analysis scheme with a short influence radius for every distinct data available. The horizontal e-folding radius of influence is set to 2 grid points and the cut off radius to 5 grid points. This value was chosen as it is equivalent to meso-scale features for mid-latitude regions.

It is noteworthy that Hu et al. (2012) recently proposed equivalent parametrization of observation error and local influence radius within the framework of an ocean color data experiment.

3.3 Validation strategy

A data assimilation experiment involves at least three sets of information: the free simulation, the data to be assimilated, and the simulation with data assimilation. While intercomparisons between these three information sets is necessary to assess the method efficiency, it will never be totally conclusive since the three information sets are intertwined (Gregg et al., 2009). An independent data set of unassimilated variables is required for an objective determination. We use here the nitrate database extracted from the historical World Ocean Atlas 2009 (WOA09) as an independent data set to validate the assimilation process. The historical nutrient measurements available in this atlas were obtained from the National Oceanographic Data Center and World Data Center archive, including all data gathered as a result of the Global Oceanographic Data Archeology and Rescue (GODAR) and the World Ocean Database (WOD) project (Boyer et al., 2006). This large-scale data set is to our knowledge the one containing the largest number of in situ nitrate concentrations. The temporal and spatial coverage of this data set allows a systematic and objective comparison with the simulation outputs.

4   Results

In this section, we examine the effects of the assimilation on the variability of the biogeochemical properties in space and time, comparing the linear and anamorphosis runs with the free model simulation. The performances of the assimilative system are first evaluated in terms of ocean surface properties, before investigating how the assimilated ocean color data modify the distribution of nutrients in the sub-surface layers.

4.1   Surface patterns of chlorophyll and nitrate concentrations

Figure 2 shows the surface chlorophyll maps obtained after time-averaging the simulation results over successive 60-day periods during the year 2006 (first row: days 1 to 60; second row: days 61 to 120; etc.). Figure 3 is organized the same way but for the surface nitrate distributions. These maps are shown for (a) the free run, (b) the SeaWiFS data or climatology, (c) the linear run and (d) the anamorphosis run. The free run shows some significant differences with the SeaWiFS data (Fig. 2). The chlorophyll bloom starts slightly later in the free run than observed (second row, corresponding to March–April). An elongated structure centered at ∼35° N appears along the southern flank of the Gulf Stream, while it is not present in the data (third row, corresponding to May–June). Inversely, SeaWiFS data show for the same period an increase of chlorophyll concentration at latitudes greater than 45° N, corresponding to the beginning of the spring bloom, while concentration values are much lower in the free simulation. The available nutrients are rapidly consumed (Fig. 3, third and fourth rows, i.e., May–August), inducing a strong increase of the chlorophyll concentration. During the peak of the chlorophyll bloom in the free run, concentrations seem to be overestimated at high latitudes although the order of magnitude remains correct. When all nitrates are consumed, the chlorophyll concentration decreases quickly (Fig. 2), while the SeaWiFS data exhibit larger values that persist later until the summer season, and to a lower extent until the end of the year.

To summarize the comparison between the free simulation and satellite chlorophyll data, the modeled chlorophyll bloom starts too late and chlorophyll concentrations increase quickly to reach values overestimating data before decreasing rapidly. It is important to note here that, in spite of these differences, the main features of the annual biogeochemical cycle are well described (chlorophyll spring bloom, oligotrophic subtropical gyre, upwelling along the Mauritanian coast). This is a crucial point since the free run is actually sampled to compute the EOF basis and the local anamorphosis transformations that are used in the assimilation scheme.

Considering the runs with data assimilation, the bloom starts almost in phase with the observations, while the elongated pattern mentioned above in the Gulf Stream area is not
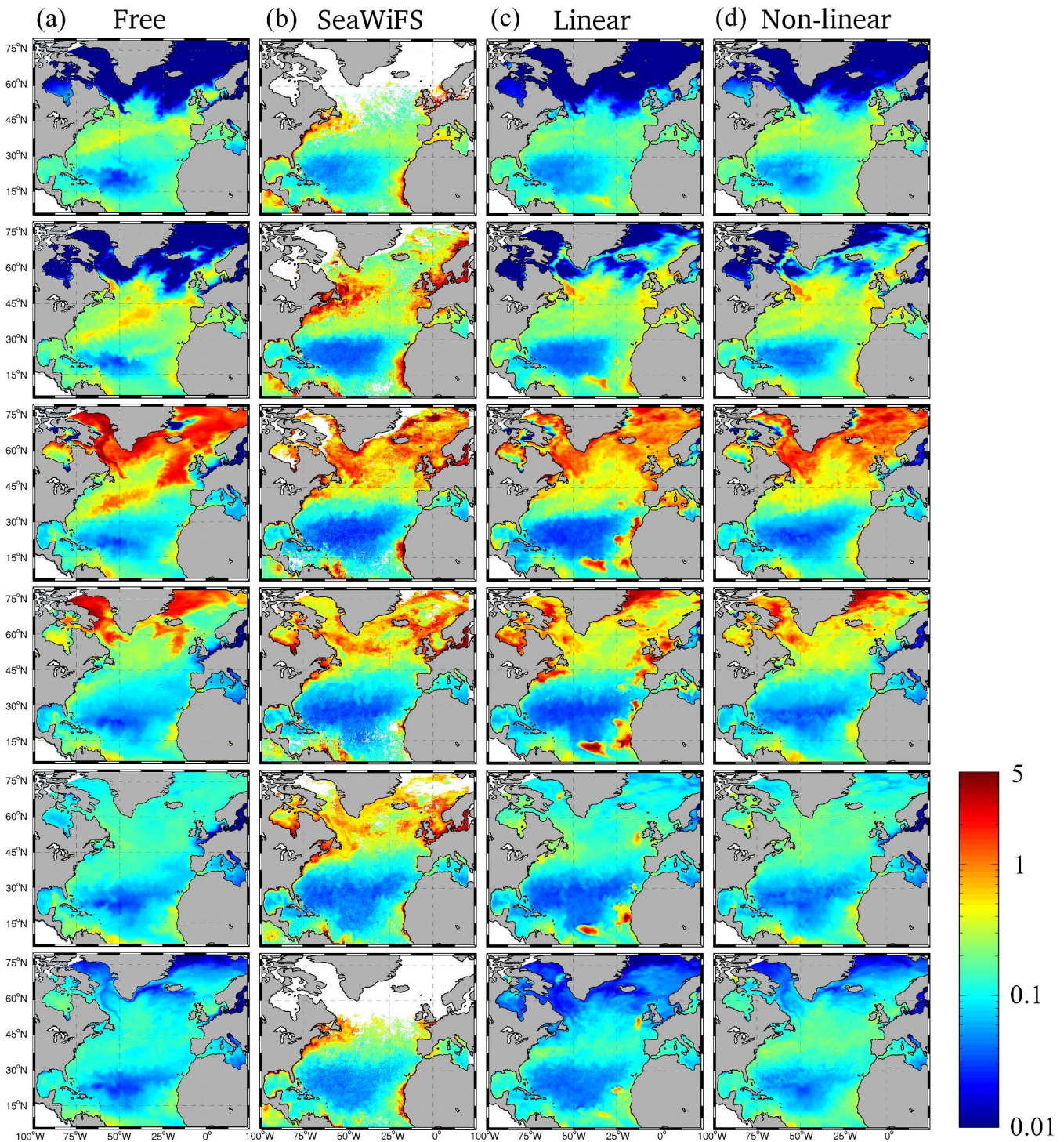
Fig. 2. Surface chlorophyll concentration binned on 60-day period (from top to bottom) for year 2006 for **(a)** free run; **(b)** SeaWiFS data; **(c)** linear run; **(d)** non-linear run. Concentrations are given in mg(Chl) m^{-3}.

