Ocean Sci., 5, 495–510, 2009
www.ocean-sci.net/5/495/2009/

Ocean Science

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

E. Simon and L. Bertino

Nansen Environmental and Remote Sensing Center, Norway

Received: 19 February 2009 – Published in Ocean Sci. Discuss.: 23 March 2009
Revised: 15 June 2009 – Accepted: 28 October 2009 – Published: 3 November 2009

Abstract. We consider the application of the Ensemble Kalman Filter (EnKF) to a coupled ocean ecosystem model (HYCOM-NORWECOM). Such models, especially the ecosystem models, are characterized by strongly non-linear interactions active in ocean blooms and present important difficulties for the use of data assimilation methods based on linear statistical analysis. Besides the non-linearity of the model, one is confronted with the model constraints, the analysis state having to be consistent with the model, especially with respect to the constraints that some of the variables have to be positive. Furthermore the non-Gaussian distributions of the biogeochemical variables break an important assumption of the linear analysis, leading to a loss of optimality of the filter. We present an extension of the EnKF dealing with these difficulties by introducing a non-linear change of variables (anamorphosis function) in order to execute the analysis step in a Gaussian space, namely a space where the distributions of the transformed variables are Gaussian. We present also the initial results of the application of this non-Gaussian extension of the EnKF to the assimilation of simulated chlorophyll surface concentration data in a North Atlantic configuration of the HYCOM-NORWECOM coupled model.

1 Introduction

The context of this work lies in the study and the forecast of the dynamics of the ocean and the evolution of its biology. Important economical stakes involve a better optimization of the management of the natural environment, especially by fisheries. So analysis and short term forecasts of the primary production will be more and more useful to environmental agencies for monitoring algal blooms and possible movement of the fish populations (Johannessen et al., 2007; Allen et al., 2008). For the particular case of Norway, an important issue is the possible movement of fish populations following the sea-ice retreat from the Norwegian Arctic to the Russian Arctic. Such perspectives have led to the developments of numerical ecosystem models during the last decades, as well as their coupling with existing physical ocean models. These couplings are made either on- or off-line, to include vertical 1-D as well as 3-D physical models and express the trade-off between our need in terms of modelling and forecast and the available computing resources.

Nevertheless these models present numerous uncertainties linked to the complexity of the processes that they try to represent and the parameterizations that they introduce. Numerical ocean models are still imperfect and present many errors due to some theoretical approximations, the numerical schemes as well as the resolution that are used. Even though many improvements have been made in the modelling of ocean ecosystems, the models are still too simple in comparison to the complexity of the ocean biology. Finally, the multi-scale interactions between the physics and the biology of the oceans are still poorly understood, leading to errors and uncertainties in the coupling of both numerical models. Numerical ocean ecosystem models alone are not sufficient for understanding and forecasting the real ocean.

Another source of information lies in the observations of the ocean biology. The use of satellites allowed the community to obtain important informations on the surface biology. The observed surface ocean color provides informations on the distribution of the surface chlorophyll for a large area of the oceans, and thus the distribution of the phytoplankton. Satellite observations are also dependent on the atmospheric conditions (for example clouds), leading to loss of data of the ocean surface. Finally, the observations can present important errors, especially for satellite data near the coast. Errors on surface chlorophyll provided from SeaWiFS chlorophyll

Correspondence to: E. Simon
(ehouarn.simon@nersc.no)

data are on average of the order of 30% of the value (Gregg and Casey, 2004), with important variations depending on the area. In the same way, in situ measurements lead to a better understanding of the vertical components of the biological systems in the interior of the ocean. Nevertheless these data have heterogeneous spatial and temporal distributions. The in situ data networks are still quite poor, mainly localized near the coast, and finally are not able to provide information covering the 3-D global ocean.

The interest for data assimilation methods focus on their ability to combine in an optimal way (in a sense to define) the heterogeneous and potentially erroneous information providing by the models and the observations. These methods can be classified in two categories: (1) the probabilistic approach based on the theory of the statistical estimation – the Kalman filter (Kalman, 1960) and its extensions – and (2) the variational approach based on the theory of the optimal control (Sasaki, 1955; Lions, 1968; Le Dimet and Talagrand, 1986; Courtier et al., 1994). These methods can be applied to important classes of problems: the optimization of parameters of the model conditionally to the observations, the sensitivity analysis of the model (to parameters, observations, etc.) and the state estimation. Both are equivalent for linear systems. Data assimilation methods have been successfully applied in the fields of meteorology and physical oceanography and some of them are now used for operational forecast. Nevertheless their application in ecosystem forecasting is quite recent: they have started to be applied to ecosystem models mainly during this last decade. Furthermore, the use of biological observations could be relevant to improve the forecast of the physical model, leading to a real interest for coupled ocean-biogeochemical models.

Data assimilation methods based on the Kalman filter have been successfully applied in numerous cases. In 1-D vertical ocean ecosystem models, real biological in situ data have been assimilated with an Ensemble Kalman Filter (EnKF) (Evensen, 1994, 2003, 2006). Allen et al. (2003) noted that an high frequency assimilation of chlorophyll data (one analysis every two days) was leading to an improvement of the chlorophyll hindcast of the ecosystem model. This study showed that the EnKF could be a suitable method for operational data assimilation systems. Assimilation of chlorophyll and nutrients data with an EnKF in an upwelling influenced estuary (Torres et al., 2006) led to a large improvement of the ecosystem solution (in comparison of the simulation without assimilation). Nevertheless improvements were required, notably on the physical dynamics, in order to achieve a good representation of the ecosystem dynamics.

In 3-D ocean ecosystem models, twin experiments of assimilation of simulated satellite surface chlorophyll data with a SEEK filter (Pham et al., 1998) in a North Atlantic configuration have been done by Carmillet et al. (2001). They demonstrated the ability of a multivariate reduced order sequential updating scheme to correct all the components of an ecosystem model observing a single surface variable only. Furthermore they pointed out the benefits to update the error covariance of the analysis according to the Kalman filter equations rather than using a fixed base of the error subspace. Twin experiments of assimilation of simulated in situ nutrients data with a SEIK filter (Pham, 2001) in the Cretan Sea led to similar conclusions (Triantafyllou et al., 2003). Finally, experiments of Carmillet et al. (2001) suggested that only variables in the upper part of the mixed-layer be corrected and allow for the propagation of the correction by the model to deepest part of the ocean, rather than using the analysis scheme in all the water column, assuming that the reduced-order initial error covariance matrix may damage the covariances on the vertical direction.

