Journal of Marine Systems 94 (2012) 145–156

Contents lists available at SciVerse ScienceDirect

Journal of Marine Systems

journal homepage: www.elsevier.com/locate/jmarsys

Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight

Jiatang Hu [a,*], Katja Fennel [a], Jann Paul Mattern [a,b], John Wilkin [c]

[a] Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
[b] Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada
[c] Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA

ARTICLE INFO

Article history:
Received 23 March 2011
Received in revised form 11 November 2011
Accepted 22 November 2011
Available online 1 December 2011

Keywords:
Data assimilation
Ensemble Kalman Filter
State estimation
Regional Ocean Modeling System
Biological model
Satellite ocean chlorophyll data

ABSTRACT

A multivariate sequential data assimilation approach, the Localized Ensemble Kalman Filter (LEnKF), was used to assimilate daily satellite observations of ocean chlorophyll into a three-dimensional physical–biological model of the Middle Atlantic Bight (MAB) for the year 2006. Covariance localization was applied to make the EnKF analysis more effective by removing spurious long-range correlations in the ensemble approximation of the model's covariance. The model is based on the Regional Ocean Modeling System (ROMS) and coupled to a biological nitrogen cycle model, which includes seven state variables: chlorophyll, phytoplankton, nitrate, ammonium, small and large detrital nitrogen, and zooplankton. An ensemble of 20 model simulations, generated by perturbing the biological parameters according to assumed probability distributions, was used. Model fields of chlorophyll, phytoplankton, nitrate and zooplankton were updated at all vertical layers during LEnKF analysis steps, based on their cross-correlations with surface chlorophyll (the observed variable). The performance of the LEnKF scheme, its influence on the model's predictive skill and on surface particulate organic matter concentrations and primary production are investigated. Estimates of surface chlorophyll and particulate organic carbon are improved in the data-assimilative simulation when compared to one without any assimilation, as is the model's predictive skill.

1. Introduction

Direct observation and numerical simulation are two important means for understanding marine ecosystems: observations contain information about the true ocean state and serve as a crucial source for model calibration and validation; and numerical ocean models are becoming increasingly powerful for predicting physical, biogeochemical and biological processes in the ocean and can be used to support marine management and decision-making. However, field measurements are spatially and temporally limited due to budget, technical and time constraints, and models are always a simplification of the truth and never completely realistic. Given this, it is essential to integrate observations and numerical models in order to achieve the most accurate estimates of the true ocean state. Such integration is best realized through data assimilation methods, which optimally merge the information contained in observations and dynamical models. In the context of biological modeling two qualitatively different approaches to data assimilation are used: 1) parameter estimation, which finds an optimal parameter set by minimizing the misfit between the model and observations (e.g., Spitz et al., 1998; Fennel et al., 2001; Friedrichs et al., 2006; Bagniewski et al., 2011), and 2) state estimation, i.e. updating the model state by utilizing the available observations as the model is integrated forward in time (e.g. Natvik and Evensen, 2003; Nerger and Gregg, 2007; Ourmieres et al., 2009; Mattern et al., 2010).

A widely used method for state estimation is the Ensemble Kalman Filter (EnKF), first introduced by Evensen (1994) to resolve problems associated with the traditional extended Kalman filter for cases of strongly nonlinear dynamics and large state spaces (Burgers et al., 1998; Evensen, 2003, 2009). The EnKF is a sequential assimilation technique that uses the Monte Carlo approach through ensemble integrations. The scheme has proven to be feasible for complex oceanographic problems including physical and biological applications and has been successfully applied to coupled physical–biological models ranging from one-dimensional (e.g., Eknes and Evensen, 2002; Allen et al., 2003; Mattern et al., 2010) to three-dimensional (e.g., Brusdal et al., 2003; Natvik and Evensen, 2003). The EnKF is able to provide updates for the whole model state even if only one variable is assimilated. This property is a consequence of the multivariate nature of the EnKF, which uses cross-correlations between different state variables. Reviews on the EnKF and its applications can be found in Evensen (2003, 2006).

In this study, satellite observations of ocean chlorophyll were assimilated into a three-dimensional physical–biological model of the

* Corresponding author. Tel.: +1 86 84114978; fax: +1 902 494 3877.
E-mail address: jiatang.hu@dal.ca (J. Hu).

0924-7963/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmarsys.2011.11.016

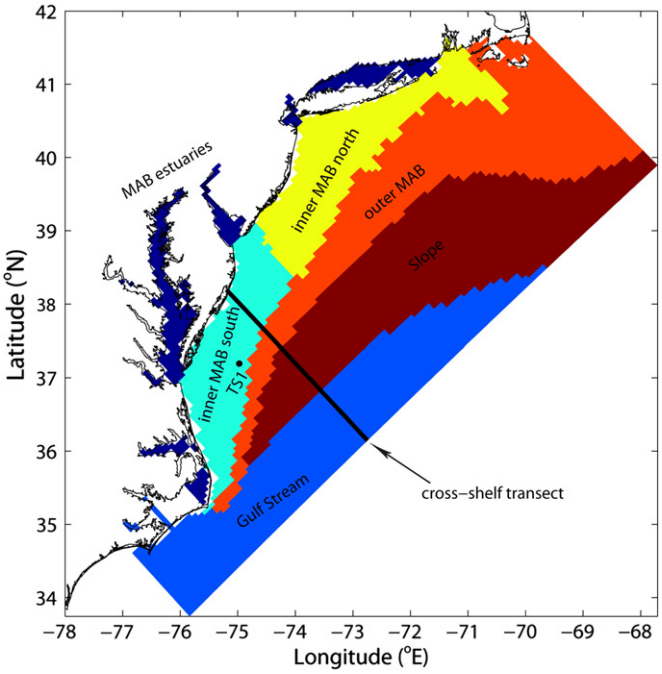

Fig. 1. Map of the Middle Atlantic Bight (MAB) and its subareas used for evaluating changes in total mass of nitrogen due to data assimilation. TS1 is a station used for showing the evolution of the ensemble distributions.

Middle Atlantic Bight (MAB, see Fig. 1) using a modified version of the EnKF, the Localized EnKF (LEnKF). The LEnKF includes a covariance localization technique (Houtekamer and Mitchell, 2001) to alleviate the known issue of the EnKF of spurious long-range correlations caused by limited ensemble sizes. The MAB is the section of the eastern North American continental shelf that extends from Nantucket Shoals in the north to Cape Hatteras in the south. Primary production in this system is nitrogen limited, and the seasonal cycle of chlorophyll and primary production is typical of a temperate continental shelf. (For more details on the MAB, see Fennel et al., 2006 and Lehmann et al., 2009). The Regional Ocean Modeling System (ROMS; http://myroms.org; Haidvogel et al., 2008) coupled with the biological nitrogen cycle model of Fennel et al. (2006, 2008) was used as the dynamical model. The main objectives of this study are 1) to investigate the performance of the LEnKF when assimilating satellite ocean chlorophyll into the biological model and to assess improvements in the predictive skill for unassimilated variables (in this case satellite-based estimates of particulate organic carbon) and for forecasts, and 2) to analyze the impact of the multivariate data assimilation on ecosystem evolution in terms of primary production and the ecological state.

This paper is organized as follows: an introduction to the coupled model, the LEnKF methodology and satellite data used for assimilation and skill assessment is provided in Section 2; details of the assimilation experiments are presented in Section 3; in Section 4 three statistical metrics used to quantify model/data consistency are presented; and results are discussed and summarized in Sections 5 and 6, respectively.

2. Materials and methods

2.1. Model description

2.1.1. Physical model

The ROMS ocean model solves the hydrostatic, Boussinesq, primitive equations in terrain-following coordinates on a structured horizontal curvilinear grid. For efficiency, it employs the split-explicit formulation that advances the depth-integrated continuity and momentum equations with a much smaller time step than the 3-dimensional baroclinic momentum and tracer equations. The ROMS computational kernel is described by Shchepetkin and McWilliams (2005, 2009) and will not be detailed here. Certain features of ROMS are attractive for coupled physical–biological modeling on continental shelves. These include a formulation of the depth-integrated mode equations that prevents aliasing (Higdon and de Szoeke, 1997) of unresolved signals into the slow baroclinic mode while accurately representing barotropic motions resolved by the baroclinic time step (e.g., tides and coastal-trapped waves). Several aspects of the kernel minimize pressure-gradient force truncation errors that would otherwise be problematic in terrain following coordinates with the steep bathymetry. A finite-volume, finite-time-step discretization for the tracer equations improves integral conservation and constancy preservation in coastal applications where the free surface displacement can be a significant fraction of the water depth. From among the several advection algorithm options available in ROMS, the MPDATA (multidimensional positive definite advection transport algorithm) scheme (Smolarkiewicz, 1984) was selected here because its positive-definite property is particularly attractive for biological tracers. A monotonized, high-order vertical advection scheme for sinking biological particulate matter integrates depositional flux over multiple grid cells so it is not constrained by the vertical CFL criterion (Warner et al., 2008). The parameterization of vertical turbulent mixing is the *k–kl* option within the Generic Length Scale scheme as implemented in ROMS by Warner et al. (2005).