present anymore (Fig. 2; columns c and d; third row). During the bloom, the model values are in the order of magnitude of the observations both in the linear and anamorphosis situations (third and fourth row), while strong spatial differences are visible (e.g., subpolar gyre). However, the model values are still underestimated on average by the end of the year, as shown above for the free run. To understand why the assimilation efficiency is weaker on chlorophyll concentrations by
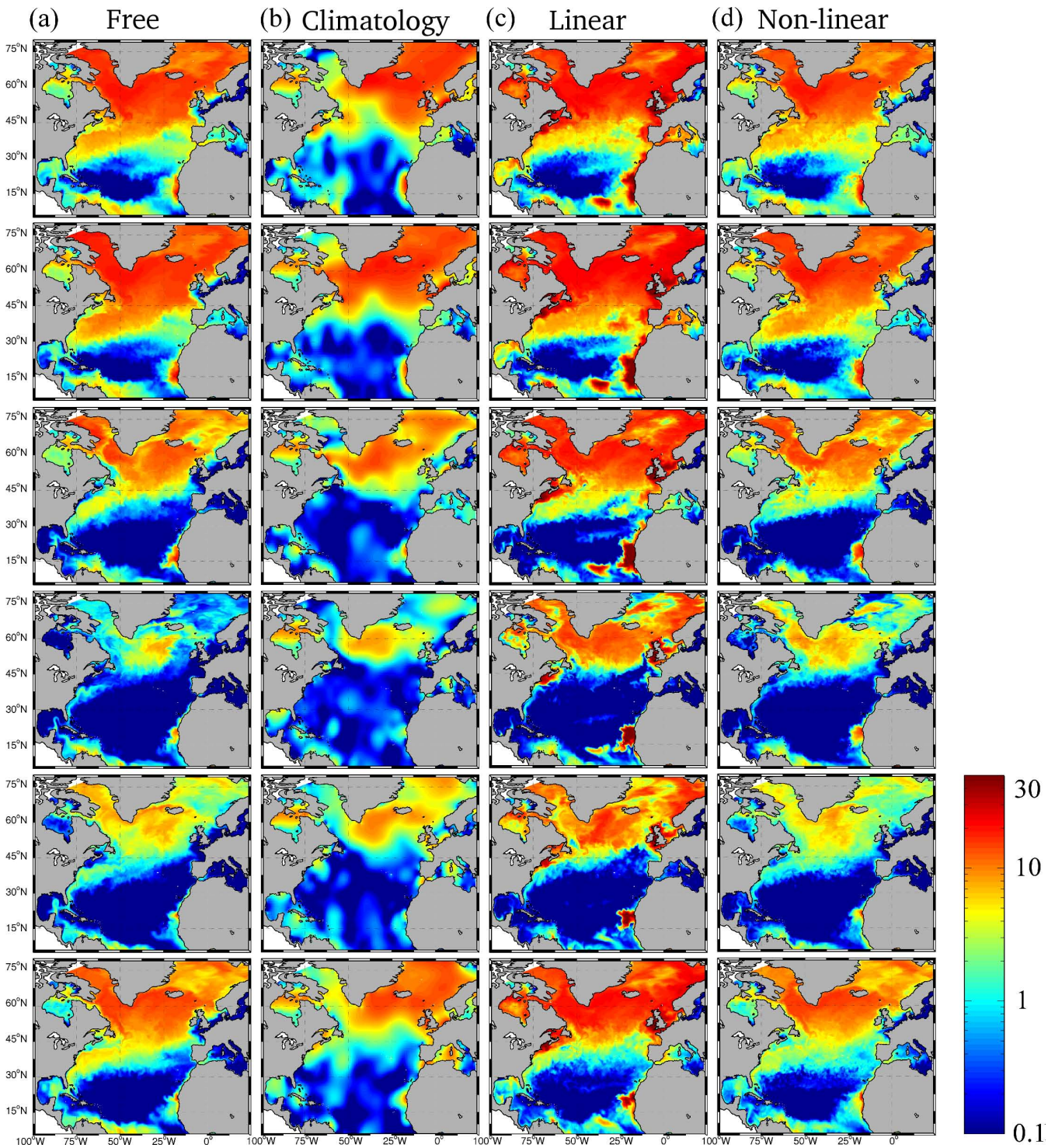
Fig. 3. Surface nitrate concentration binned on 60-day period (from top to bottom) for year 2006 for **(a)** free run; **(b)** WOA09 climatology; **(c)** linear run; **(d)** non-linear run. Concentrations are given in $\mathrm{mmol(NO_3)\,m^{-3}}$.

the end of the year, one should consider the way the error covariance matrix is specified. The error covariance matrix is computed using free run model outputs on a 2-month period running window over the year. As the free run variabil-ity is weaker than the variability revealed by observations on a 2-month period, the EOF basis is not able to capture the actual variability of the ecosystem in an efficient manner. This issue is related to the fixed-based variant of the SEEK

filter chosen to assimilate data and could be attenuated by increasing the temporal windows during which the EOF are computed. However, one must be aware that it could lead to erroneous state corrections as it will include in the EOF computation numerous states having an extremely low probability to happen. This seasonal variation in the spreading of the non-stochastic ensemble used here is illustrated by Fig. 13 of Brankart et al. (2012).

The spreading of the ensemble obtained after the computational setup of the EOF basis as described above raises the question of the performance of anamorphosis transformations in such a situation. In the non-linear assimilation procedure, the analysis is constrained to be in the range of values defined by the historical ensemble. As a result, the transformation back to the original space may induce by itself a truncation of the analyzed values. To assess whether the truncations play an important role in the non-linear procedure, it is relevant to check how every assimilated observation compares to the maximum of the historical ensemble used to compute the anamorphosis transformation for the phytoplankton variable (assimilated variable in the CPBM). However, computing strictly how many observations are above the ensemble maximum does not account for observation errors, as the observations are characterized by a probability distribution rather than by an absolute value. We therefore computed the number of assimilated data for which the 95 % confident interval of the observation PDF does not overlap the concerned local ensemble. We previously defined the observation error as 30 % of the considered data, so the minimum bound of this 95 % confident interval corresponds to 40 % of the considered data (observation minus 2 times the standard deviation). Thus, when 40 % of the observation value is above the ensemble maximum, we consider that the observation PDF does not overlap the local ensemble, indicating that the ensemble spread resulting from the EOF computation within 2-month periods is possibly too small.

Figure 4 shows the percentage of model grid points where the observation PDF does not overlap the local ensemble used to compute the anamorphosis transformation. This percentage is shown for the eight regions considered with respect to time. The percentage of observations for which 40 % of the value exceeds the local ensemble is generally weak (below 10 %) for regions 1, 4, 5, 6, 7, and 8. It should be noted, however, that the actual percentage of observations discarded by anamorphosis truncations is evidently larger than the numbers shown in Fig. 4. The seasonal distribution of this percentage is centered around the bloom period, especially for high latitude regions (1 and 2). regions 2 (North Sea) and 3 (Gulf of St. Lawrence) show inversely high percentages, up to 40 % of the whole assimilated data. In addition, the mean averaged percentage of observations exceeding the local ensemble for regions 2 and 3 are 16 % and 32 %, respectively. These regions are also the ones showing the most contrasted differences between the linear and non-linear run (see Figs. 2 and 3), indicating that the truncations

of the assimilated observation are, at least partly, responsible for these divergences. While this link is clear for regions 2 and 3, it is less clear for region 8 but cannot be totally discriminated. Indeed, in this region the percentage of observations out of the ensemble remains low along the simulation period but the region is spatially extended. In that configuration, even a low percentage could nevertheless hide the same process as highlighted for regions 2 and 3.

It is likely from Fig. 4 that the spread of the ensemble obtained for regions 2 and 3 is too small to capture the observation information available over these regions. By increasing the time window for computing the EOFs above two months, the spread of the ensemble could become larger but, at the same time, spurious correlations might occur more frequently, e.g., between phytoplankton and nitrate concentration values. From a global point of view, surface nitrate concentrations as modeled by the linear run (Fig. 3c) clearly differ from the climatology, the free run and the anamorphosis simulations along a seasonal cycle. The minimum surface nitrate concentrations reached after the bloom remain significantly higher than the climatology, suggesting that the multivariate linear analysis fails to estimate coherent surface nitrate patterns. Same conclusions (as for chlorophyll variable) about observations truncation can be drawn for nitrates regarding regions 2, 3 and 8.

Nevertheless, Fig. 3 shows a global drift in nitrate concentrations toward unrealistic values, for high latitude regions (1–4) along the whole seasonal cycle. However, as already discussed above, the amount of surface chlorophyll observations discarded by the anamorphosis transformation for regions 1 and 4 are relatively low and localized in space and time. So these truncations are not sufficient to explain the better performance of the non-linear scheme in terms of nitrate estimations on the whole domain for a complete seasonal cycle.