Finally for realistic experiments in 3-D ocean ecosystem models, Natvik and Evensen (2003a,b) successfully assimilated SeaWiFS data (surface ocean color) with an EnKF over a short period (2 months) in a North Atlantic configuration: updated states were consistent with data in the surface and, as expected, the analysis steps were reducing the variance fields for different ecosystem components (in the surface and sub-surface). However, long term trends of the ensemble statistics were not investigated, as well as the improvement of the analyzed estimates (non-observed variables). Nerger and Gregg (2007) noted a significant improvement of the surface chlorophyll estimate when assimilating daily SeaWiFS data with a univariate static SEIK filter in a global ocean configuration. Only the surface chlorophyll concentration was directly modified by the assimilation. Furthermore the assimilation used a logarithm transformation of the chlorophyll, according to the assumption of log-normal distribution of the chlorophyll and errors in chlorophyll (Campbell, 1995). Similarly, Gregg (2008) demonstrated the capabilities of a monovariate assimilation of SeaWiFS data with a simple method (Conditional Relaxation Scheme Method) over long periods. For a more important overview of works dealing with the problem of data assimilation in ocean ecosystem model, we refer to Gregg et al. (2009).

The focus of this present paper is the application of the EnKF for state estimation in coupled ocean ecosystem models. Considering that the EnKF performs multivariate analysis and allows an evolution of the covariance errors according to the nonlinear dynamics of the system, it appears to be one of the most advanced data assimilation method able to deal with the assimilation of surface satellite data in ecosystem models. Nevertheless application of data assimilation methods based on linear statistical analysis to such models in an efficient way is a theoretically and practically challenging issue.

On the one hand, the strongly nonlinear behavior of ecosystem models (especially during the period of the spring bloom) raises the question of which stochastic model to be used (Bertino et al., 2003). Nonlinear methods like particle filters seem attractive for such models as they appear to be a variance minimizing schemes for any probability density function. Losa et al. (2004) applied successfully a Sequential

Importance Particle filter (see Doucet et al., 2001) for a combined parameters-state estimation in a 1-D ecosystem model. Nevertheless for realistic configurations, the size of the ensemble required for an efficient application of such a filter is too important to be considered. On the other hand one is also confronted with the model constraints: the analysis state has to be consistent with the model, especially under the constraints of positiveness of some variables. Most variables of ecosystem models are concentrations of a given tracer, and so cannot be negative. Nevertheless this problem is also known for the assimilation in physical ocean models. One thinks for example to the correction of layer thickness while assimilating data in hybrid coordinates model (HYCOM). Several solutions have been suggested to deal with such problems. The one of Thacker (2007) introduces inequality constraints via Lagrange multipliers, leading to a 2-passes 3D-Var. Such approach can also be applied to a Kalman filter. Into the framework of stochastic methods, Lauvernet et al. (2009) developed a truncated Gaussian filter with inequality constraints. But positiveness is only one example of non-Gaussianity among many others. We focus here on a more general approach to non-Gaussianity.

Finally the non-Gaussian distributions of most biogeochemical variables break an important assumption of the linear analysis, leading to a loss of optimality of the EnKF (and other filters). The optimality of the linear statistical analysis is proved under some assumptions, notably an assumption of Gaussianity made on the distribution of the variables (of the model and the observations) and the errors.

In the context of Kalman filtering, a way to deal with these last two difficulties is the introduction of anamorphosis functions in the filter, as suggested by Bertino et al. (2003). They presented an EnKF in which they introduce non-linear changes of variables (anamorphosis function) in order to realize the analysis step in a Gaussian space. Numerical experiments with a 1-D ocean ecosystem model led to promising results. The present paper comes within the continuity of these works and deals with the application of this extension of the EnKF in a more realistic 3-D ocean ecosystem models. Even if our experimental framework appears to be close to the works of Natvik and Evensen (2003a), important differences remain: in this present study, we realized a twin experiment to investigate the influence of the assimilation methodology over longer term trends (one year) both on observed and non-observed variables of the model.

The outline of the paper is as follows. We present the EnKF with Gaussian anamorphosis and a way to build a monovariate anamorphosis function in Sect. 2. We describe our experimental framework in Sect. 3. Results of the methods are discussed in Sect. 4, and we present our conclusions in Sect. 5.

2 The Ensemble Kalman filter with Gaussian anamorphosis

We describe in this section the algorithm of the EnKF with Gaussian anamorphosis suggested by Bertino et al. (2003). The principle is simple and consists of introducing non-linear changes of variables in order to realize the analysis step in a "Gaussian" space, while the forecast step is realized in the physical space.

The main benefit of such algorithm is to alleviate in one pass two important limitations of the application of linear statistical analysis scheme in ecosystem models (described in introduction). The assumption of a Gaussian distribution of the variables appears now to be relevant for the transformed variables during the analysis step. Furthermore there is no "physical" constraint (constraint of positiveness, etc.) on the transformed variables during the analysis, removing post-processing steps that are compulsory when the analysis state vector is not consistent with the physical model.

2.1 Algorithm

The algorithm is based on the skeleton of the EnKF and divides into two steps:

Forecast: the forecast step is a propagation step in the EnKF that uses a Monte-Carlo sampling to approximate the forecast density by N realizations:

$$\forall i = 1:N, \quad \mathbf{x}_n^{f,i} = f_{n-1}(\mathbf{x}_{n-1}^{a,i}, \epsilon_n^{m,i}) \tag{1}$$

with $\mathbf{x}_n$ the state vector at time t_n, f_{n-1} the nonlinear model and ϵ_n^m the model error.

Analysis: the analysis step conditions each forecast member to the new observation $\mathbf{y}_n$ by a linear update. The anamorphosis functions are introduced in this step.

For each variable of the model, at time t_n, we apply a function ψ_n which is a nonlinear bijective function from the physical space to a Gaussian space. We treat each variable separately. In order to simplify the notations, we assume that we have one variable in our model (so one function ψ_n). It reads:

$$\forall i = 1:N, \quad \tilde{\mathbf{x}}_n^{f,i} = \psi_n(\mathbf{x}_n^{f,i}) \tag{2}$$

In practice, it means that we apply the changes of variable for each variable in every point of the discretized domain.