Air–sea fluxes of momentum and heat were computed using standard bulk formulae (Fairall et al., 2003) using atmospheric marine boundary layer conditions from the North American Regional Reanalysis (Mesinger et al., 2006) in conjunction with the sea surface temperature computed by ROMS. The vertical profile of solar shortwave radiation is parameterized by two exponential functions following Paulson and Simpson (1977) for assumed Jerlov water type 1. River inflows are based on daily average observations of river discharge from U.S. Geological Survey gauging stations on the Hudson, Delaware and Connecticut rivers, and the four largest rivers entering the Chesapeake Bay, modified to include ungauged portions of the watershed.

Open boundary conditions for temperature, salinity, and sub-tidal frequency velocity and sea level are taken from the same MAB and Gulf of Maine (MABGOM) regional model (R. He and K. Chen, unpublished manuscript) used as boundary conditions to the MAB shelf-break front simulations of Chen and He (2010). Tides are added to the low frequency boundary velocity and sea level variability using harmonics from a regional tide model (Mukai et al., 2002). The model domain (see Fig. 1) has 130×82 grid cells in the horizontal direction and 36 layers in the vertical direction. The horizontal resolution of the model grid varies from ~5.5 km in the inner MAB to ~8.0 km in the outer MAB.

2.1.2. Biological model

The biological ROMS module used has been applied extensively in studies of nitrogen and carbon cycling for the North American east coast continental shelves (Fennel et al., 2006, 2008; Fennel and Wilkin, 2009; Previdi et al., 2009; Druon et al., 2010). The model has seven state variables: chlorophyll, phytoplankton, nitrate, ammonium, small and large detrital nitrogen, and zooplankton. Photoacclimation is included through the use of a time-varying ratio of chlorophyll to phytoplankton biomass (Geider et al., 1996). For details of the model structure and governing equations see Fennel et al. (2006). The biological model parameters used here are as in Fennel et al. (2008); Table 1 lists some of the key parameters to which the biological model is highly sensitive (Mattern, 2008). These parameters are those varied (by adding perturbations drawn from the log-normal distributions given in Table 1) in order to create

Table 1
Assumed probability distributions of biological parameters used for ensemble integrations.

Parameter description	Expected value (E)	Distribution	Unit
Phytoplankton growth rate at 0 °C	0.69	Log-N(−0.526, 0.309)	d^{-1}
Phytoplankton mortality rate	0.15	Log-N(−2.388, 0.982)	d^{-1}
Maximum chlorophyll to carbon ratio	0.0535	Log-N(−3.796, 1.736)	mg Chl mg C^{-1}
Initial slope of the photosynthesis–irradiation curve	0.025	Log-N(−4.888, 2.398)	mg C (mg Chl W m^{-2} $d)^{-1}$
Vertical sinking velocity for phytoplankton	0.1	Log-N(−2.929, 1.253)	m d^{-1}

Note that the expected values (E) are taken from Fennel et al. (2006). The log-normal distributions, Log-N(μ, σ), use E as the mean and $E/4$ as the variance (V_{ar}): $\mu = \ln(E) - \frac{1}{2}\ln\left(1+\frac{V_{ar}}{E^2}\right)$, $\sigma^2 = \ln\left(1+\frac{V_{ar}}{E^2}\right)$.

an ensemble of model states for the assimilation. Note that the perturbed parameters are mostly involved in the dynamics of phytoplankton and thus they have a direct impact on the concentrations of chlorophyll and phytoplankton as well as related processes (e.g., primary production). Other state variables including ammonium, nitrate, detrital nitrogen, and zooplankton, although are affected indirectly as they will respond to the changed fields during model integration.

The coupled biological model was initialized with model fields of chlorophyll, phytoplankton, nitrate, ammonium, detrital nitrogen and zooplankton for January 1, 2006 from a larger scale simulation of the Northeast North American (NENA) shelf model (Fennel et al., 2008) and run to December 31, 2006. Concentrations of the biological state variables along the open boundaries are also obtained from the NENA simulation. River concentrations of ammonium, nitrate, and organic nitrogen from the U.S. Geological Survey monitoring database were used to derive a monthly climatology for these variables and subsequently multiplied with the freshwater flux to yield the river inputs of nutrients and detrital nitrogen (Fennel et al., 2006).

2.2. The LEnKF methodology

The EnKF (Evensen, 1994) is a sequential assimilation method that uses ensemble integrations in order to approximate the evolution of error statistics (i.e., estimates of the model's mean state and error covariances) through time. By integrating an ensemble of model states forward in time one can approximate the time evolution of the model state's probability density function and its associated error statistics. The method proceeds in two steps: whenever observations become available, the statistical information contained in the ensemble is combined with the observations to update the model states (referred to as analysis step); after this update, the model states are integrated forward in time using the model (referred to as forecast step). Before the forecast step, new biological parameters are redrawn from their respective distributions (as described above in Section 2.1.2) for each model run in the ensemble. Instead of using the same set of parameters for all forecast steps, the resampling of parameters is a strategy that prevents a possibly disadvantageous combination of extreme parameters to prevail over the whole assimilation period and has been used previously (Mattern et al., 2010). The sequence of EnKF analysis and model forecast steps is iterated throughout the simulation period. A brief description of the EnKF analysis scheme follows below (for a more detailed presentation of the EnKF implementation see Evensen, 2003, 2006).

2.2.1. EnKF analysis scheme

The EnKF is based on the update equations of the Kalman Filter (KF); its analysis step for a model state $\mathbf{x}$ at a particular measurement time is given as

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left(d - \mathbf{H}x^f\right) \tag{1}$$

where the superscripts a and f represent the analyzed (i.e. updated) and the forecast (i.e. prior to the update) estimates, respectively; $\mathbf{K}$ is the Kalman gain matrix, $\mathbf{d}$ contains the observations, and $\mathbf{H}$ is the measurement operator that maps the model state onto the available observations. The Kalman gain matrix $\mathbf{K}$ is computed as

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \tag{2}$$

where $\mathbf{P}^f$ and $\mathbf{R}$ denote the error covariance matrices for the forecast estimate and the measurements, respectively; the superscript T represents the matrix transpose. It can be seen from the above equations that the analyzed estimate $\mathbf{x}^a$ is a weighted combination of the forecast estimate $\mathbf{x}^f$ and the residual between the observations and the forecast (i.e., $\mathbf{d} - \mathbf{H}\mathbf{x}^f$). By letting $\mathbf{x}^t$ denote the true (unknown) state, the error covariance for the model forecast is defined as

$$\mathbf{P}^f = \left\langle (\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T \right\rangle \tag{3}$$

where <> denotes the expected value.

The analysis scheme in the EnKF follows the original update equations of the KF, except that the Kalman gain matrix is computed from the error covariances estimated by an ensemble of model states. Let us define a matrix containing an ensemble with m members

$$\mathbf{A} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m). \tag{4}$$

The analysis equation for ensemble member $\mathbf{x}_i$ can be written as

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}\left(d_i - \mathbf{H}x_i^f\right) \tag{5}$$

where i runs from 1 to m; $\mathbf{d}_i$ is the ith member of an ensemble of observation vectors. As noted by Burgers et al. (1998), the treatment of observations as random variables is of importance to obtain a consistent formulation of the error covariances after the analysis. Each member of the ensemble of observations is defined as

$$d_i = d + \varepsilon_i^{\,i} \tag{6}$$

where $\boldsymbol{\varepsilon}_i$ is a vector of measurement errors with zero mean and a covariance equal to $\mathbf{R}$.

Within the EnKF methodology, the error covariance for the model forecast $\mathbf{P}^f$ is approximated by using the ensemble mean as the best estimate of the true state and the spread of the ensemble around its mean as the error variance

$$\mathbf{P}^f \simeq \mathbf{P}_e^f = \frac{1}{m-1}\left(\mathbf{A}^f - \overline{\mathbf{A}^f}\right)\left(\mathbf{A}^f - \overline{\mathbf{A}^f}\right)^T. \tag{7}$$

Here the overline denotes an ensemble mean; $\mathbf{P}_e^f$ represents the ensemble error covariance for the model forecast, providing an approximation to $\mathbf{P}^f$. By defining the matrix holding the ensemble of observations as

$$\mathbf{D} = (d_1, d_2, \cdots, d_m) \tag{8}$$

the EnKF analysis scheme can be expressed in terms of ensemble matrices as follows

$$\mathbf{A}^a = \mathbf{A}^f + \mathbf{P}_e^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T + \mathbf{R}\right)^{-1} \left(\mathbf{D} - \mathbf{H}\mathbf{A}^f\right). \quad (9)$$

The EnKF analysis equations formally read as those in the standard KF except for the use of $\mathbf{P}_e^f$ instead of $\mathbf{P}^f$. The error of such approximation to the error covariances will decrease proportional to $1/\sqrt{m}$ as the number of ensemble members m increases (Evensen, 1994).