Another interesting point is that the higher percentages of observations exceeding the local ensemble distribution are found for continental waters (North Sea, Gulf of St. Lawrence and also probably Senegal upwelling). The CPBM used in this paper is clearly not designed to produce realistic results in coastal waters. As an example, river plumes are simply considered as "pure water" runoffs without extra nutrient inputs. However, these specific processes are most relevant in the context of coastal ocean color data assimilation experiments (Fontana et al., 2009, 2010; Ciavatta et al., 2011; Hu et al., 2012). Thus, the ensemble used to compute the anamorphosis transformation does not overlap the real biogeochemical state of ocean in these areas. In that sense, the truncation of the observational information avoids any kind extrapolation outside of the range of the ensemble which may lead to incoherent multivariate correlation structure in the peripheral regions of the state space that have not been explored by the ensemble (see, e.g., Fig. 3c, fourth and fifth rows). And finally, prior to the definition of a complex parametrization of the tails of the anamorphosis function, a
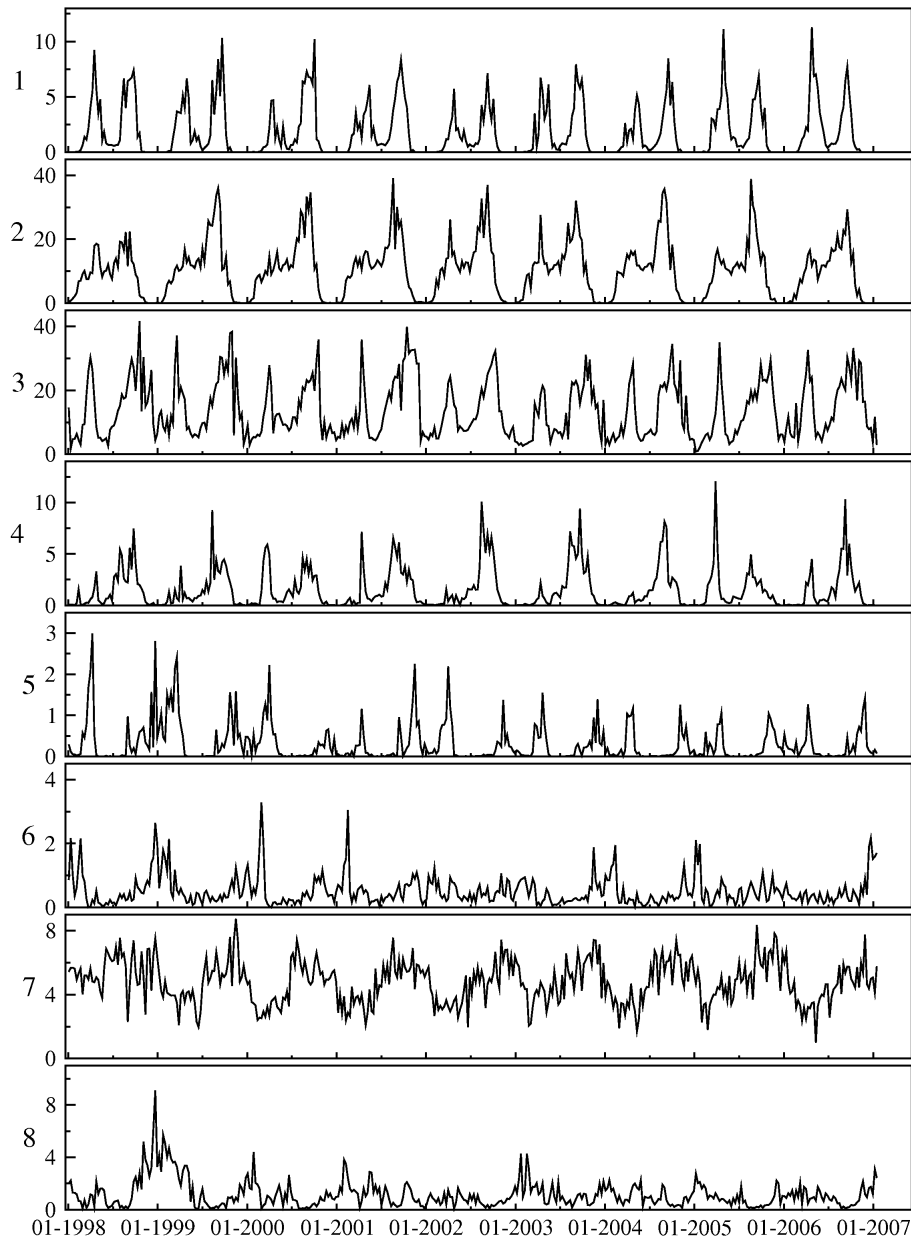
Fig. 4. Percentage of grid cells for which 40 % of the assimilated data exceed the upper bound of the historical ensemble for each biogeochemical region with respect to time.

more realistic simulation of these areas (and thus an ensemble overlapping the real system state) will naturally lower the number of observations truncated.

4.2 Seasonal and interannual variability of the surface chlorophyll

After having investigated the annual cycle of the surface biogeochemical properties of a particular year (2006), we extend the diagnostics to focus on the seasonal-to-interannual variability of the primary production between 1998 and 2007. Figure 5 shows the temporal evolution of the horizontally av-

eraged chlorophyll concentration in the first layer of the coupled model during the simulation period. The data and the model outputs are plotted after time-averaging over 16-day periods. The temporal evolution is shown for 8 biogeochemical provinces adapted from the Longhurst (1995) classification, as defined in Fig. 1.

The green fringe under the SeaWiFS curve is an indicator of the satellite data spatial coverage in the considered region. The fringe thickness vanishes when the considered region is fully covered by satellite data. Conversely, the fringe thickness increases linearly with respect to the number of

missing observations, approaching 100 % of the observation value when almost no data are available. When no data at all are present, no dots and no fringe are drawn. The thickness of the green fringe can thus be interpreted as an observation error index associated with the lack of data.

In region 1, the free run almost systematically overestimates the maximum peak of chlorophyll that occurs during the bloom period (late spring/early summer). This bias is efficiently corrected by the assimilation system, both in the linear and non-linear experiments. Unfortunately, during other seasons the number of chlorophyll observations is sometimes too low (as shown by the large green fringe) to enable efficient corrections of the model estimates. As an example, in 2000 the satellite was not able to collect any ocean color measurement in region 1 during several months in winter. Similar conclusions can be drawn in other high latitude regions, e.g., in region 2 that includes the Baltic and North Seas where the seasonal signal looks very weak. A first conclusion that can be drawn for these regions is that the maximum chlorophyll concentrations during the bloom are efficiently constrained in the range defined by the satellite data. Another point to be underlined is the difference between the linear and non-linear runs, which is generally small, suggesting that the multivariate corrections have similar effects in both experiments.

Region 3 exhibits a seasonal cycle in the SeaWiFS data, with well-marked peaks of chlorophyll in the early spring period. This cycle is reproduced by the free run, but with maximum chlorophyll values that remain much below the measurements. Nevertheless, an overestimation of the chlorophyll content by the SeaWiFS data set cannot be excluded in these coastal waters, notwithstanding the rather poor performance of the CPBM for such coastal areas as discussed previously. The non-linear run slightly increases the chlorophyll concentrations toward the measured values. This is not the case for the linear run, probably because the temporal dynamics are completely modified as a result of strong increments on nitrate concentrations (see Fig. 3c, third and fourth row).

Considering region 4 in the open ocean, it appears that the bloom in the free run starts slightly too late, does not last as long as it should when compared to the data, and reaches values that overestimate the observed peak. The free run and the data are satisfyingly reconciled by the assimilation procedure during the bloom period. The corrections applied during winter are very modest, once again due to a lack of data in this zone. However, the absolute values of chlorophyll concentration are consistently improved all along the simulation period thanks to the assimilation process.

In the mid-latitude region 5, the bloom period is well reproduced by the free run, however with values that are overestimated during the peak of the bloom. This flaw is likely due to the presence of the Gulf Stream pattern discussed above. For this region, the linear as well as anamorphosis runs help to constrain the chlorophyll evolution with a good accuracy. Interestingly, a drift of the chlorophyll concentra-

tions appears after several years in the free simulation, while this trend is removed when applying the assimilation scheme.

In the mid-latitude region 6, the free run performs well in terms of both timing and maximum values of the bloom. The anamorphosis run outstandingly increases the accuracy of the chlorophyll description all along the simulation period. The results are different when considering the linear run. Indeed, after the first year of simulation, the estimated chlorophyll becomes very different from the data and the other simulations. It is a consequence of the high chlorophyll spots that occur in that region after several years of simulation (as discussed in the previous paragraph). The unrealistic corrections applied to non-observed variables in the case of the linear run induce secondary effects on the biogeochemical dynamics that eventually lead to the failure of the method.

In the Gulf of Mexico (region 7), all simulations underestimate the satellite-estimated chlorophyll content while the temporal variability differs from the observations, for similar reasons as in the coastal region 3. This bias remains stable all along the experiment and the corrections applied to chlorophyll are modest.

The subtropical gyre is defined as region 8, for which the free run performs well and the non-linear run slightly better. Corrections applied directly to chlorophyll concentrations are low as a consequence of the low seasonal variability of phytoplankton content in that region and the correspondingly low variability contained in the EOFs. A chlorophyll peak is visible by the end of year 1998; it is actually the signature of an exceptionally extended Northwest African upwelling (not shown) probably linked to the 1997–1998 El Niño changes that may intensify this process (Demarcq, 1998), implying a biogeochemical response of the system (Ohde and Siegel, 2010). Once again, spots of strong nitrate concentrations appear in that region as a consequence of the inappropriate multivariate linear analysis, inducing unrealistic chlorophyll concentrations by the end of the simulation period (see also Fig. 3c; third and fourth row).

Overall, the assessment of the data assimilation experiments shows satisfactory results on the chlorophyll representation along the simulation period. The realism of the temporal evolution of chlorophyll concentrations is improved by the assimilation of chlorophyll SeaWiFS data in the open ocean regions considered here. Moreover, the non-linear assimilation scheme performs better than the linear one. The results are less convincing in coastal regions, a statement that is not surprising since the numerical system (both model and assimilation components) was not specifically set up to perform well in these regions.