In the same way, we introduce an anamorphosis function χ_n for the observations $\mathbf{y}_n$ at time t_n:

$$\tilde{\mathbf{y}}_n = \chi_n(\mathbf{y}_n). \tag{3}$$

Given the observation operator $\mathbf{H}$ links the physical variables and the observations. We define the observation operator $\tilde{\mathbf{H}}_n$ linking the transformed variables and observations by the formula

$$\tilde{\mathbf{H}}_n = \chi_n \circ \mathbf{H} \circ \psi_n^{-1} \tag{4}$$

where $\circ$ defines the function composition. By assuming that $\tilde{\mathbf{H}}_n$ is linear (this assumption is discussed in the remarks that follow), the linear analysis equation in the Gaussian space reads formally as the classical linear analysis equation:

$$\forall i = 1:N, \quad \tilde{\mathbf{x}}_n^{a,i} = \tilde{\mathbf{x}}_n^{f,i} + \tilde{\mathbf{K}}_n(\tilde{\mathbf{y}}_n - \tilde{\mathbf{H}}_n \tilde{\mathbf{x}}_n^{f,i} + \epsilon_n^{o,i}) \tag{5}$$

with $\tilde{\mathbf{K}}_n$ the classical Kalman gain matrix in the Gaussian space and $\epsilon_n^{o,i}$ the observation errors in the Gaussian space which follow a normal law ($\epsilon_n^{o,i} \sim \mathcal{N}(0, \tilde{\Sigma}^o)$). The transformed Kalman gain matrix $\tilde{\mathbf{K}}_n$ is built on the forecast error covariance matrix $\tilde{\mathbf{C}}_n^f$ approximated by the covariance of $(\tilde{\mathbf{x}}_n^{f,i})_{i=1:N}$.

The pull-back to the physical space is realized by using the inverse of the anamorphosis function:

$$\forall i = 1:N, \quad \mathbf{x}_n^{a,i} = \psi_n^{-1}(\tilde{\mathbf{x}}_n^{a,i}) \tag{6}$$

The analyzed mean $\mathbf{x}_n^a$ and the covariance matrix $\mathbf{C}_n^a$ are approximated by the ensemble average and covariance of $(\mathbf{x}_n^{a,i})_{i=1:N}$.

Remarks

1. The construction of relevant anamorphosis functions χ_n and ψ_n is not straightforward. Analytic functions as log or Cox-Box can be used for variables which initially have a "good" distribution, but are not guaranteed to improve the distribution in general. A more general way to build relevant anamorphosis function can be obtained from the empirical marginal distribution. More details about their constructions are given later.

2. The use of nonlinear functions may introduce non linearities on the transformed observation operator $\tilde{\mathbf{H}}$. In some practical cases, a "good" choice of $\mathbf{H}_n$ and χ_n leads to a linear operator. In the case when observed variables are part of the state vector, $\tilde{\mathbf{H}}$ is obviously linear. It can not be guaranteed for general cases. For a nonlinear $\tilde{\mathbf{H}}$, we suggest to use the EnKF analysis scheme for nonlinear measurements suggested by Evensen (2003, 2006).

3. This algorithm based on the use of monovariate anamorphosis functions does not handle multivariate non-Gaussianity of the state vector. Even if each transformed variables follows a Gaussian distribution, their bivariate (and more generally their multivariate) distributions will not be necessarily bi-Gaussian (resp. multi-Gaussian). In practice this property is really difficult to check due to the large size of the vectors. We assume that the improvements of the monovariate distributions will improve the multivariate distribution. More sophisticated transformations should be investigated in the future (see Schölzel and Friedrichs, 2008).

2.2 Construction of a monovariate anamorphosis function

The performances of the extended EnKF described above are strongly dependent on the choice of the anamorphosis functions ψ_n and χ_n. Several strategies can be applied to the construction of functions that improve the Gaussianity of the distribution of the variables. A first idea is to use "classical" analytic function as the logarithmic function or the Cox-box functions.

Rather than using analytic functions that require prior knowledge of the distribution of variables, we construct the anamorphosis functions directly from a sample of variables.The idea is to build the anamorphosis functions from the empirical marginal distributions of the variables. For that we assume that the variables at different locations and on a limited time period are identically distributed conditionally to the past observations and the physics. The algorithm of the construction of a monovariate anamorphosis function (one function per variable) divides into three parts:

1. **Construction of the experimental anamorphosis function based on the empirical marginal distribution**. Such functions and the way to build these are well known in the geostatistical community. A brief description of the algorithm is given in Appendix A. More details can be found in Chilès and Delfiner (1999). The computational costs of this step are negligible in comparison with the costs of forecast steps in the EnKF.

2. **Interpolation of the experimental anamorphosis function**. Classical polynomial interpolations can be used. Nevertheless, high order polynomial interpolations generate oscillations (close to the extrema of the empirical anamorphosis) that need a particular treatment when defining the tails of the monotonic function. We choose linear interpolation instead.

3. **Definition of the tails of the function**. It is an important step due to the fact that one defines the bounds of the physical variables. The definition of the physical bounds is the way to introduce the physical constraints of the model (for example a minimum value equal to zero will correspond to a constraint of positiveness). For the bounds of the Gaussian space, one has to take unrealistic high values of the analysis into account which causes the tails to extend towards infinity.

These three steps of the construction of the anamorphosis function for the chlorophyll-*a* variable are summarized in Fig. 1.