In summary, the EnKF data assimilation methodology relies on the representation of error statistics by an ensemble of model states. The fundamental approximations inherent in the EnKF are the assumption of Gaussian error statistics at analysis times and the use of a finite ensemble.

2.2.2. Localization of the error covariances

The reliability of the EnKF technique primarily depends on whether the ensemble size is sufficient to provide an adequate representation of the error covariances. Given the computational cost of model integration, the ensemble size is often limited in practical applications, especially in 3-dimensional models with large state spaces. The limitation on ensemble size can cause spurious correlations between greatly distant grid points, which introduce noise in the analysis and may result in a filter divergence. Houtekamer and Mitchell (1998) pointed out that estimates of the background error covariances between greatly distant grid points were often exaggerated when using a small ensemble size. They also noted that a simple approach for avoiding this problem is to exclude remote observations from the analysis of the local grid point that is being analyzed. A method for "covariance localization" was proposed by Houtekamer and Mitchell (2001). The idea is to localize the ensemble error covariances by applying a Schur product (an element by element multiplication) with a distance-dependent correlation function (Gaspari and Cohn, 1999). Thus the error covariances associated with remote observations are removed and the conditioning of error covariance matrices is improved.

Previous studies on the localization of background error covariances (e.g., Hamill et al., 2001; Houtekamer and Mitchell, 2001) have shown that the EnKF analysis can be substantially improved with covariance localization. Based on the formulation introduced by Houtekamer and Mitchell (2001), the EnKF analysis scheme with the incorporation of covariance localization can be expressed as

$$\mathbf{A}^a = \mathbf{A}^f + \left[\boldsymbol{\rho} \circ \left(\mathbf{P}_e^f \mathbf{H}^T\right)\right] \left[\boldsymbol{\rho} \circ \left(\mathbf{H}\mathbf{P}_e^f \mathbf{H}^T\right) + \mathbf{R}\right]^{-1} \left(\mathbf{D} - \mathbf{H}\mathbf{A}^f\right) \quad (10)$$

where $\boldsymbol{\rho}$ denotes a correlation matrix holding correlations of local support; the notation $\boldsymbol{\rho} \circ \mathbf{B}$ represents the Schur product of the correlation matrix $\boldsymbol{\rho}$ and an error covariance matrix $\mathbf{B}$. Here $\boldsymbol{\rho}$ is determined by using a fifth-order piecewise rational function, as given by Gaspari and Cohn (1999). By defining l as influence radius and e as Euclidean distance between an analyzed grid point and an observation location, the correlation ψ between a grid point and an observation, i.e., an element in $\boldsymbol{\rho}$, is calculated as

$$\psi(l,e) = \begin{cases} 1 - \frac{1}{4}\left(\frac{e}{l}\right)^5 + \frac{1}{2}\left(\frac{e}{l}\right)^4 + \frac{5}{8}\left(\frac{e}{l}\right)^3 - \frac{5}{3}\left(\frac{e}{l}\right)^2, & 0 \le e \le l; \\ \frac{1}{12}\left(\frac{e}{l}\right)^5 - \frac{1}{2}\left(\frac{e}{l}\right)^4 + \frac{5}{8}\left(\frac{e}{l}\right)^3 + \frac{5}{3}\left(\frac{e}{l}\right)^2 - 5\left(\frac{e}{l}\right) + 4 - \frac{2}{3}\left(\frac{e}{l}\right)^{-1}, & l < e \le 2l; \\ 0, & e > 2l. \end{cases} \quad (11)$$

The entries of the matrix $\boldsymbol{\rho}$ are very similar to the values of a Gaussian function (Gaspari and Cohn, 1999). This allows one to retain short-range correlations in the error covariance matrices and removes the spurious long-range correlations. More specifically, $\boldsymbol{\rho}$ is a distance-dependent function, which varies from 1 at the observation location to 0 at the distance greater than twice of the influence radius l.

It is important to note that the filter algorithm shown in Eq. (10) is an approximation to that in Eq. (9), and that due to the localization only observations located within a specified distance (defined by the influence radius) from an analyzed grid point will contribute to the analysis in this grid point.

2.3. Satellite ocean data: chlorophyll and particulate organic carbon (POC)

Chlorophyll data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and the Moderate-resolution Imaging Spectroradiometer (MODIS) are used for the data assimilation experiment described in Section 3. These satellite data were provided by the Northeast Fisheries Science Center (Ecosystem Assessment Program) which obtained SeaWiFS and MODIS-Aqua data from the NASA Ocean Biology Processing Group at http://oceancolor.gsfc.nasa.gov/ (Kimberly Hyde, personal communication, 2010). The ocean color scenes were processed using SeaDAS version 5.2 standard processing and mapped at 1.0 km pixel resolution using a Lambert conic conformal map projection. SeaWiFS and MODIS chlorophyll-*a* data were averaged to create daily images. The daily chlorophyll fields were then interpolated onto the model grid and assimilated into the three-dimensional biological model daily at model noon. It should be noted that often data points in daily fields are missing due to clouds and inter-orbit gaps, leading to irregular spatial and temporal sampling.

Satellite POC data from the MODIS sensor were also collected and processed as described above. The daily POC fields were remapped to the model grid and subsequently used as an independent validation dataset in order to evaluate the predictive skill of the chlorophyll assimilation and its influence on the unassimilated variables.

3. Experiment design

3.1. Ensemble size

A data assimilation experiment was performed for a model simulation from January 1 to December 31 of 2006, using 20 ensemble members. This ensemble size was chosen based on initial tests with 20, 40, and 80 ensemble members. These tests indicated that with an ensemble size of 20 adequate estimates of statistical properties (i.e. ensemble mean and covariances) were produced with little improvement for larger ensemble sizes. Therefore, considering the balance between computational requirements and accuracy, an ensemble size of 20 was chosen for the data assimilation experiment presented here.

3.2. Generation of the initial ensemble

All ensemble members start from identical initial conditions of the three-dimensional ocean state and share the same atmospheric forcing and boundary conditions. While absorption by chlorophyll modifies the water column profile of radiation available for photosynthesis in the biological model, for simplicity no corresponding adjustment is made to the absorption of shortwave radiation that internally heats the water column. Consequently there is no feedback of the biological state on ocean physics, and the ensemble members are identical in terms of their physical circulation. Perturbing key biological parameters to which the biological model was found to be sensitive (see Table 1) generated the ensembles of the biological model state, as described in Section 2.1.2. The ensemble was spun up for 10 days without any data assimilation; this allowed the model dynamics to develop in response to the stochastic parameters. The LEnKF analysis scheme was applied for the first time on January 11, 2006 and repeated daily until December 31, 2006.

3.3. Observation errors, influence radius and inflation factor

Observation errors need to be specified in order to calculate the measurement error covariance **R**, but are generally poorly known. Here, an error estimate of 35%, which was originally the target error of the SeaWiFS project (Hooker et al., 1992), was used for the satellite chlorophyll observations, which were also assumed to be spatially independent. Thus **R** turns into a diagonal matrix and its construction is straightforward. Assuming 35% for the total errors seems reasonable, since it results in larger absolute errors in areas where high chlorophyll concentrations are observed, such as the estuaries and inner shelf of the MAB. This is consistent with the recognition that satellite chlorophyll observations are less accurate close to the coasts. Although a better representation of the observation errors would be desirable for a data assimilation system, the current assumption is sensible and appears to be of sufficient validity for the data assimilation experiment presented here. The influence radius for the covariance localization was set to 100 km. Additionally, an inflation factor as introduced by Anderson and Anderson (1999) was used to inflate the ensemble around its mean in order to account for underestimates of the variance due to the small ensemble size. The inflation factor was set to 1.01; that is, the forecast ensemble $\mathbf{A}^f$ in Eq. (9) is modified according to

$$\mathbf{A}^f = 1.01\left(\mathbf{A}^f - \overline{\mathbf{A}^f}\right) + \overline{\mathbf{A}^f}. \tag{12}$$

3.4. Data-assimilative and deterministic model runs

A simulation with the LEnKF was performed on an ensemble of 20 model runs as described above, and is subsequently referred to as the "data-assimilative run". Chlorophyll observations were assimilated at a daily interval, and chlorophyll, phytoplankton, zooplankton and nitrate were updated at all vertical levels during the analysis steps, based on their covariance with surface chlorophyll. Ammonium as well as small and large detrital nitrogen was not included in the LEnKF updates in favor of increased computational efficiency. Including these variables would have almost doubled the size of $\mathbf{P}_e^f$, the ensemble error covariance matrix (see Eq. (7)), which is involved in the multiplication in the analysis step (see Eq. (9)). As mentioned earlier, during model integration these components will react on changes induced by data assimilation according to the model dynamics.