4.3 Surface chlorophyll concentration forecast

Additional diagnostics of surface chrorophyll estimates are performed to assess the prediction capacity of the assimilative system, investigating how the model is able to preserve after 8 days some benefit of the assimilation increments
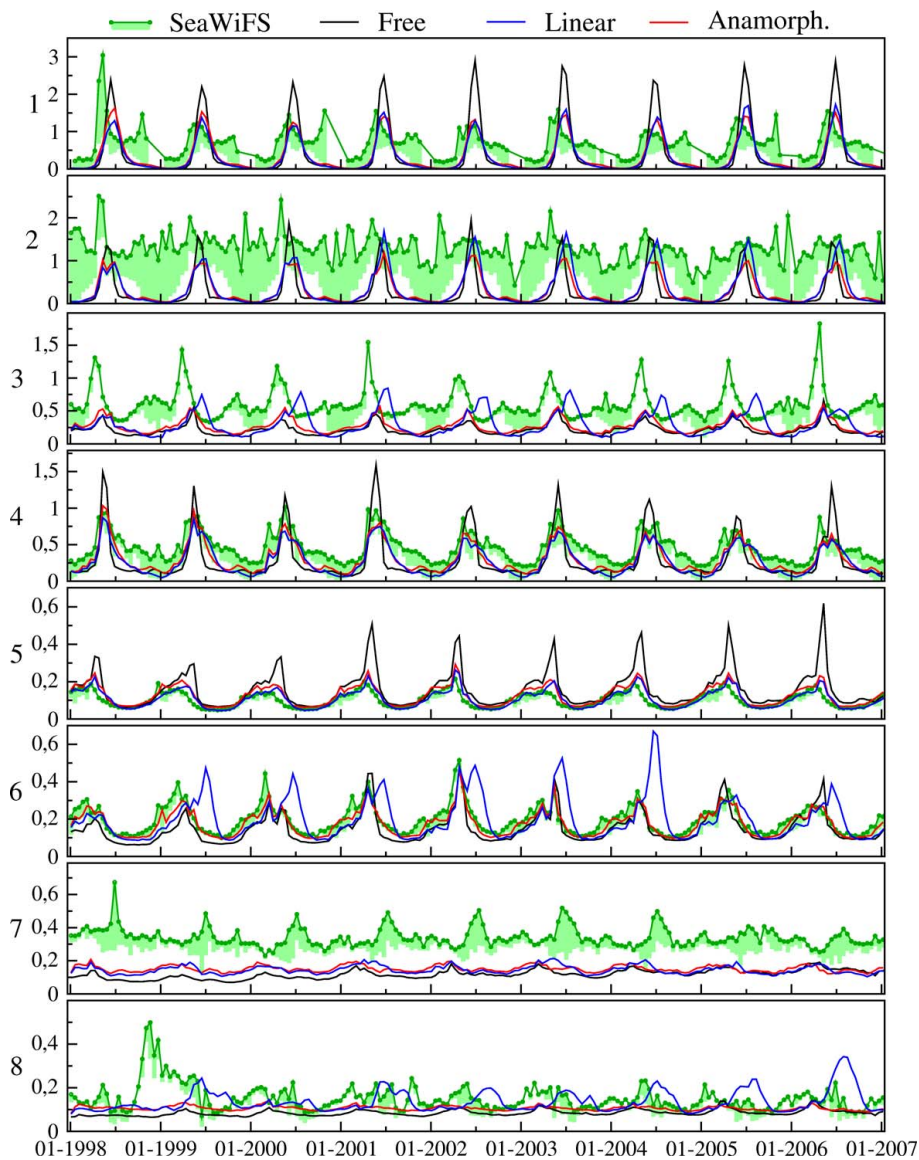
Fig. 5. Surface chlorophyll temporal evolution spatially binned on biogeochemical regions as defined on Fig. 1. Concentrations are given in mg(Chl) m^{-3}.

injected in initial conditions. This forecast diagnostic is a first indication of the relevance of the multivariate correction scheme, as it is frequently observed that initial conditions poorly balanced with respect to the model governing equations tend to reject the assimilation increments very quickly (Hemmings et al., 2008; Ford et al., 2012).

In order to derive a statistical measure of the forecast skill, every individual observation pixel assimilated in the model was compared to its equivalent 8-day forecast computed in the free, linear and anamorphosis runs. The comparison is based on 410 model snapshots and more than 1.9 10^7 SeaWiFS individual pixel data. Figure 6 shows the probability density function (PDF) of the log($C_{SeaWiFS}$/C_{model}) function, where $C_{SeaWiFS}$ and C_{model} are the SeaWiFS and model con-

centration, respectively. It appears that the 8-day forecast of the free run is slightly biased, with an overestimation of the chlorophyll content by the model. This general behavior is mainly due to the overestimation of primary production at high latitudes during the spring bloom. The linear run shows an improvement of these diagnostics by reducing the PDF dispersion, while a strong bias remains visible. The assimilation impact is further improved when considering the non-linear run. Indeed, while the non-linear PDF is equivalent to the linear PDF for extreme values (log error lower than –2 and higher than 2), the PDF maximum is now centered close to 0.
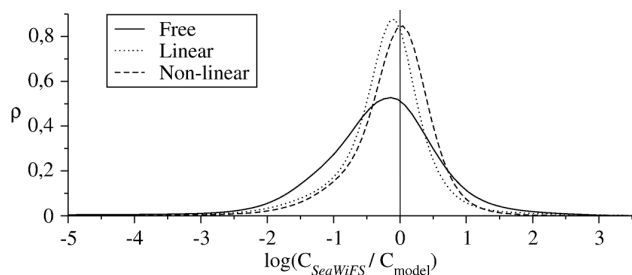
Fig. 6. Normalized histogram of $\log(C_{SeaWiFS}/C_{model})$ function where $C_{SeaWiFS}$ and C_{model} stand for the SeaWiFS and the modeled concentration of chlorophyll, respectively.

We show here that the prediction capacity of the model over a period of 8 days is improved with the assimilation system. In the case where the CPBM would have completely "forgotten" the increment from the previous assimilation step, the PDF as defined above would be slightly modified between the free and assimilation run. Conversely, the persistence of the analysis increment between two assimilation steps is an indication that the forecast skill of the system is improved in the assimilation run. More generally, Fig. 6 illustrates the benefit that can be expected in terms of forecast skill by assimilating ocean color data into a basin-scale CPBM.

4.4 Comparison with independent in situ nitrate measurements

In this section, we investigate how the assimilation is able to propagate the observed information to non-observed quantities (surface and sub-surface nutrients). The nitrate model compartment is chosen here because in situ nitrate observations have been collected with a good coverage in the North Atlantic during the period of reanalysis. These diagnostics will provide further indications that the multivariate scheme is well suited for combining ocean color observations with CPBM predictions.

The WOA09 data set was used to objectively evaluate the reanalysis of nutrient distributions. For comparison, only data measured in deep sea waters (i.e., bottom model topography deeper than 500 meters) were kept in the process. This selection was made to ensure an objective determination of the method efficiency since the CPBM used here is not well designed for shallow waters. Figure 7 shows the in situ data available in the WOA09 data set during the simulation period (depth less than 10 meters). It is apparent that the data set covers the North Atlantic area well, permitting a valuable assessment of the method efficiency within the framework of this realistic experiment. The number of data per years inside of the modeled domain and for a bathymetry higher than 500 m for the period 1998 to 2006 is 1901, 1228, 530, 108, 758, 1405, 619, 0, 0.

A histogram of the $\log(C_{in\ situ}/C_{model})$ function where $C_{in\ situ}$ and C_{model} are the in situ and colocalized model concentrations is shown in Fig. 8 for the surface data (depth less than 10 meters) represented in Fig. 7. It is important to note that the nitrate model concentrations used for the comparison correspond to model forecasts from day 1 to day 8 (actually between two assimilation steps). This histogram, based on 1759 measurements, is shown for the free run, the linear run, the non-linear run and the WOA09 climatology. The free run shows a centered function while some extreme mismatches of both over- and under-estimations appear, resulting in a logarithmic root mean square error (RMS) of 0.82. The histogram of the linear run shows that the number of instances where the model underestimates the measured concentration is reduced by the assimilation process. By contrast, the number of instances where the model overestimates the observations is increased. In this configuration, the RMS of the linear run is 0.87, attesting that the assimilation of satellite chlorophyll did not help to improve the nitrate representation. The result is objectively different when considering the non-linear assimilation scheme, as overestimations remain more or less unchanged compared to the free run while strong underestimations are reduced, yielding an RMS value of 0.72. This result demonstrates that, in terms of nitrate, the forecast was improved by the assimilation of satellite chlorophyll data.

A more interesting point to be underlined is that the histogram of the climatology is very similar to the non-linear one, with a RMS of 0.66. The difference between the climatology and the data from which the climatology was computed may appear surprising. There are several reasons that could explain this difference. Firstly, only a limited number of data included in the WOA09 data set were used to compute the WOA09 climatology as a result of numerous data quality control tests (e.g., range and gradient check; statistical check; subjective flagging; see Garcia et al., 2010) to eliminate questionable data from the climatology computation. Secondly, the temporal binning of data used to compute the climatology at monthly timescale intrinsically induces temporal representativeness errors. The fact that the histogram of the non-linear run looks very similar to the climatology histogram indicates that the differences that may occur between the non-linear run and the in situ data are mainly due to questionable data. Therefore, this comparison should be more significant when considering only confident data.