Remarks

1. The anamorphosis function of a Gaussian variable is linear.

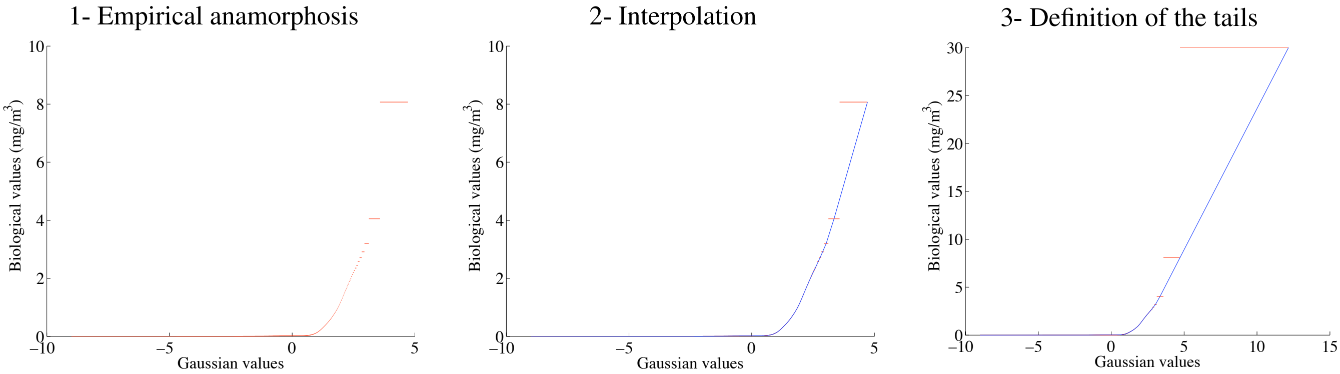


Fig. 1. Surface chlorophyll-a observations: the steps of the construction of a monovariate anamorphosis function.

2. The anamorphosis functions as constructed here are designed for continuous distribution functions and may not improve "pathological" distributions such as Dirac or bimodal.

3. Without Monte-Carlo sampling the introduction of non-linear functions in order to realize the linear analysis estimation in another space can lead to an assimilation bias as follows.

$$E[\psi_n^{-1}(\tilde{\mathbf{x}}_n^a)] \neq \psi_n^{-1}(E[\tilde{\mathbf{x}}_n^a]) \quad (7)$$

   The bias only has an explicit expression in a few particular cases, like the exponential. One general way to avoid the bias is to randomly sample the forecast distribution. In the EnKF, this sampling is realized by using an ensemble during the forecast step. Nevertheless for the other methods such as the Ensemble Optimal Interpolation (EnOI) or the Extended Kalman Filter (EKF), samplings are compulsory.

4. We assume that the variables at different locations in space are identically distributed. In practice, this assumption can not be checked for localized events, leading to a loss of relevance of anamorphosis functions. The spatial refinements of these functions is still an open issue and has to be investigated.

3 Description of the experimental framework

3.1 The coupled ocean ecosystem model

The experiments were performed in a North Atlantic and Arctic configuration of the HYCOM-NORWECOM coupled model. We describe briefly this configuration, which corresponds to the coarse resolution one in Hansen and Samuelsen (2009).

The domain of the model covers the North Atlantic and the Arctic oceans from 30° S. The grid was created using the conformal mapping algorithm outlined in Bentsen et al. (1999).

The physical model used is the HYbrid Coordinate Ocean Model, HYCOM, (Bleck, 2002). The vertical coordinates are isopycnal in the open, stratified ocean, and change to z-level coordinates in the mixed layer and/or unstratified seas. The model uses 23 layers with a minimum thickness of 3 m at the top layer. The model presents 216 × 144 horizontal grid points which corresponds to a horizontal resolution of 50 km. This is sufficient to broadly resolve the large-scale circulation.

The evolution of the ice cover in the North part of the domain (mainly in the Arctic Ocean) is taken into account by an on-line coupling between the physical ocean model and an ice module including a thermodynamic model (Drange and Simonsen, 1996) and a dynamic model (using the elastic-viscous-plastic rheology of Hunke and Dukowicz, 1999). Finally the ERA40 synoptic fields and climatological river runoff (excluding nutrients) are used to force the model.

The ecosystem model is the NORWegian ECOlogical Model system, NORWECOM, (Skogen and Søiland, 1998; Aksnes et al., 1995). This model includes two classes of phytoplanktons (diatoms and flagellates), several classes of nutrients, and includes oxygen, detritus, inorganic suspended particulate matter (ISPM) and yellow substances classes. Nevertheless in our experiments ISPM and yellow substances were not activated. The ecosystem state vector is made up of 7 variables.

This configuration is illustrated in Fig. 2 by a snapshot of surface chlorophyll-a on 22 October 1997.

3.2 Data assimilation experiments

We focus on data assimilation in the ecosystem model. The multivariate assimilation of both physical and biological states is a challenging work and remains an open issue. The state vector corresponds to the ecosystem state vector only, namely seven 3-D variables. Due to the lack of feedback in

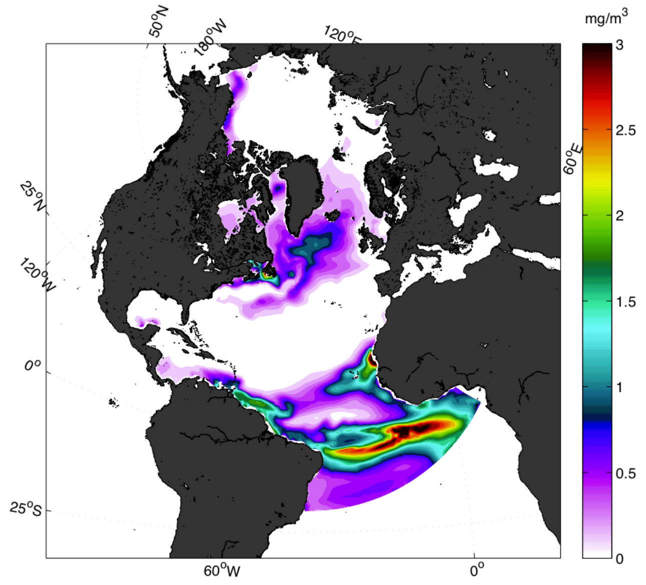


Fig. 2. Arctic and North Atlantic configuration: surface chlorophyll-a concentration (mg/m^3) on 22 October 1997.