It should be mentioned that the data assimilation was performed on the actual chlorophyll concentrations, thus the assimilation is suboptimal because it violates the optimality-assumptions of the Kalman Filter (KF). The errors in chlorophyll are approximately log-normally distributed, while the analysis equations of the KF assume a normal distribution. In addition, after each analysis step possible negative concentrations that were produced by the LEnKF updates were simply set to zero; such treatment of negative concentrations was performed only on 0.3% of the total grid cells throughout the assimilation period.

In order to assess the effectiveness of the assimilation scheme and its impacts on ecosystem evolution, the data-assimilative run is compared to deterministic model runs without any data assimilation. Specifically, a deterministic run was performed for the whole time period and is referred to as the "free run". Additionally, 1-month deterministic runs were performed starting from the optimal model states (i.e. the analyzed ensemble mean of the data-assimilative run) at the beginning of each month and are referred to as "monthly runs".

4. Statistical metrics of model/data fit

Three statistical metrics were used to quantify model/data consistency: model bias, root mean square error (RMSE) and model efficiency (ME), all using spatial averaging only (see Lehmann et al., 2009). The analyzed ensemble mean is used for the data-assimilative run since it is regarded as the best guess estimate of the true state.

The model bias quantifies the mean deviations between model results (M) and observations (O):

$$Bias(t) = \frac{1}{n}\sum_{k=1}^{n}[M(k,t) - O(k,t)] \tag{13}$$

where t and k are the temporal and spatial indices, respectively; n denotes the number of model/data pairs. The bias is positive when the model overestimates the observations (considering the whole

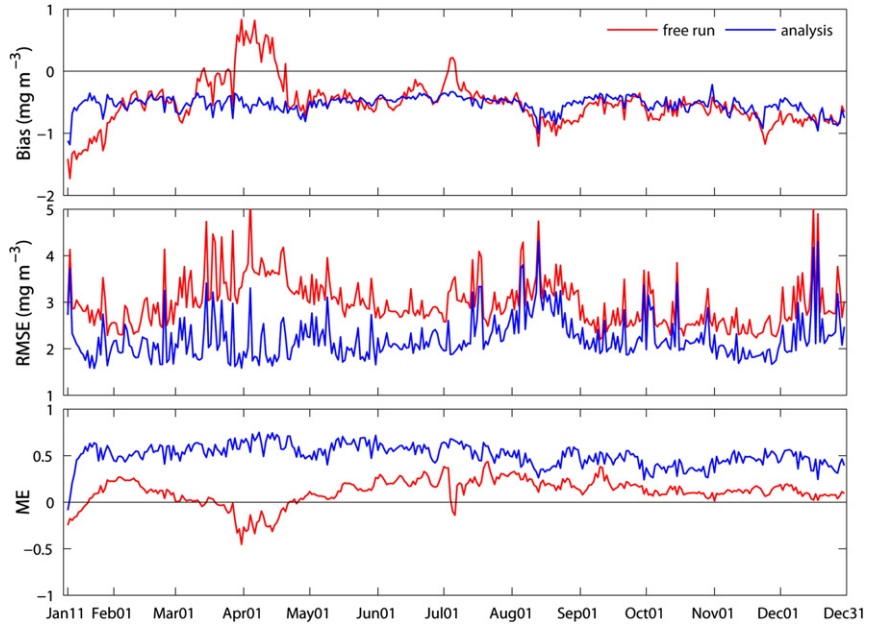

Fig. 2. Model bias, root mean square error (RMSE) and model efficiency (ME) between the observed and simulated surface chlorophyll for the data-assimilative (using the analyzed ensemble mean) and the free (without any data assimilation) runs from January to December of 2006.

Table 2
RMSE and correlation coefficients between the seasonal means of observed and simulated variables for the free and the data-assimilative runs (see Fig. 3): (a) surface chlorophyll and (b) surface POC.

Model run	Winter (Dec–Feb)		Spring (Mar–May)		Summer (Jun–Aug)		Autumn (Sep–Nov)	
	RMSE	Corr	RMSE	Corr	RMSE	Corr	RMSE	Corr
(a) Chl (mg m^{-3})								
Free run	0.29	0.79	0.36	0.53	0.31	0.79	0.35	0.69
Ens. mean	0.17	0.93	0.14	0.93	0.15	0.94	0.15	0.95
(b) POC (mg l^{-1})								
Free run	0.22	0.86	0.24	0.67	0.28	0.76	0.23	0.74
Ens. mean	0.19	0.92	0.19	0.78	0.21	0.84	0.12	0.91

domain), while a negative bias reflects underestimation of the observations.

The RMSE quantifies the deviations between model results and observations in a least-squares sense:

$$RMSE(t) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}[M(k,t)-O(k,t)]^2} \quad (14)$$

and is always greater than or equal to zero. The smaller the RMSE the better the model/data fit.

ME measures the deviations between model and observations relative to the variability in the observations:

$$ME(t) = 1-\sum_{k=1}^{n}[M(k,t)-O(k,t)]^2 \Big/ \sum_{k=1}^{n}\left[O(k,t)-\frac{1}{n}\sum_{k=1}^{n}O(k,t)\right]^2, \quad (15)$$

and is always less than or equal to one, with ME = 1 indicating a perfect model prediction. Positive values of ME suggest that the model is a better predictor than the observational climatology, while negative values of ME indicate that the observational climatology is a better predictor than the model.

5. Results and discussion

Results of the model experiments are presented and discussed in this section. First, the overall performance of the data assimilation system is investigated, i.e., the impact of the LEnKF scheme on the observed variable (surface chlorophyll) and on the unobserved variables. This is done 1) by comparing surface chlorophyll from the free run (without any assimilation) and from the assimilative run to the chlorophyll observations, 2) by comparing simulated particulate matter with independent satellite POC observations, and 3) by illustrating the temporal evolution of the ensemble distributions. In addition, predictive skill of the assimilative model is analyzed by comparing surface chlorophyll between the monthly runs and the free run. Finally the influence of assimilating surface chlorophyll observations on chlorophyll at depth and on other model components (specifically nitrate, phytoplankton biomass and primary production) is assessed.

5.1. Performance of the data assimilation system

5.1.1. Comparison with the assimilated chlorophyll data

Model bias, RMSE and ME (see Eqs. (13)–(15)) were calculated for surface chlorophyll for the free and assimilative runs to quantitatively evaluate the effect of the chlorophyll assimilation (Fig. 2). As expected, the LEnKF scheme drives the model state closer to the observations, yielding smaller RMSE and larger ME values for the analysis relative to the free run. However, the model bias is not improved consistently (during March the bias is smaller for the free run), which is a consequence of the spatial averaging involved in the bias calculation.

RMSE and correlations of the seasonal mean surface chlorophyll between the observations and the free and the assimilative runs are given in Table 2. It is clear that the free run provides fairly good

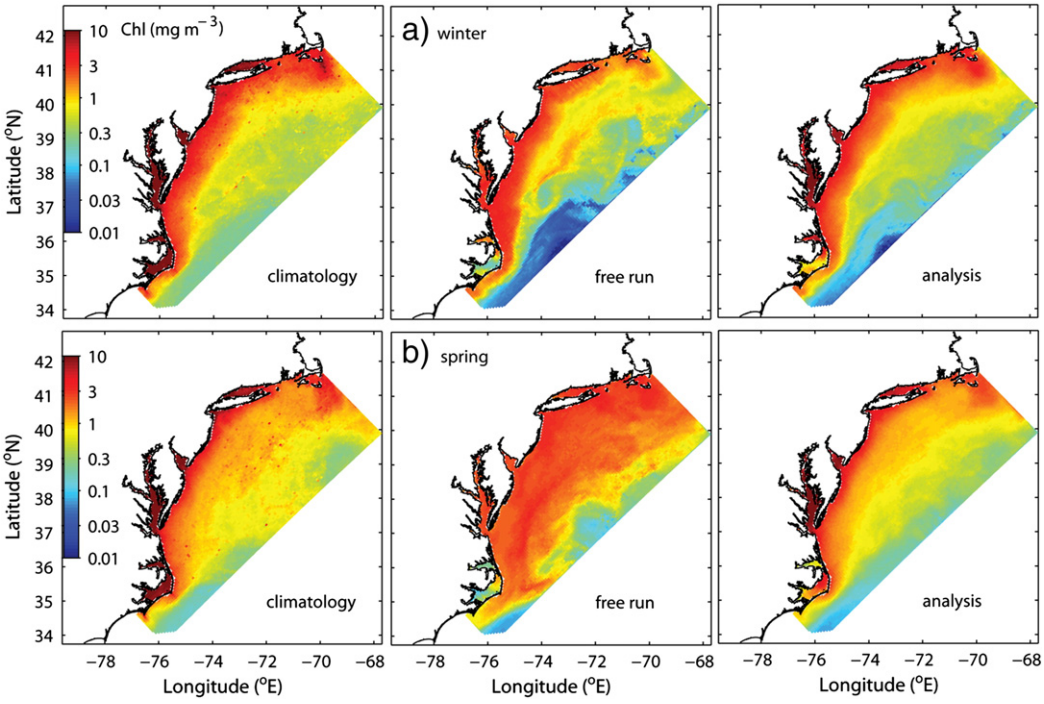

Fig. 3. Seasonal mean distribution of surface chlorophyll for the observations (denoted "climatology"), the free run and the assimilative run in: (a) winter (December–February); and (b) spring (March–May). See Table 2 for quantitative measures of agreement (RMSE and correlation) between these fields.