Following the same methodology, an identical RMS index was computed for several running depth intervals in the euphotic layer. Considered intervals are 0–5, 5–10, 10–50, 30–70, 50–90, 70–110, 90–130, 110–150 and 130–170 meters. Figure 9 shows the mean of the considered depth interval with respect to the corresponding RMS where all data entered the computation (a), and where only data higher than $1\,\mathrm{mmol(NO_3)\,m^{-3}}$ entered the computation (b). A total of 1198 distinct data entered the computation for Fig. 9a while only 646 were kept for Fig. 9b.
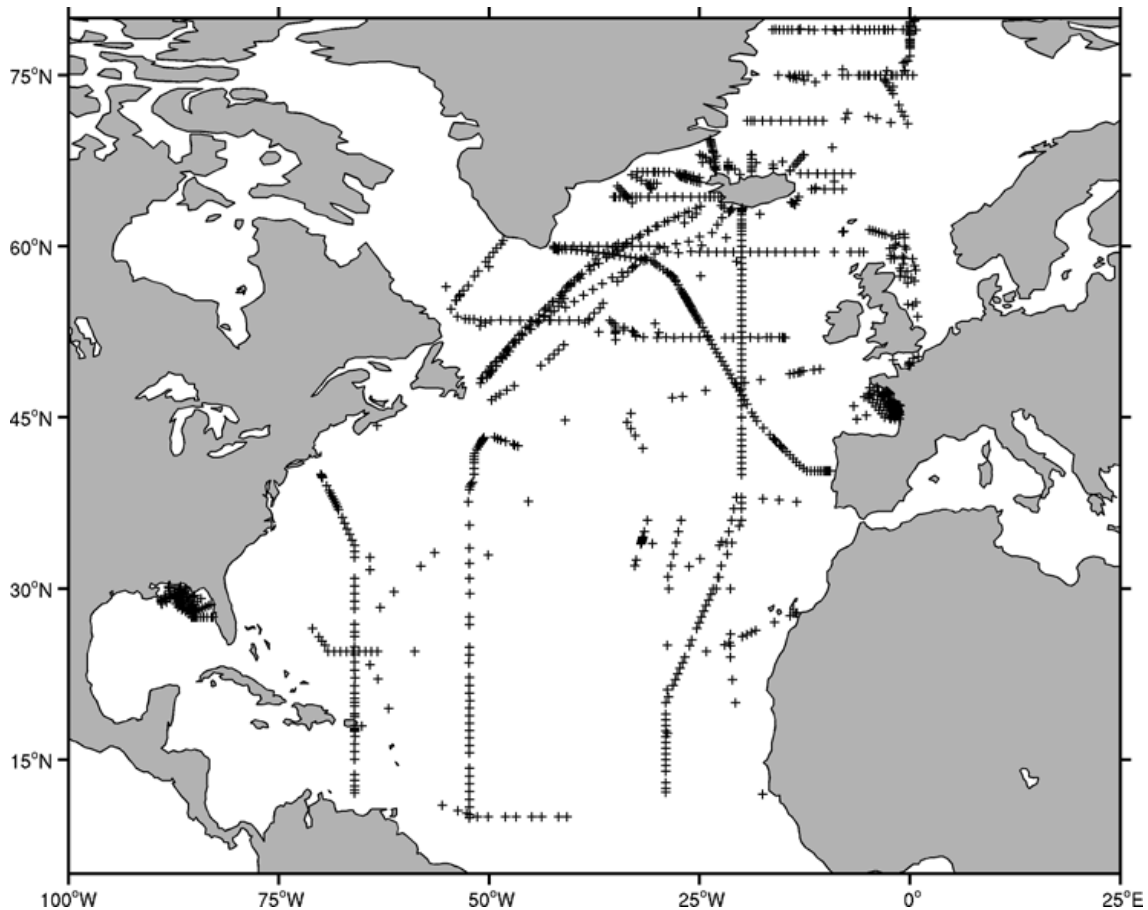
Fig. 7. Nitrate measurements contained in the World Ocean data set for the period 1998–2006 for depths lower than 10 m.

It appears that the assimilation reduces the error of the free run in the 0–5 and 5–10 intervals for the non-linear run and only in the first 0–5 m for the linear run (Fig. 9a). For the deeper part of the euphotic layer, the results are more contrasted and the non-linear assimilation process even increases the error at some depth intervals. Results are improved when considering only measured data greater than $1\,\mathrm{mmol(NO_3)\,m^{-3}}$ (Fig. 9b). In this case, the RMS profile in the non-linear situation is close to the climatology throughout the water column up to the 130–170 m interval. In the 0–5 m interval, the RMS is outstandingly reduced from 0.75 to 0.31 thanks to the assimilation process. The linear run reduces the error of the free run in the first two depth intervals, but increases it almost everywhere in the euphotic layer. There are several reasons to explain why the assimilation process performs better for observed values greater than $1\,\mathrm{mmol(NO_3)\,m^{-3}}$. Indeed, these data are generally measured outside the oligotrophic sub-tropical gyre, where the low temporal variability of biogeochemical concentrations for the free run (used to compute the EOFs) does not allow the assimilation system to correct strong differences between model and data. Secondly, data lower than

$1\,\mathrm{mmol(NO_3)\,m^{-3}}$ are also measured in strong nutrient concentration gradients (around 45° N) where a well-reproduced ocean circulation is essential to obtain satisfying biogeochemical modeling. As the physics was not constrained by data assimilation, these transition zones may not be located at their exact position, explaining the poor performance of the assimilation system.

Figure 10 shows the spatial distribution of the $\log(C_{\mathrm{in\;situ}}/C_{\mathrm{model}})$ function computed for all data of the WOA09 data set included in the 0–10 m depths interval. Here, white dots indicate a weak difference between the model (or climatology) and the observations, blue dots indicate overestimation by the model, and red dots indicate underestimation by the model.

Maps are plotted for the free run (a), the linear run (b), the non-linear run (c) and the climatology (d). Maps were divided into 5 frames to make the discussion of the results clearer. At high-latitude regions (frame 1), the nitrate concentration remains roughly unchanged in the different experiments (a, b, c) as a consequence of the lack of ocean color data in this part of the ocean. The nutrient-enriched area (frame 2) shows the most significant differences between
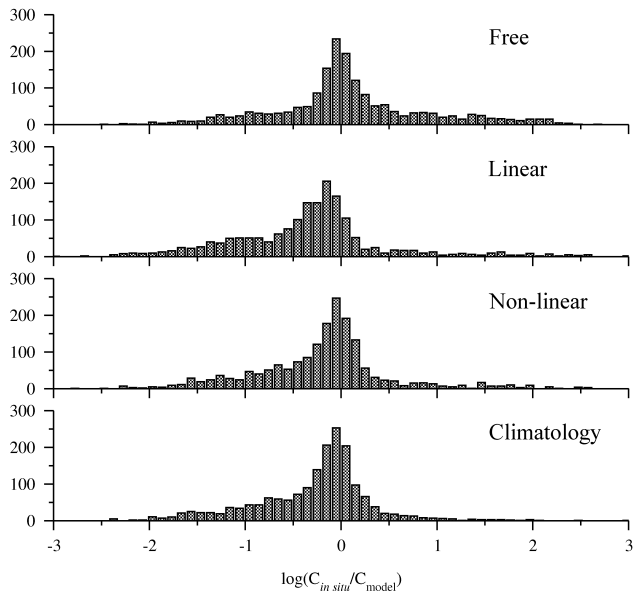
Fig. 8. Histogram of log($C_{\text{in situ}}/C_{\text{model}}$) function where $C_{\text{in situ}}$ and C_{model} stand for the in situ and the modeled concentration of nitrate, respectively.
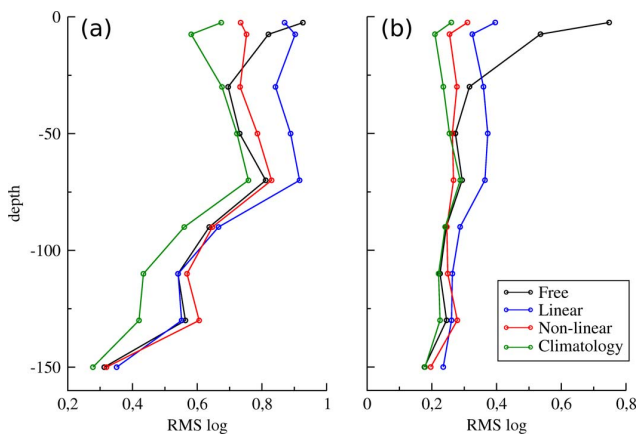

Fig. 9. Nitrate logarithmic RMS between in situ data and model with respect to the depth for: **(a)** all available data; **(b)** only data higher than 1 mmol (NO_3) m^{-3}.

each experiment. While the nitrate concentrations are overestimated and underestimated by (resp.) the free (a) and the linear run (b), the non-linear run (c) shows a clear improvement. In this region, we can see that the non-linear run performs as well as the climatology (d) both in terms of magnitude and error bias. In the western part of the subtropical gyre (frame 3), results remain unchanged for the various experiments, as previously observed for the chlorophyll variable. In the eastern part of the subtropical gyre (frame 4), the situation is the same except in the region of the Northwest African upwelling (25° W – 15° N) where nutrient inputs into the superficial layers of the water column induce a

rather strong biogeochemical response captured by the EOF decomposition. This explains the good behavior of the non-linear run in that part of the ocean, performing even better than the climatology (d). In the Gulf of Mexico (frame 5), none of the modeled situations or climatology perform well, showing that even the WOA climatology is not designed to fit the biogeochemical ocean state in coastal regions.

4.5 Assimilation impact on sub-surface biogeochemical description

To understand the assimilation impact on the biogeochemical variables in the water column, it is interesting to consider a zonal vertical section at 58° N from 60° W to 10° W. This vertical section is plotted for chlorophyll (Fig. 11) and nitrate concentrations (Fig. 12), showing the free (a), linear (b) and non-linear (c) runs. These sections were produced during a chlorophyll bloom (30 days temporal binning between 1 June 2006 and 1 July 2006). As previously stated, this period is the one offering the best assimilation efficiency. It is apparent that the vertical distributions of chlorophyll keep a similar shape (high in the euphotic layer, low deeper) even after data assimilation, whatever method used.