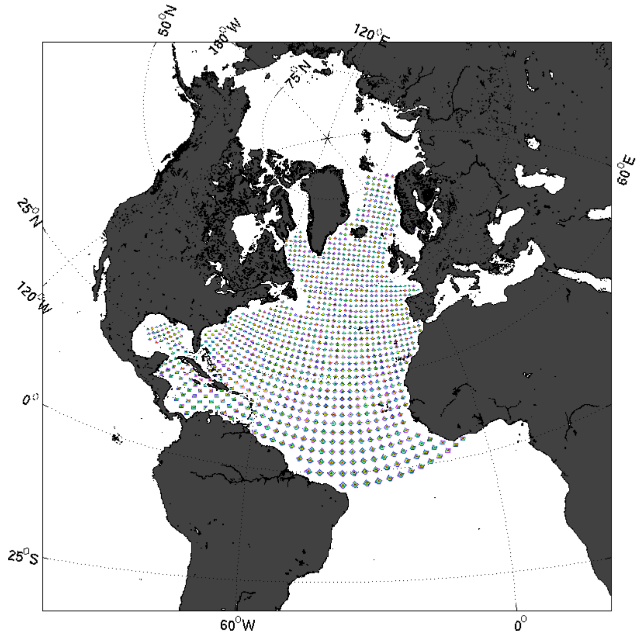


Fig. 3. Surface chlorophyll observations: network of available observations on 31 December 1997.

the coupling from the ecosystem model to the physical one, the assimilation does not correct the ocean physical state.

Our aim is to compare the performances of the extended EnKF with Gaussian anamorphosis to those of a "classical" EnKF. In that way twin experiments have been realized: the true state and the observations are issued from a simulation of the coupled model. The benefits of such a framework is the knowledge of all the components of the solution which leads us to check the impact of the assimilation, in space as well as in time, over all the variables of the model.

Two assimilation systems have been implemented in the same configuration described bellow. The first one called ECO corresponds to the direct application of the EnKF. A post-processing step is added to remove negative values as well as too important values: negative values are increased to zero while unlikely high values are replaced by an arbitrary upper bound (this value corresponds to the biological maximum bound introduced in the construction of the anamorphosis functions, cf. Table 1). The second one called ANA corresponds to the application of the EnKF with Gaussian anamorphosis. No post-processing step is included, as the method does not require any.

The temporal linking of the experiments is as follows. Started from an already spun-up simulation at the date of 10 July 1997, the true state is generated by running the model without perturbation, while the ensemble is generated by running the same model with perturbations (more details about the generation of the ensemble come below). This simulation is issued from the work of Hansen and Samuelsen (2009) and corresponds to the results of a spin-up started in 1958. At this date the spring bloom is at a late stage and the concentration of phytoplankton starts to decrease. Then data assimilation is included as from 24 September 1997. At this date the spring bloom is over and the global concentration of phytoplankton is low and decreases. Assimilation cycles are then performed over one year with a frequency of one analysis step per week.

The synthetic observations are the surface chlorophyll-a obtained by a spatial sampling of the noised true state (Eq. 8) of every third grid index. Furthermore the observations under ice or too close to coasts (the depth of the water column must be greater than 300 m) are not assimilated in order to take into account several constraints of the assimilation of realistic satellite data. Finally the observations present in the southern boundary area (last 15 grid points in the y-direction) are not assimilated either, nor are the observations present in the Arctic ocean (first 50 grid points in the y-direction). It leads to a time evolutive network of observations illustrated in Fig. 3 on 31 December 1997.

The observations are defined as follows

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n^t \times e^{(Z_n - \sigma^2/2)} \tag{8}$$

with $Z_n \sim \mathcal{N}(0, \sigma = 0.3)$. It means that we construct the observations by adding to the true surface chlorophyll-a, which is assumed to have a lognormal distribution, an observation error with a spatial average around 30%, which corresponds to the "usual" error of real satellite data. However, the observation error may locally reach high values (around 75%) as noted for the case of real data. $\frac{\sigma^2}{2}$ is a bias reduction term (observation error).

Table 1. Anamorphosis functions: maximal biological bounds.

Variables	NIT	PHO	SIL	DET	SIS	FLA	DIA	CHLA
$\mathrm{mg\,m^{-3}}$	1000	210	4000	100	200	150	150	30

The strategy for estimating the observation error ϵ^o in the EnKF changes with the assimilating systems. In the ECO system, the observation error at each observation point p is assumed to have a Gaussian distribution with a mean of zero and a standard deviation of 30% of the value of the observation: $\epsilon^o(p) \sim \mathcal{N}(0, \sigma = 0.3 \times \mathbf{y}_n(p))$. It prevents from negative perturbed observations $(\mathbf{y}_n + \epsilon_n^0)$ that are normally truncated to zero, leading to less frequent unrealistic negative values in the analysis ensemble. Even if it may artificially increase the uncertainties of the observations with high value, this approach leads to a significant improvement of the performances of the EnKF comparing to a observation error built on an average value of the observations (not shown). In the ANA system, the observation error in the transformed space has a Gaussian distribution with a mean of zero and a standard deviation of 0.3: $\epsilon^o \sim \mathcal{N}(0, \sigma = 0.3)$. The anamorphosis functions being designed to generate transformed variables with a normal distribution, the observation error in the transformed space is supposed to be around 30% of the transformed observation.

At an observation point, **H** relates linearly the chlorophyll-*a* concentration CHLA to the model diatoms and flagellates concentrations (DIA and FLA) by Eq. (9).

$$\mathrm{CHLA} = \frac{\mathrm{DIA} + \mathrm{FLA}}{11.} \tag{9}$$

The initial ensemble as from 24 September 1997 is the same for both systems (ECO and ANA). It is made up of 100 members obtained by running the model from 10 July 1997 with perturbations of the atmospheric fields in the physical model only (as done in Natvik and Evensen, 2003a). The perturbations induced in the physics then cascade in the ecosystem component of the coupled model. As the state vector is made of the biological component only, the assimilation cannot correct the errors induced by the perturbations in the physical component of the coupled model. Nevertheless the context of twin experiments in a coarse resolution model leads to a low bias in the physical component, the main structure being similar in the ensemble and in the reference simulation. It allows for us to focus only on the improvement of the ecosystem component of the coupled system. For the future realistic framework, a first step will consist to correct the errors in the physical component by assimilating physical data, as already done in the TOPAZ operational forecast and monitoring system (Bertino and Lisæter, 2008), and then the assimilation of chlorophyll-*a* satellite data will be done in the ecosystem component of the coupled model. Direct perturbations of the ecosystem component can also be added. This strategy may appear simplistic, nevertheless the multivariate biophysical assimilation is still an open issue.