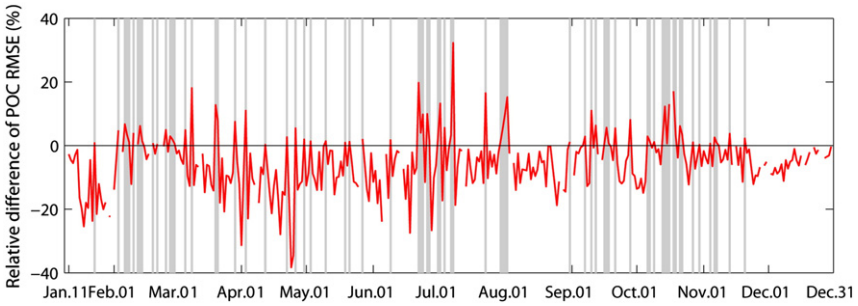

Fig. 4. Relative difference between the POC RMSE from the data-assimilative run (using the analyzed ensemble mean) and from the free run from January to December of 2006. The light gray areas mark periods when the RMSE of the data-assimilative run is larger than that of the free run.

estimates of the surface chlorophyll during the winter, summer and autumn, with the best model performance observed in summer; compared to the free run, the assimilation provides substantial improvements to the chlorophyll estimates in all four seasons. Globally, the model captures the observed inshore–offshore gradient in surface chlorophyll (see Fig. 3 for the winter and spring), with high concentrations in the nearshore areas of the MAB and decreasing toward the offshore. The model performs less well in spring (see Fig. 3b) compared to the other seasons: in spring, the free run overestimates surface chlorophyll on the outer shelf and the slope water; in the data-assimilative run, the model/data consistency is notably improved, as shown in Table 2 (also indicated by RMSE and ME in Fig. 2).

It should be noted that the results presented here are a direct consequence of the specified observation errors, since the strength of the state update by the LEnKF scheme depends to large degree on the uncertainty in the observations relative to that in the model state. The smaller the observation error the closer the updated model state will resemble the observations.

5.1.2. Predictive skill of the assimilation system

The chlorophyll comparisons in the previous section show that the assimilation system works as expected. Here it is tested whether the assimilation system improves the representation of independent observations and whether predictive skill is improved. To test the former, comparisons are performed with satellite POC observations and with the deterministic monthly runs (as described in Subsection 3.4). Specifically, the model-predicted particulate organic nitrogen was calculated by summing phytoplankton, zooplankton and the two detrital nitrogen components, and was then used to estimate POC assuming the Redfield ratio of 106C:16N; the estimated POC fields from the free and assimilative runs are then compared to the POC observations. In addition, it would have been desirable to validate the assimilation system against independent in situ observations. However, this was not possible due to the lack of such observations. In order to test whether predictive skill is improved, surface chlorophyll from the monthly runs is compared with satellite data that was not yet used for assimilation.

Spatially averaged RMSEs were calculated for the observed and simulated surface POC for the data-assimilative and free runs and their relative differences are shown in Fig. 4. Most of the time (but not always) the POC estimates are improved by the chlorophyll assimilation relative to the free run (indicated by negative RMSE differences in Fig. 4); the assimilation improves ~78% of the POC estimates over time. The improvement is well visible in spring, when large corrections to the surface chlorophyll are observed (see RMSE in Fig. 2).

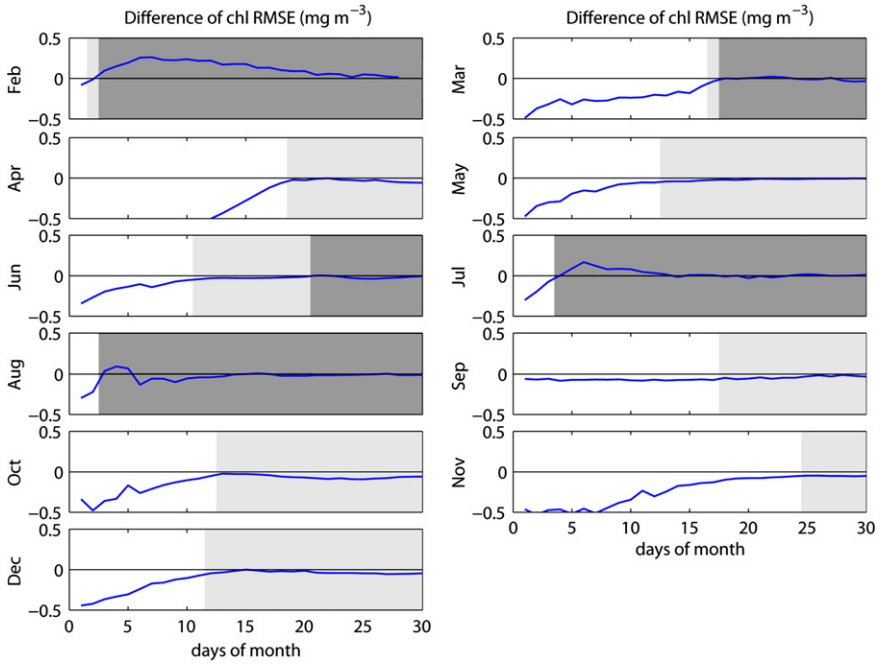

Fig. 5. Chlorophyll RMSE of the monthly runs minus that of the free run for February to December of 2006. The white portion of each panel indicates for how long the monthly run, which started from optimal initial conditions, has a smaller RMSE than the free run by at least −0.05 [mg m^{-3}]. The light and dark gray regions mark periods when the RMSE differences for the first time become smaller than or equal to −0.05 and 0, respectively.

However, alterations of the model state due to the LEnKF updates also lead to deterioration of the POC estimates on some occasions (indicated by positive RMSE differences and gray shading in Fig. 4). This may be partly due to the fact that the two detrital nitrogen components are not directly updated by the assimilation, and therefore no systematic improvement can be expected for these two components as well as the related POC. It is also possible that estimates of phytoplankton or zooplankton deteriorate during the updates. RMSE and correlations of the seasonal mean surface POC between the observations and the two model runs are displayed in Table 2 as well. It is shown that the assimilation improves the POC estimates over all four seasons compared to the free run; corrections to the surface POC are a bit weaker than corrections to the surface chlorophyll (see Table 2), which is expected given that no POC data was assimilated.

Fig. 5 shows the differences of spatially averaged RMSEs of surface chlorophyll between the monthly and the free runs. From these plots it is evident that the LEnKF updates have a positive impact on the predictive skill with smaller RMSEs in the monthly runs for up to 25 days after initialization with an optimized state. The improvement in the predictive skill is time-dependent; it is most pronounced and persistent in March, April and November where improvements are apparent for about two or three weeks after initialization. In some months like February and July, however, the influences of the assimilation on the predictive skill are small; occasionally the simulation was even dragged away from the observations. As noted earlier, state variables other than chlorophyll including those directly updated by the assimilation and those affected only indirectly could improve or deteriorate the model state in the assimilative run. As the impact of the assimilation registered in the updated variables gradually diminishes with time, the simulation of the monthly runs reverts back to a state close to that predicted by the free run.