For each assimilation situation, chlorophyll concentrations are also bound by realistic values, relatively close to those of the free run. However, this is not the case when considering nitrate concentrations in the same vertical section (Fig. 12). Indeed, the linear run (b) does not follow the vertical distribution obtained in the other situtations (a and c) and exhibits larger values, as was already visible in Fig. 3 for the first layer of the model. In order to understand the assimilation mechanism responsible for this behavior, one must consider the correlations specified in the analysis scheme between observed and non-observed variables. Figure 13, for instance, shows the correlation of the ensemble on the geographic location 40° W–58° N between the surface phytoplankton (proxy of chlorophyll in our study) and the nitrate concentration along the water column for 1 June 2006. The linear case is represented as black dots while the non-linear one is represented as red dots. The linear run shows extremely high negative correlations in the first two layers of the model, while the correlation rapidly decreases as depth increases. In the deeper part of the water column, the correlation remains stable around –0.2, attesting that the surface phytoplankton concentration is correlated to nitrate even below the euphotic layers. Nervertheless, these correlations are not a sufficient proof of the method efficiency, as we previously showed that nitrate description was not improved in the deeper part of the water column by the linear assimilation process. When looking at the non-linear correlation profile, we see that a significant correlation (between –0.6 and –0.8) up to 50 m depth is visible while it rapidly decreases to approximately 0 below this depth. This 50 m depth is the one for which the method was proved to be efficient for nitrate data (Fig. 9) and is also the typical depth of the euphotic layers, as visible in Figs. 11
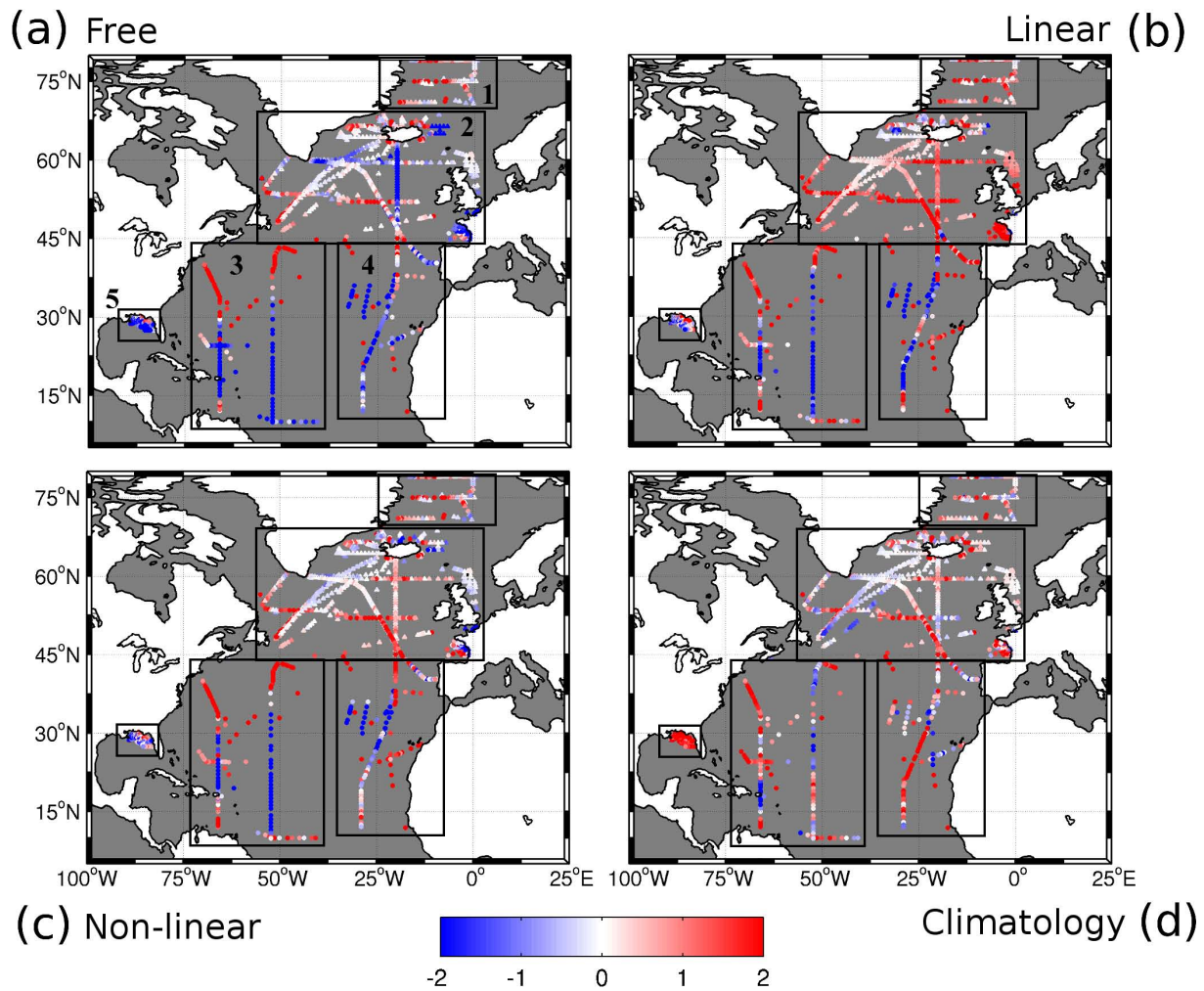
Fig. 10. Spatial distribution of $\log(C_{in\ situ}/C_{model})$ function where $C_{in\ situ}$ and C_{model} stand for the in situ and the modeled concentration of nitrate (resp). Frames a, b, c and d correspond to the free run, the linear run, the non-linear and the climatology, respectively.

and 12. Therefore, mitigated performances of the data assimilation system below the euphotic layer appear to be related to weak correlations between surface phytoplankton and nitrate concentration deeper in the water column. We can thus argue that the correlation profile showed here between surface phytoplankton concentration and nitrate concentration is more realistic when using an anamorphosis transformation. Indeed as expected intuitively, the correlation is high in the euphotic layer and almost null below. This substantial increase in the spatial correlations description is discussed in Brankart et al. (2012) when considering several data assimilation experiments using anamorphosis transformation, including the present one. This analysis of vertical sections and correlation profiles highlights a limitation of the methodology setup within the framework of the present study. Indeed, results indicate that when using an efficiently defined assimilation method (as previously shown by model to independent data set comparison), the surface phytoplankton con-

centration is not correlated to the nitrate concentration below the euphotic layers. And thus we could not expect to control the three-dimensional CPBM using only superficial information such as those brought by remote sensing of ocean color without defining a priori information about biogeochemical concentration vertical distributions.

5 Conclusions and perspectives

In this paper, a state-of-the-art assimilation system was developed to assimilate satellite-derived chlorophyll data into a three-dimensional CPBM of the North Atlantic Ocean. Different simulations were conducted during a 9-year period (1998–2006), allowing the assimilation of 410 SeaWiFS-estimated maps of chlorophyll temporally binned every 8 days. The simulations were performed with a fixed variant of the reduced-rank Kalman filter (SEEK) to limit the computational burden of the assimilation process. Several key
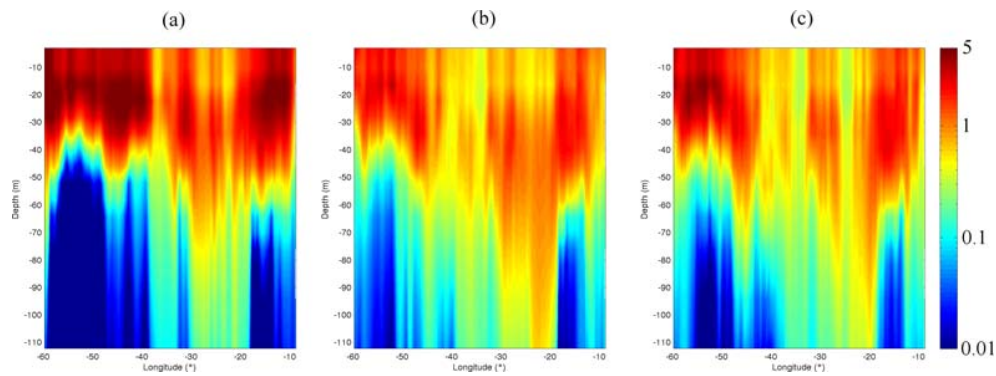
Fig. 11. Vertical section of chlorophyll concentration on 58° N between 60° W and 10° W temporally binned on 30 days between 1 June 2006 and 1 July 2006 for: **(a)** free run; **(b)** linear run ; **(c)** non-linear run. Concentrations are given in mg(Chl) m^{-3}.
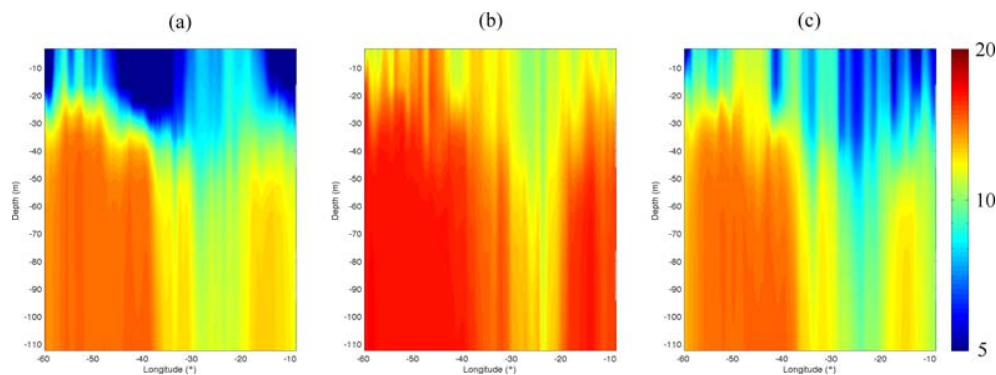

Fig. 12. Vertical section of nitrate concentration on 58° N between 60° W and 10° W temporally binned 30 days between 1 June 2006 and 1 July 2006 for **(a)** free run; **(b)** linear run; and**(c)** non-linear run. Concentrations are given in mmol(NO$_3$) m^{-3}.

parameters entering the analysis scheme (e.g., model and observation error parameterizations, local influence radius) were carefully tuned to maximize the benefit of the assimilation process. Comparisons were made between a free run, an assimilation run using a linear updating scheme, and an assimilation run using a non-linear updating scheme with anamorphic transformations.