The random perturbations are generated by a spectral method (Evensen, 2003) in which the residual error is simulated using a spatial decorrelation radius of 250 km. The decorrelation time-scale is of five days. The standard deviations of the fields perturbed are: $0.03\,\mathrm{N\,m^{-2}}$ for the eastward and northward drag coefficient, $\sqrt{2.5}\,\mathrm{m\,s^{-1}}$ for the wind speed, $\sqrt{0.005}\,\mathrm{W\,m^{-2}}$ for the radiative fluxes and 3° Celsius for the air temperature. These values correspond to the ones use in the TOPAZ operational forecast and monitoring system.

Finally both systems use localization as suggested by Evensen (2003). The radius is constant and equal to 500 km (10 cell-grids in the two horizontal directions) therefore at each point we assimilate between 2 and 10 observations depending on the area. The aim of this work being the comparison of the intrinsic behavior of the two assimilation systems, we have not introduced advanced operational processes as the decrease of the radius close to the coast for example, in order to have a better understanding of the benefits of anamorphosis functions.

3.3 Construction of the monovariate anamorphosis functions

We assume that each variable and the chlorophyll-*a* at different locations in space are identically distributed in a time period of three months centered on the datum of the analysis step. In that way we obtain time evolving anamorphosis functions. The choice of three months is motivated by the time scale of bloom phenomena which is about 4 months. Such a moving window allows for a representation of the differences of distribution at the beginning and the end of the bloom in the construction of the anamorphosis functions.

The experimental anamorphosis functions are computed from weekly output from a four year integration of the model. The anamorphosis function is piecewise linear, using linear interpolation of the experimental anamorphosis function. The middle of steps are used to interpolate the empirical anamorphosis functions, with the exception of the last right step for which the maximal value of the data set is used. The tails of the anamorphosis are defined as follows:

- Biological bounds: the minimum values are equal to zero (constraint of positiveness) and the maximum values are unlikely high values summarized in Table 1.

– Gaussian bounds: the minimum values are equal to -9 (value with a probability around 1×10^{-19}). We do not define maximum values, the right tails extending towards infinity.

Remark

> In case of model bias (which would occur with assimilation of real data), the model-based anamorphosis functions may be impaired by the bias, especially when using a short moving window. For example, the main bloom could be modeled too early or too late by a couple of weeks, which would make high concentrations of plankton too likely or too unlikely at different stages of the bloom. Thus the moving time window should be shorter than the bloom, but not too short by comparison to usual ecosystem model delays. We consider three months as a reasonable compromise.

The interpolated anamorphosis functions (step 2) of chlorophyll-a, diatoms and flagellates (phytoplankton) and silicate (nutrient) are shown in Fig. 4 during three periods of the year: in winter (31 December 1997) when the primary production is low, during the spring bloom (14 May 1998) and in fall (3 September 1998) when the concentration of phytoplankton decreases slowly.

We note that the shape of the anamorphosis functions of the chlorophyll-a and the two phytoplanktons are quite similar (see in Fig. 4). The anamorphosis presents a curvature in the interval $[-1, 1]$ of the Gaussian space, affecting around 65% of the values (the transformed variables have a normal distribution $\mathcal{N}(0,1)$). Had the distribution been a truncated-Gaussian, the anamorphosis would have been a straight line, intersecting the abscissa. Furthermore the impact of the season appears mainly on the localization around zero of the strong non-linearity of the functions, and on the maximum value present in the biological data set. Finally the anamorphosis functions of the silicate variable present many nonlinearities all along the shape of the functions, and particularly near the high values of the biological data set. It is also the case for the other nutrient variables (not shown).

The results of the application of anamorphosis functions on the distribution of the diatoms and the silicates are shown in Fig. 5 during the same three periods of the year previously shown. In this present study, we focus on diatoms which are linked to the chlorophyll-a (observation) by a linear relation and on the silicates which limit the rate of the production of diatoms but not the production of flagellates.

First we note that the time evolving anamorphosis functions provide more Gaussian distributed variables as expected. This is globally true for the other variables of the ecosystem model (not shown). Nevertheless the histogram of the transformed diatoms during the spring bloom allows for the appearance of the superimposition of two Gaussian functions. It can be explained by the bloom in the eastern part of the North Atlantic (mainly off Spain) in the ensemble which is earlier than the blooms present in the data set used for building the anamorphosis functions. So it means that we reach the problem of the bias of anamorphosis functions based on moving windows. A way to deal with this problem would be to include more extreme events in the data set used for the construction of the anamorphosis functions.

4 Data assimilation results

4.1 Observation error

At first we are interested in the evolution with time of the spatial averages of the true observation error and its estimate by the filter in both systems (Fig. 6). For the case of the EnKF with Gaussian anamorphosis (ANA configuration), the spatial average is computed in the transformed space, while this value is computed in the physical space for the true observation error and the plain EnKF (ECO configuration).

First we note that the curve of the spatial average of the true observation error presents large deviations around the specified value (30%). We note also the presence of more important errors in the observation at the beginning of the spring bloom in March–April. These variations of the observation error introduce difficulties for its estimation by the filter. The specification of relevant estimate of the observation error is an important problem reached when dealing with real observations.

For the case of the ECO configuration, the evolution of the spatial average of the observation error estimate is almost constant around 30%, according to the observation error variance specified in the filter. This value corresponds to the average value of the true observation error. However, the presence of variations in the true observation error leads to a succession of under- and overestimate of the observation error in the analysis steps.