An open question is whether the multivariate EnKF is capable of improving the evolution of unobserved variables that are part of the state vector and updated during assimilation steps, i.e. chlorophyll below the surface mixed layer and phytoplankton, zooplankton and nitrate. In twin experiments it has been shown the EnKF (Eknes and Evensen, 2002) and the Singular Evolutive Extended Kalman (SEEK) filter (Carmillet et al., 2001) produce successful updates for unobserved variables. However, these findings do not necessarily translate to realistic applications where real observations are used. For example, Triantafyllou et al. (2007) and Fontana et al. (2009) showed that assimilation of surface variables does not necessarily improve state variables below the surface mixed layer. The results described here show that the multivariate LEnKF analysis results in a marked improvement in surface chlorophyll compared to the free run, improves the model's predictive skill, and appears to have a positive impact on the unassimilated variables (indicated by improvements in POC estimates). However, it should be noted that the unobserved variables and the predictive skill are not always improved and that an improvement in POC does not necessarily imply an improvement in

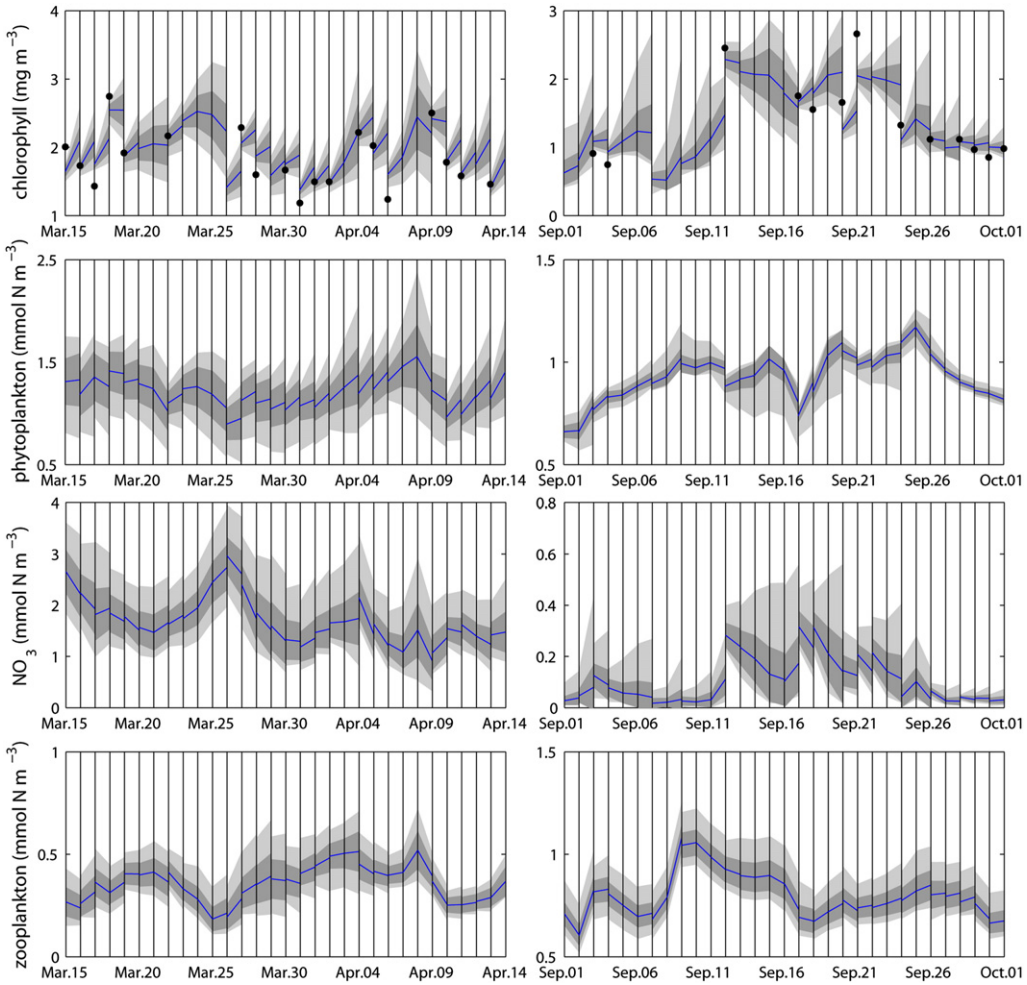

Fig. 6. Ensemble distributions of the concentrations of chlorophyll, phytoplankton, nitrate and zooplankton at the surface for station TS1 (see Fig. 1 for its location) during March–April (left panel) and during September (right panel). Black dots correspond to the observations. The solid blue line indicates the ensemble mean. The light gray area marks the region between the ensemble minimum and maximum, and the dark gray area marks the region showing a standard deviation of the ensemble. Black vertical lines indicate the data assimilation steps. Note that scales change.

its individual components (phytoplankton, zooplankton, small and large detritus).

5.1.3. Time evolution of the ensemble distributions

The temporal changes in the ensemble distributions for the station TS1 (see Fig. 1 for its location) are shown in Fig. 6 for two periods, including one during March–April and one during September. These two selections correspond to a period with relatively high primary production (and relatively large updates of surface chlorophyll by the LEnKF analysis, see Fig. 2) and a period with low primary production (and weak updates of surface chlorophyll), respectively. Note that due to the localization (see Section 2.2.2) the analysis in TS1 is only affected by observations located within the specified influence radius from TS1; whenever no observation is available within this radius (e.g., during March 24–25), the ensemble remains and passes onto the next model forecast step. Sometimes no observations are available at the exact location of TS1, but within the influence radius. From Fig. 6 it can be seen that at the observation times the shifts of the ensemble mean and variance due to the LEnKF updates are evident. The LEnKF analysis is able to force the ensemble closer to the observations (even moving out of its previous range, see e.g., the update on September 12), even if the ensemble spread is small. The small estimated variance suggests that the model error (based on parameter perturbations following a log-normal distribution, see Table 1) may have been chosen too small. The variance of the ensemble decreases during the assimilation step, and then increases again during the forecast step in response to nonlinear model dynamics. The multivariate LEnKF scheme also affects the unassimilated variables, including phytoplankton, nitrate and zooplankton. The variance of the ensemble for these variables generally decreases during the assimilation, although much less so than for chlorophyll. When compared to autumn, the ensemble spread is relatively large during spring, which can primarily be attributed to the relatively intense biological processes in spring (i.e., spring bloom).

As noted earlier, the analyzed estimates of model variables rely to a large degree on the errors in the observations and in the model. Thus, assuming a smaller observation error could lead to larger updates during the analysis. Furthermore, this could also result in a higher degree of convergence of the ensemble, that is, a larger reduction of the ensemble spread. However, it is unclear how such a reduction of the variance would affect the effectiveness of the date assimilation. Further work is required to evaluate the influence of the observation error on the performance of the assimilation system, but this is beyond the scope of the present study.

5.2. Impact of the assimilation on the ecological state and primary production

During LEnKF analysis steps, unobserved variables are updated according to the ensemble-prescribed covariances between the unobserved and observed state variables. In order to assess the impact of the assimilation system on subsurface layers and on other biological variables, differences of chlorophyll, nitrate and phytoplankton biomass along a cross-shelf transect (see Fig. 1 for its location) between the assimilative and the free runs as well as the cumulative direct changes of the biological variables during LEnKF updates are displayed for winter and spring in Fig. 7. This figure allows one to differentiate between direct changes during LEnKF updates and indirect changes that result from the model dynamics acting upon the updated fields. As can be seen in the absolute cumulative changes for chlorophyll in spring, the LEnKF propagates surface information only within the top 50 m. Changes in nitrate and phytoplankton biomass, however, appear to be mostly due to indirect responses of the model dynamics to the updated fields. The assimilation mainly affects the ecological state in the upper 50 m of the water column, which is most evident for chlorophyll in spring (Fig. 7b). Interestingly, in all seasons (see Fig. 7 for the winter and spring, other seasons not shown) phytoplankton biomass tends to increase while chlorophyll tends to decrease in the data-assimilative run compared to the free run. This implies a decrease in the ratio of chlorophyll to phytoplankton biomass. In the data-assimilative run, nitrate tends to decrease due to the uptake by phytoplankton for growth; that is, an increase

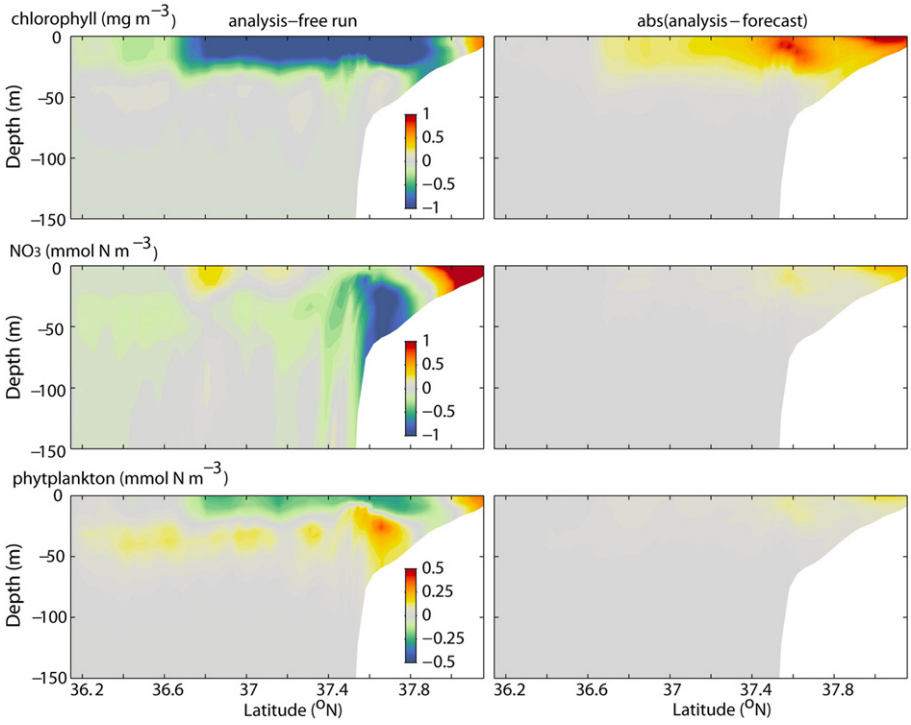

Fig. 7. Differences of chlorophyll, nitrate and phytoplankton biomass for the upper 150 m of the water column along a cross-shelf transect (see Fig. 1 for its location) between the data-assimilative and the free runs (left panel) and the absolute cumulative changes due to the LEnKF updates (right panel) in the winter (a) and spring (b). The cumulative changes were calculated by averaging the daily absolute differences between the analyzed and the forecast ensemble means over each season.