These experiments show that the application of anamorphosis yields a non-linear analysis scheme which is identified as a key element of the assimilation performance, without requiring significant increase of computing resources. The assimilation of chlorophyll data in the non-linear configuration efficiently improves the description of the seasonal cycles of surface chlorophyll along the simulation period. The skill of the model to forecast surface chlorophyll concentrations after 8 days without assimilation was also improved by the assimilation process. Temporal evolution of spatially-binned surface chlorophyll concentration showed that the spatial coverage of ocean color data remains a critical point as no ocean color data are available for high-latitude regions during several months each year.

Comparisons between model outputs and an independent set of in situ measurements showed that surface nitrate con-

centration is more accurately estimated by the assimilation of satellite-derived chlorophyll concentration. Indeed, the modeled surface nitrate concentration fields were closer to the climatology generated from the WOA09. The model performance was good in the chlorophyll bloom area (north of 45° N) while it was not systematically improved in the region of the sub-tropical gyre. The gain was limited to the upper layers of the euphotic zone, while the deeper part of the water column was not strongly affected by the assimilation process. This is due to the weak correlations between surface phytoplankton and nitrate below the euphotic layer. Truncations resulting from the definition of the tails of the anamorphosis was found to have a direct impact on the assimilation results for specific regions of the domain. A high rate of truncations was found in areas where the local ensemble used to compute the anamorphosis transformation does not overlap the real system state (as indicated by the observations versus model mismatch), so that truncations avoid extrapolation of system state outside of the range explored by the ensemble. In that sense, a better match between observation and model will naturally lower the number of truncated observations. This dependency between anamorphosis truncations and definition of the historical ensemble deserves to

be more precisely assessed. Nevertheless, the global and persistent difference between the linear and non-linear assimilation system over the model domain cannot be explained only by these truncations. This attests that the correlation between observed and non-observed variables was improved thanks to the anamorphosis transformation. A qualitative look to a multivariate correlation profile between nitrate and phytoplankton along the water column also confirmed this statement. While we have identified the anamorphosis tranformation as a key ingredient for the method's success, some pragmatic choices made here on its tuning (particularly the local ensemble definition) deserve to be further investigated.

Overall, the assessment of the non-linear experiment shows that the assimilation system can be seen as a first prototype, opening perspectives toward reanalyses of the North Atlantic Ocean biogeochemistry during the satellite ocean color era. However, this study also highlights that the full control of a three-dimensional CPBM trajectory is likely to be hopeless with the assimilation of surface chlorophyll data only. Equivalent conclusions regarding the mitigated impact of the satellite chlorophyll assimilation below the surface were recently drawn by Hu et al. (2012) and Ford et al. (2012).

Several solutions can be envisaged to overcome this issue. The first one could be to define a priori assumptions about vertical profiles of biogeochemical variable concentrations. This means that the observation vector would not be limited to the first layer of the model but integrated along the water column. Efforts were recently undertaken to characterize relationships between spatially-sensed chlorophyll concentrations and vertical distribution of phytoplankton content (see, e.g., Uitz et al., 2006). The use of such relationships to propagate information brought by surface chlorophyll concentrations to deeper parts of the water column would require further investigations. While this solution could improve the assimilation performance in open sea waters, a universal a priori assumption on biogeochemical vertical profiles can hardly be defined consistently for coastal areas as various local factors may influence directly the vertical distribution of biogeochemical concentrations (river plumes, waves).

In the long-term, a more promising approach would be to explicitly include deeper observations into the assimilation process. In this way, no a priori assumption about the vertical distribution of biogeochemical variable concentrations would be required, while information would be assimilated explicitly at depth. The critical point here is that it requires a refined array of in situ sensors systematically measuring biogeochemical properties at basin-scale; to some degree an equivalent to the ARGO float dedicated to biogeochemical measurements. Such a sampling array currently does not exist but efforts are ongoing to deploy autonomous biogeochemical sensors in deep sea waters that are able to measure precisely chlorophyll and nitrate profiles (Claustre et al., 2010a, b). Since such a large-scale data set is intended in the foreseeable future, investigations to optimally combine
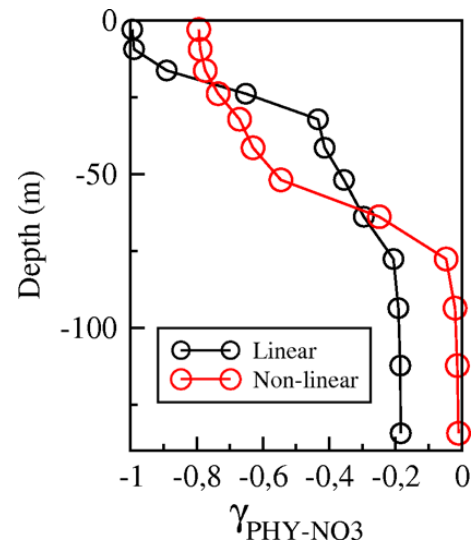

Fig. 13. Model correlation between surface phytoplankton and nitrate concentration with respect to the depth at location 58° N–40° W on 1 June 2006.

satellite and in situ biogeochemical data through observing system simulation experiments would be a straightforward extension of the present study.

A large-scale data set is also required to better assess assimilation efficiency with respect to other components of the biogeochemical model (e.g., ammonium, zooplankton). While one could reasonably hope that the assimilation of satellite chlorophyll should have a positive impact on the rest of the biogeochemical model, we were not able to prove it as things stand today.

Upgrading the simplified assimilation scheme used here to a more sophisticated one should also help in improving the reanalysis presented in this paper. Indeed, we computed uncertainties based on the free simulation without error propagation along the simulation. As a consequence, assimilation showed poor performance in areas where model variability was lower than data variability. Specifying the uncertainties in a more complex way (e.g., ensemble Kalman filter, error propagation) will certainly lead to improving the assimilation performance.

Edited by: M. Hoppema

References

Barnier, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Le Sommer, J., Beckmann, A., Biastoch, A., Boning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., and De Cuevas, B.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution, Ocean Dynam., 56, 543–567, doi:10.1007/s10236-006-0082-1, 2006.

Berline, L., Brankart, J.-M., Brasseur, P., Ourmières, Y., and Verron, J.: Improving the physics of a coupled physical-biogeochemical model of the North Atlantic through data assimilation: Impact on the ecosystem, J. Mar. Syst., 64, 153–172, doi:10.1016/j.jmarsys.2006.03.007, 2007.

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, Int. Stat. Rev., 71, 223–241, doi:10.1111/j.1751-5823.2003.tb00194.x, 2003.

Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation, Mon. Wea. Rev., 138, 2997–3023, doi:10.1175/2010MWR3164.1, 2010.

Boyer, T. P., Antonov, J., Garcia, H., Johnson, D., Locarnini, R., Mishonov, A., Pitcher, M., Baranova, O., and Smolyar, I.: World Ocean Database 2005, U.S. Government Printing Office, Washington D.C., 2006.

Brankart, J.-M., Testut, C., Béal, D., Doron, M., Fontana, C., Meinvielle, M., Brasseur, P., and Verron, J.: Towards an improved description of ocean uncertainties: effect of local anamorphic transformations on spatial correlations, Ocean Sci., 8, 2147–2195, doi:10.5194/osd-8-2147-2011, 2012.

Brasseur, P. and Verron, J.: The SEEK filter method for data assimilation in oceanography: a synthesis, Ocean Dynam., 56, 650–661, 2006.

Brasseur, P., Ballabrera-Poy, J., and Verron, J.: Assimilation of altimetric data in the mid-latitude oceans using the Singular Evolutive Extended Kalman filter with an eddy-resolving, primitive equation model, J. Mar. Syst., 22, 269–294, doi:10.1016/S0924-7963(99)00044-5, 1999.

Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y., and Verron, J.: Characterization of mixing errors in a coupled physical biogeochemical model of the North Atlantic: implications for nonlinear estimation using Gaussian anamorphosis, Ocean Sci., 6, 247–262, doi:10.5194/os-6-247-2010, 2010.

Carmillet, V., Brankart, J.-M., Brasseur, P., Drange, H., Evensen, G., and Verron, J.: A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical-biochemical model of the North Atlantic ocean, Ocean Model., 3, 167–192, doi:10.1016/S1463-5003(01)00007-5, 2001.