Finally we note a continuous overestimation of the observation error in the ANA configuration, exception to few analysis steps during the spring bloom. This is explained by the chlorophyll-a anamorphosis function not being exactly an exponential function. It leads to persistent weaker corrections in the Gaussian space than the ones that could have been obtained with a more relevant estimate and weaker than in the ECO configuration. Furthermore, we note significant variations with time around 35% of the observation error estimate, which seem to follow the low frequency oscillations of the true observation error. We have no explanation for these similar trends and this result may not be observed in future experiments. However, transformed observations with a normal distribution would have led to an almost constant estimate of the observation error around 30% in average (rather 35% in the present experiments). It means that the chlorophyll-a anamorphosis function cannot produce transformed variable with a normal distribution as expected. This

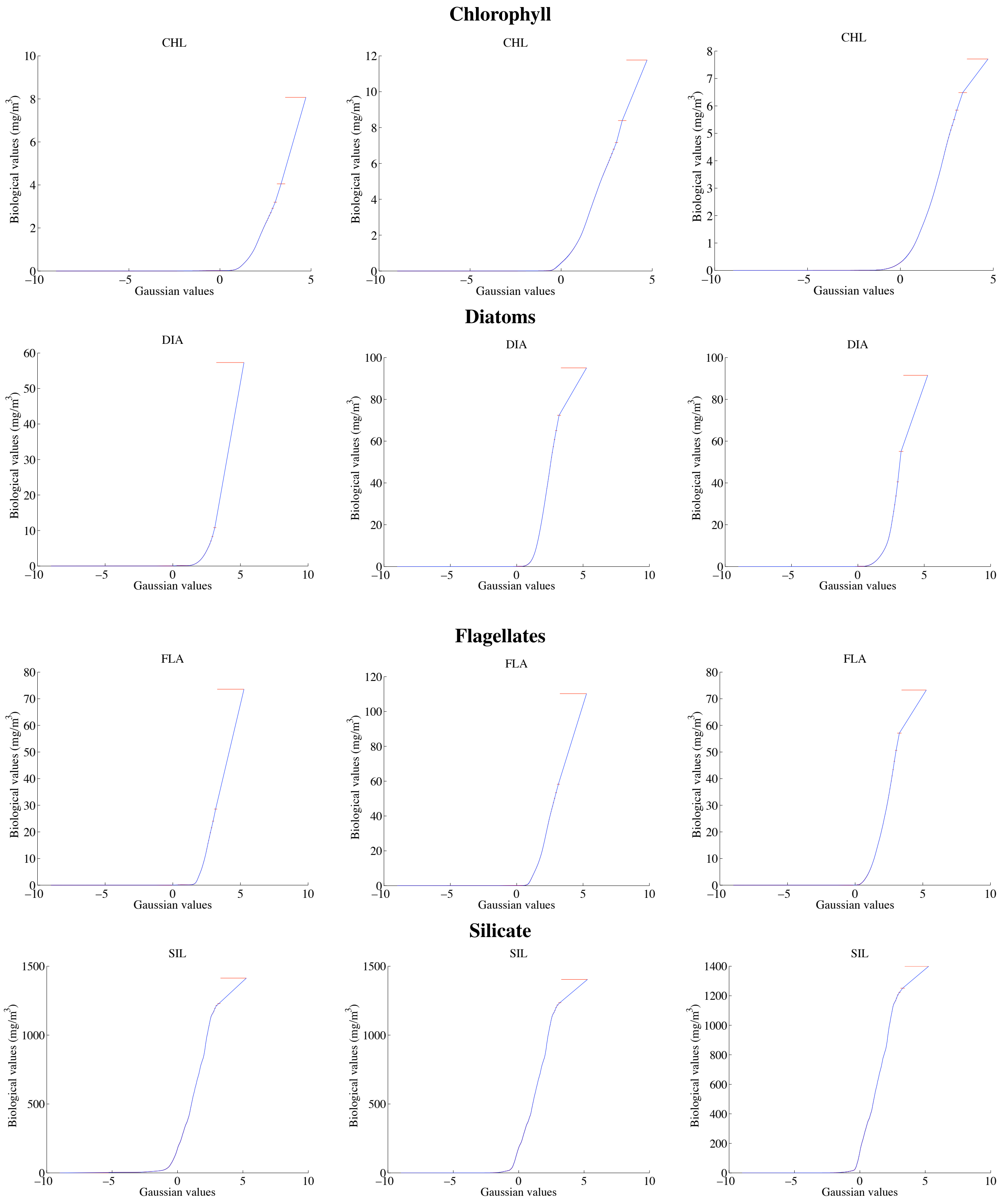


Fig. 4. Interpolated anamorphosis functions. Left: 31 December 1997; center: 14 May 1998; right: 3 September 1998. The right tails are not plotted (same slope that the last segment).