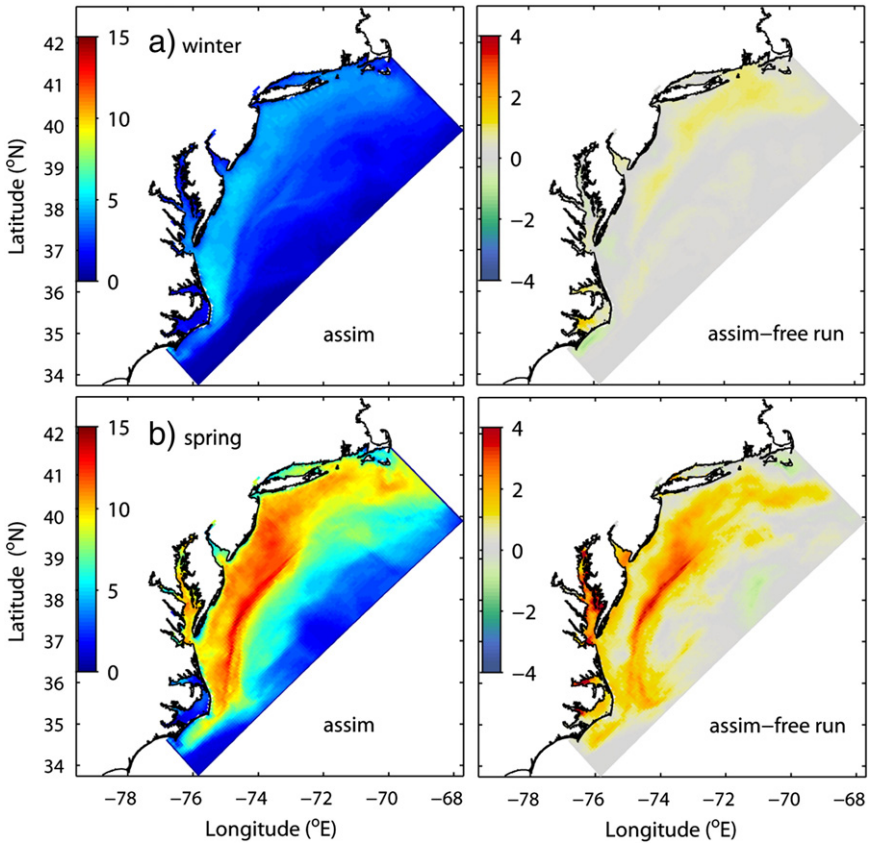

Fig. 8. Mean distribution of vertically integrated primary production (unit in mmol N m^{-2} d^{-1}) for the data-assimilative run and the differences between the data-assimilative and the free runs in the winter (a) and spring (b).

in phytoplankton biomass leads to a decrease in nitrate. Zooplankton generally follows the seasonal pattern of phytoplankton due to the direct link of the two through predation (results not shown).

Note that the data assimilation includes all vertical layers in the LEnKF updates, but, as stated above, changes of the ecological state due to the assimilation are notable only in the upper 50 m of the water column, probably because correlations between the between mixed layer and waters below are small. In order to quantitatively evaluate whether these changes improve the model solution vertical profiles of chlorophyll, nutrients, etc. would be required, but are not available here.

Associated with changes in phytoplankton biomass, chlorophyll and nutrient concentrations are changes in primary production. Primary production depends on the growth rate of phytoplankton and its biomass, where the growth rate is a function of temperature, photosynthetically available radiation and nutrient concentrations. Note

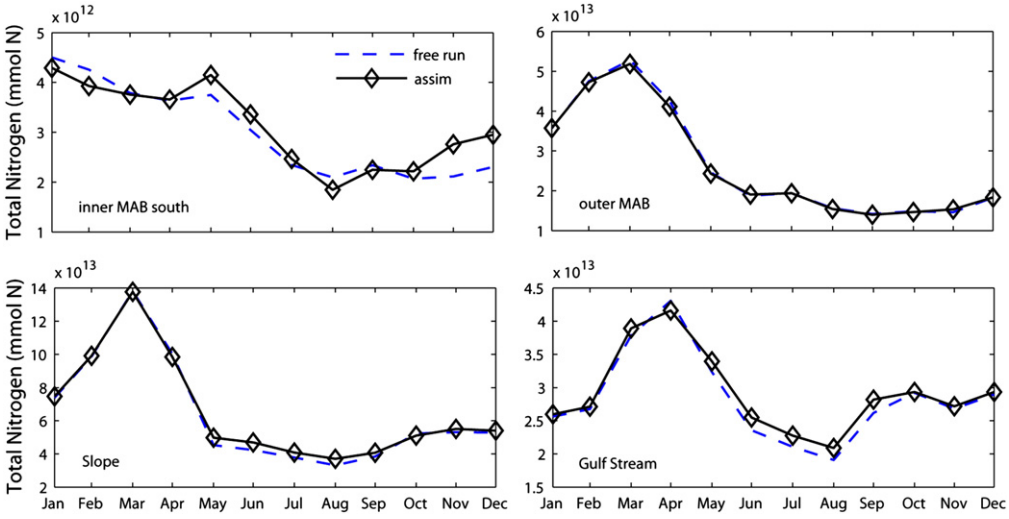

Fig. 9. Comparison of monthly means of total mass of nitrogen in the upper 100 m of the water column in the four MAB subareas (see Fig. 1 for their locations) for the free and the data-assimilative runs. Note that scales change.

that changes in chlorophyll and nutrient concentrations can affect primary production via light limitation (due to the effect of self-shading) and nutrient limitation, respectively, although changes in phytoplankton biomass typically dominate. A comparison of the seasonal means of vertically integrated primary production for the free and the data-assimilative runs is given for winter and spring in Fig. 8. Simulated primary production is highest on the MAB shelf and lowest in the slope waters and Gulf Stream adjacent to the oligotrophic Sargasso Sea. Globally, primary production in the data-assimilative run is higher than in the free run, but has a similar spatial pattern. The data assimilation elevates phytoplankton biomass across the upper water column (as illustrated in Fig. 7), which contributes to the increased primary production evident on the MAB shelf. The increase in primary production due to data assimilation is most pronounced in spring (Fig. 8b).

Another important question is the impact of the data assimilation on the total mass of nitrogen, given that the LEnKF does not ensure mass conservation. In order to assess whether and to what extent nitrogen was added or removed, the total nitrogen content (TN) was calculated for the subareas shown in Fig. 1. TN was calculated by summing phytoplankton, ammonium, nitrate, zooplankton, and small and large detrital nitrogen, and integrating them over each subarea. Fig. 9 presents the monthly mean TN in the upper 100 m of the water column (i.e. the region where changes of the model state due to data assimilation are notable) for the free and the assimilative runs in the southern inner MAB shelf, the outer MAB shelf, the slope waters, and the Gulf Stream. The data assimilation increases TN in all subareas (except on the outer MAB shelf), primarily by adding phytoplankton biomass and nitrate. The increase in TN is more pronounced on the slope waters and the Gulf Stream than on the inner MAB shelf. When compared to the free run, TN in the data-assimilative run increases by 2.5%, 3.0%, and 3.8% on the slope waters, the Gulf Stream, and the inner MAB shelf, respectively, over the assimilation period, while on the outer MAB shelf TN decreases by 0.7%. The TN added in these subareas amounts to about 16, 8.5, 1.2, and -1.9×10^{11} mmol N, respectively, which is about an order of magnitude smaller than the nitrogen discharged by the rivers ($\sim2.7\times10^{13}$ mmol N).