Ciavatta, S., Torres, R., Saux-Picart, S. and Allen, J. I.: Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? J. Geophys. Res., 116, C12043, doi:10.1029/2011JC007219, 2011.

Claustre, H., Morel, A., Hooker, S. B., Babin, M., Antoine, D., Oubelkheir, K., Bricaud, A., Leblanc, K., Quéguiner, B., and Maritorena, S.: Is desert dust making oligotrophic waters greener?, Geophys. Res. Lett., 29, 1469, doi:10.1029/2001GL014056, 2002.

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., and Yoder, J.: Guidelines Towards an Integrated Ocean Observation System for Ecosystems and Biogeochemical Cycles, In Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1), Venice, Italy, 21–25 September 2009, edited by: Hall, J., Harrison, D. E., and Stammer, D., ESA Publication WPP-306, doi:10.5270/OceanObs09.pp.14, 2010a.

Claustre, H., Bishop, J., Boss, E., Stewart, B., Berthon, J.-F., Coatanoan, C., Johnson, K., Lotiker, A., Ulloa, O., Perry, M.-J., D'Ortenzio, F., Hembise Fanton D'Andon, O., and Uitz, J.: Bio-Optical Profiling Floats as New Observational Tools for Biogeochemical and Ecosystem Studies: Potential Synergies with Ocean Color Remote Sensing, In Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21–25 September 2009, edited by: Hall, J., Harrison, D. E., and Stammer, D., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.17, 2010b.

Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M., Mogensen, K., Oke, P., Testut, C., Verron, J., and Weaver, A.: Ocean data assimilation systems for GODAE, Oceanography, 22, 96–109, doi:10.5670/oceanog.2009.69, 2009.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Demarcq, H.: Spatial and temporal dynamics of the upwelling off Senegal and Mauritania: local change and trend, Global versus local changes in upwelling systems, Orstom ed. Paris, 149–166, 1998.

Doney, S. C.: Major challenges confronting marine biogeochemical modeling, Global Biogeochem. Cy., 13, 705–714, 1999.

Doron, M., Brasseur, P., and Brankart, J.-M.: Stochastic estimation of biogeochemical parameters of a 3D ocean coupled physical-biogeochemical model: Twin experiments, J. Mar. Syst., 87, 194–207, doi:10.1016/j.jmarsys.2011.04.001, 2011.

Feldman, G. and McClain, C.: Ocean Color Web, SeaWiFS Reprocessing 2009.1, NASA Goddard Space Flight Center, edited by: Kuring, N. and Bailey, S. W., http://oceancolor.gsfc.nasa.gov/, 2010.

Fontana, C., Grenz, C., Pinazo, C., Marsaleix, P., and Diaz, F.: Assimilation of SeaWiFS chlorophyll data into a 3D-

coupled physical-biogeochemical model applied to a freshwater-influenced coastal zone, Cont. Shelf Res., 29, 1397–1409, doi:10.1016/j.csr.2009.03.005, 2009.

Fontana, C., Grenz, C., and Pinazo, C.: Sequential assimilation of a year-long time-series of SeaWiFS chlorophyll data into a 3D biogeochemical model on the French Mediterranean coast, Cont. Shelf Res., 30, 1761–1771, doi:10.1016/j.csr.2010.08.003, 2010.

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., and Demaria, J.: Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model, Ocean Sci. Discuss., 9, 687–744, doi:10.5194/osd-9-687-2012, 2012.

Fukumori, I.: Assimilation of TOPEX sea level measurements with a reduced-gravity, shallow water model of the tropical Pacific Ocean, J. Geophys. Res., 100, 25027–25039, 1995.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., and Johnson, D. R.: World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate), edited by: Levitus, S., NOAA Atlas NESDIS 71, U.S. Government Printing Office, Washington, D.C., 398 pp, 2010.

Gregg, W. W.: Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, J. Mar. Syst., 69, 205–225, doi:10.1016/j.jmarsys.2006.02.015, 2008.

Gregg, W. W. and Casey, N. W.: Global and regional evaluation of the SeaWiFS chlorophyll data set, Remote Sens. Environ., 93, 463–479, doi:10.1016/j.rse.2003.12.012, 2004.

Gregg, W. W., Friedrichs, M. A., Robinson, A. R., Rose, K. A., Schlitzer, R., Thompson, K. R., and Doney, S. C.: Skill assessment in ocean biological data assimilation, J. Mar. Syst., 76, 16–33, doi:10.1016/j.jmarsys.2008.05.006, 2009.

Hemmings, J. C., Barciela, R. M., and Bell, M. J.: Ocean color data assimilation with material conservation for improving model estimates of air-sea CO2 flux, J. Mar. Res., 66, 87–126, doi:doi:10.1357/002224008784815739, 2008.

Hu, J., Fennel, K., Mattern, J. P., and Wilkin, J.: Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight, J. Mar. Syst., 94, 145–156, doi:10.1016/j.jmarsys.2011.11.016, 2012.

Johnson, K., Berelson, W., Boss, E., Chase, Z., Claustre, H., Emerson, S., Gruber, N., Körtzinger, A., Perry, M., and Rise, S.: Observing biogeochemical cycles at global scales with profiling floats and gliders: Prospects for a global array, Oceanography, 22, 216–225, 2009.

Large, W. and Yeager, S.: Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies, NCAR Technical Note: NCAR/TN-460+STR. CGD Division of the National Center for Atmospheric Research, 2004.

Lavender, S., Pinkerton, M., Moore, G., Aiken, J., and Blondeau-Patissier, D.: Modification to the atmospheric correction of SeaWiFS ocean colour images over turbid waters, Cont. Shelf Res., 25, 539–555, doi:10.1016/j.csr.2004.10.007, 2005.

Levitus, S., Boyer, T. P., Conkright, M. E., O'Brien, T., Antonov, J., Stephens, C., Stathoplos, L., Johnson, D., and Gelfeld, R.: NOAA Atlas NESDIS 18, WORLD OCEAN DATABASE, Vol. 1: Introduction, 1998.

Longhurst, A.: Seasonal cycles of pelagic production and consumption, Prog. Oceanogr., 36, 77–167, doi:10.1016/0079-6611(95)00015-1, 1995.

Lévy, M., Gavart, M., Mémery, L., Caniaux, G., and Paci, A.: A four-dimensional mesoscale map of the spring bloom in the northeast Atlantic (POMME experiment): Results of a prognostic model, J. Geophys. Res., 110, C07S21, doi:10.1029/2004JC002588, 2005.

Natvik, L. J., Eknes, M., and Evensen, G.: A weak constraint inverse for a zero-dimensional marine ecosystem model, J. Mar. Syst., 28, 19–44, doi:10.1016/S0924-7963(00)00082-8, 2001.

Natvik, L. J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments, J. Mar. Syst., 40-41, 127–153, doi:10.1016/S0924-7963(03)00016-2, 2003.

Nerger, L. and Gregg, W. W.: Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter, J. Mar. Syst., 68, 237–254, doi:10.1016/j.jmarsys.2006.11.009, 2007.

Ohde, T. and Siegel, H.: Biological response to coastal upwelling and dust deposition in the area off Northwest Africa, Cont. Shelf Res., 30, 1108–1119, doi:10.1016/j.csr.2010.02.016, 2010.

Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J.-M., and Verron, J.: On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean, J. Mar. Syst., 75, 100–115, doi:10.1016/j.jmarsys.2008.08.003, 2009.

Pham, D. T., Verron, J., and Roubaud, M. C.: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Mar. Syst., 16, 323–340, doi:10.1016/S0924-7963(97)00109-7, 1998.

Simon, E. and Bertino, L.: 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 Sci., 5, 495–510, doi:10.5194/os-5-495-2009, 2009.

Simon, E. and Bertino, L.: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model, J. Mar. Sys., 89, 1–18, doi:10.1016/j.jmarsys.2011.07.007, 2012.

Stammer, D., Köhl, A., Awaji, T., Balmaseda, M., Behringer, D., Carton, J., Ferry, N., Fischer, A., Fukumori, I., Giese, B., Haines, K., Harrison, E., Heimbach, P., Kamachi, M., Keppenne, C., Lee, T., Masina, S., Menemenlis, D., Ponte, R., Remy, E., Rienecker, M., Rosati, A., Schröter, J., Smith, D., Weaver, A., Wunsch, C., and Xue, Y.: Ocean information provided through ensemble ocean synthesis, in: Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21–25 September 2009, edited by: Hall, J., Harrison, D. E., and Stammer, D., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.85, 2010.

Treguier, A. M., Barnier, B., de Miranda, A. P., Molines, J. M., Grima, N., Imbard, M., Madec, G., Messager, C., Reynaud, T. and Michel, S.: An eddy-permitting model of the Atlantic circulation: Evaluating open boundary conditions, J. Geophys. Res., 106, C10, 22115–22129, doi:10.1029/2000JC000376, 2001.

Uitz, J., Claustre, H., Morel, A., and Hooker, S. B.: Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll, J. Geophys. Res., 111, C08005, doi:10.1029/2005JC003207, 2006.

Wilson, C.: The rocky road from research to operations for satellite ocean-colour data in fishery management, ICES J. Mar Sci., 68, 677–686, doi:10.1093/icesjms/fsq168, 2011.