6. Conclusions

Satellite ocean chlorophyll data was assimilated into a three-dimensional biological model of the MAB at a daily time step from January to December of 2006. The data assimilation is performed by using the EnKF method combined with a covariance localization approach, which was implemented to filter out spurious long-range correlations in the EnKF analysis. Experiments were conducted in order to assess the effectiveness of this data assimilation system, its predictive skill and effect on unobserved variables, and its influence on the temporal and spatial evolution of the ecological state and primary production. Results indicate that the data assimilation improved the model behavior at the surface when compared to the free run, for example, model estimates of surface POC, which was not assimilated, improved notably. The assimilation also had a positive impact on the model's predictive skill, in the sense that simulations that started from optimal analysis fields had smaller RMSEs than a non-assimilative free run. The improvement in the predictive skill is time-dependent and varies greatly with the seasons. The data assimilation caused notable changes in the MAB ecosystem in terms of primary production and the ratio of chlorophyll to phytoplankton biomass. As the LEnKF does not ensure mass conservation, changes in total nitrogen were quantified and found to be an order of magnitude smaller than nitrogen inputs form rivers.

Acknowledgments

This work was supported by the ONR MURI grant N00014-06-1-0739 to KF and JW. KF was also supported by NSERC and CFI. We thank Kimberly Hyde for making the satellite observations of ocean chlorophyll and POC available, and two anonymous reviewers for their constructive comments on an earlier version of this manuscript.

References

Allen, J.I., Eknes, M., Evensen, G., 2003. An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea. Ann. Geophys. 21, 399–411.
Anderson, J.L., Anderson, S.L., 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Weather Rev. 127, 2741–2758.
Bagniewski, W., Fennel, K., Perry, M.J., D'Asaro, E.A., 2011. Optimizing models of the North Atlantic spring bloom using physical, chemical and bio-optical observations from a Lagrangian float. Biogeosciences 8, 1291–1307.
Brusdal, K., Brankart, J.M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P.J., Dombrowsky, E., Verron, J., 2003. A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. J. Mar. Syst. 40–41, 253–289.
Burgers, G., Leeuwen, P.J.v., Evensen, G., 1998. Analysis scheme in the Ensemble Kalman Filter. Mon. Weather Rev. 126, 1719–1724.
Carmillet, V., Brankart, J.M., Brasseur, P., Drange, H., Evensen, G., Verron, J., 2001. 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.
Chen, K., He, R., 2010. Numerical investigation of the Middle Atlantic Bight frontal circulation using a high resolution ocean hindcast model. J. Phys. Oceanogr. 40, 949–964.
Druon, J.N., Mannino, A., Signorini, S., McClain, C., Friedrichs, M., Wilkin, J., Fennel, K., 2010. Modeling the dynamics and export of dissolved organic matter in the Northeastern U.S. Continental shelf. Estuarine Coastal Shelf Sci. 88, 488–507.
Eknes, M., Evensen, G., 2002. An Ensemble Kalman filter with a 1-D marine ecosystem model. J. Mar. Syst. 36, 75–100.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162.
Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dyn. 53, 343–367.
Evensen, G., 2006. Data Assimilation: The Ensemble Kalman Filter. Springer, New York.
Evensen, G., 2009. The ensemble Kalman filter for combined state and parameter estimation. Control Syst. Mag., IEEE 29, 82–104.
Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J.B., 2003. Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm. J. Clim. 16, 571–591.
Fennel, K., Wilkin, J., 2009. Quantifying biological carbon export for the northwest North Atlantic continental shelves. Geophys. Res. Lett. 36, L18605.
Fennel, K., Losch, M., Schröter, J., Wenzel, M., 2001. Testing a marine ecosystem model: sensitivity analysis and parameter optimization. J. Mar. Syst. 28, 45–63.
Fennel, K., Wilkin, J., Levin, J., Moisan, J., O'Reilly, J., Haidvogel, D., 2006. Nitrogen cycling in the Middle Atlantic Bight: results from a three-dimensional model and implications for the North Atlantic nitrogen budget. Global Biogeochem. Cycles 20, GB3007. doi:10.1029/2005GB002456.
Fennel, K., Wilkin, J., Previdi, M., Najjar, R., 2008. Denitrification effects on air–sea CO_2 flux in the coastal ocean: simulations for the northwest North Atlantic. Geophys. Res. Lett. 35, L24608. doi:10.1029/2008GL036.
Fontana, C., Grenz, C., Pinazo, C., Marsaleix, P., Diaz, F., 2009. 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.
Friedrichs, M.A.M., Hood, R.R., Wiggert, J.D., 2006. Ecosystem model complexity versus physical forcing: quantification of their relative impact with assimilated Arabian Sea data. Deep-Sea Res. II Top. Stud. Oceanogr. 53, 576–600.
Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. R. Meteorol. Soc. 125, 723–757.
Geider, R., MacIntyre, H., Kana, T., 1996. A dynamic model of photoadaptation in phytoplankton. Limnol. Oceanogr. 41, 1–15.
Haidvogel, D.B., Arango, H., Budgell, W.P., Cornuelle, B.D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W.R., Hermann, A.J., Lanerolle, L., Levin, J., McWilliams, J.C., Miller, A.J., Moore, A.M., Powell, T.M., Shchepetkin, A.F., Sherwood, C.R., Signell, R.P., Warner, J.C., Wilkin, J., 2008. Ocean forecasting in terrain-following coordinates: formulation and skill assessment of the Regional Ocean Modeling System. J. Comput. Phys. 227, 3595–3624.
Hamill, T., Jeffrey Whitaker, J., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an Ensemble Kalman Filter. Mon. Weather Rev., 129, pp. 2776–2790.
Higdon, R.L., de Szoeke, R.A., 1997. Barotropic–baroclinic time splitting for ocean circulation modeling. J. Comput. Phys. 135, 31–53.
Hooker, S.B., Esaias, W.E., Feldman, G.C., Gregg, W.W., McClain, C.R., 1992. An overview of SeaWiFS and ocean color. In: Hooker, S.B., Firestone, E.R. (Eds.), NASA Technical Memorandum 104566. SeaWiFS Technical report Series, vol. 1. NASA, Goddard Space Flight Center, Greenbelt, Maryland.

Houtekamer, P.L., Mitchell, H., 1998. Data assimilation using an Ensemble Kalman Filter technique. Mon. Weather Rev. 126, 796–811.
Houtekamer, P.L., Mitchell, H., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129, 123–137.
Lehmann, M.K., Fennel, K., He, R., 2009. Statistical validation of a 3-D bio-physical model of the western North Atlantic. Biogeosciences 6, 1–14.
Mattern, J.P., 2008. Ensemble-based data assimilation for a physical–biological ocean model near Bermuda. Master's thesis, Universität zu Lübeck.
Mattern, J.P., Dowd, M., Fennel, K., 2010. Sequential data assimilation applied to a physical–biological model for the Bermuda Atlantic time series station. J. Mar. Syst. 79, 144–156.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North American Regional Reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360.
Mukai, A.Y., Westerink, J.J., Luettich, R.A., Mark, D., 2002. Eastcoast 2001, a tidal constituent database for the western North Atlantic, Gulf of Mexico and Caribbean Sea, Tech. Rep. ERDC/CHL TR-02-24. 196 pp.
Natvik, L.J., Evensen, G., 2003. 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.
Nerger, L., Gregg, W.W., 2007. Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. J. Mar. Syst. 68, 237–254.
Ourmieres, 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. J. Mar. Syst. 75, 100–115.
Paulson, C., Simpson, J., 1977. Irradiance measurements in the upper ocean. J. Phys. Oceanogr. 7, 952–956.
Previdi, M., Fennel, K., Wilkin, J., Haidvogel, D.B., 2009. Interannual variability in atmospheric CO_2 uptake on the northeast U.S. continental shelf. J. Geophys. Res. G04003. doi:10.1029/2008JG000881.
Shchepetkin, A., McWilliams, J., 2005. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Model. 9, 347–404.
Shchepetkin, A., McWilliams, J., 2009. Computational Kernel Algorithms for Fine-scale, Multi-process, Long-term Oceanic Simulations. In: Temam, R., Tribbia, J. (Eds.),
Smolarkiewicz, P.K., 1984. A fully multidimensional positive-definite advection transport algorithm with small implicit diffusion. J. Comput. Phys. 54, 325–362.
Spitz, Y.H., Moisan, J.R., Abbott, M.R., Richman, J.G., 1998. Data assimilation and a pelagic ecosystem model: parameterization using time series observations. J. Mar. Syst. 16, 51–68.
Triantafyllou, G., Korres, G., Hoteit, I., Petihakis, G., Banks, A.C., 2007. Assimilation of ocean colour data into a biogeochemical flux model of the eastern Mediterranean Sea. Ocean Sci. 3, 397–410.
Warner, J., Sherwood, C., Arango, H., Signell, R., 2005. Performance of four turbulence closure models implemented using a generic length scale method. Ocean Model. 8, 81–113.
Warner, J., Sherwood, C., Signell, R., Harris, C., Arango, H., 2008. Development of a three-dimensional, regional, coupled wave, current, and sediment-transport model. Comput. Geosci. 34, 1284–1306.