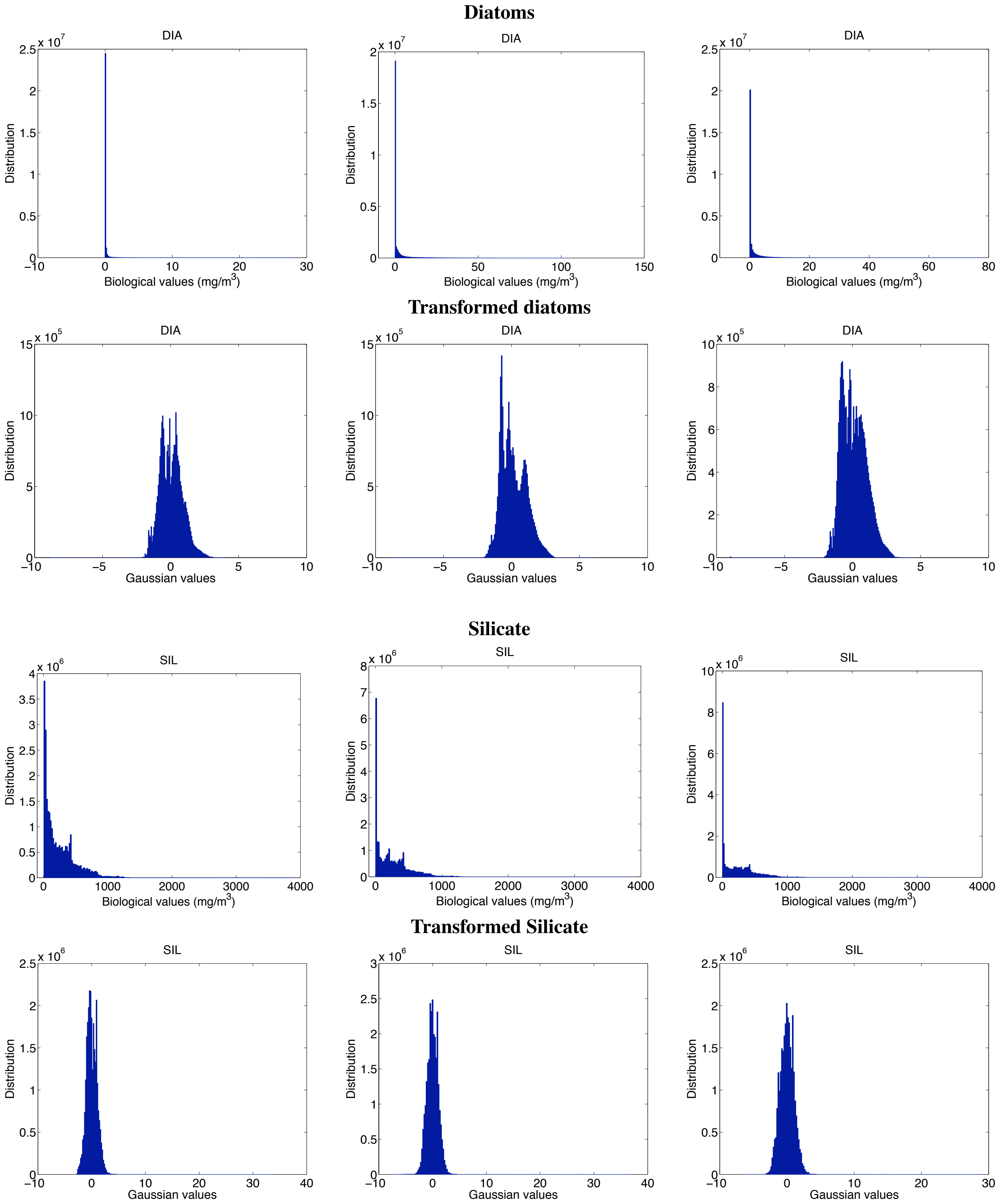


Fig. 5. Distributions of 3-D biological and transformed variables. Left: 31 December 1997; center: 14 May 1998; right: 3 September 1998.

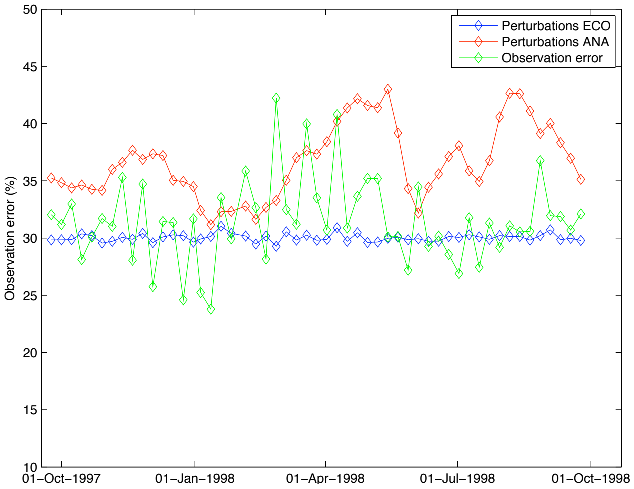


Fig. 6. Observation error: one year evolution of the spatial averages of the true observation error and the estimated observation errors by the filters (%).

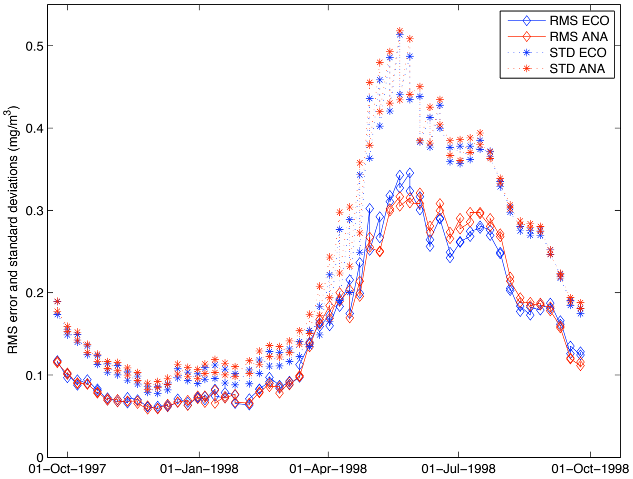


Fig. 7. Surface chlorophyll-a: one year evolution of the RMS error and the standard deviations (mg/m^3).

should improve when including observations in the data set used to build the anamorphosis functions.

4.2 Overall error evolution

We are interested in the evolution in time of the true Root Mean Square error (RMS) and the ensemble standard deviations (STD) of the solution of the two systems. The expression at time t_n of these two quantities is as follows:

$$\begin{aligned} \mathrm{RMS}(t_n) &= \sqrt{\tfrac{1}{\#\Omega}\sum_{\mathbf{k}\in\Omega}(\mathbf{x}^t(t_n,\mathbf{k})-\bar{\mathbf{x}}(t_n,\mathbf{k}))^2} \\ \mathrm{STD}(t_n) &= \sqrt{\tfrac{1}{N-1}\tfrac{1}{\#\Omega}\sum_{\mathbf{k}\in\Omega}\sum_{m=1}^{N}(\mathbf{x}^m(t_n,\mathbf{k})-\bar{\mathbf{x}}(t_n,\mathbf{k}))^2} \end{aligned} \qquad (10)$$

with Ω the domain of computation, $\#\Omega$ the number of grid points of the domain used for the computation of the RMS and STD, N the number of members, $\mathbf{x}^t$ the true state, and $\bar{\mathbf{x}}$ the mean of the ensemble.

Figure 7 represents the evolution of the RMS error and the standard deviations over one year for the surface chlorophyll-a (what we observe). In that case Ω is the top layer of the model. We note that both systems present the same evolution of RMS error and standard deviations, even if slight differences are observed during the period of the spring bloom (April–August). We note also that the standard deviation is higher than the RMS error for both systems, expressing an over-estimation of the error by the filters.

Furthermore we observe three phases in the evolution of the curves. The first one corresponds to the end of the bloom and the winter (October 1997–March 1998). During that phase, the RMS error is low and the assimilation of observations does not significantly improve the solution, indeed may damage it when the observation error locally reaches high values. The second phase corresponds to the spring bloom. The RMS error and the standard deviation increase from March to June. During that period, the analysis steps are efficient and lead to a significant decrease of the RMS error and standard deviations of the solutions. Furthermore, we note that the RMS error in the ANA experiment is slightly lower than in the ECO configuration. In the second part of the bloom (June–August), the RMS error and STD start to decrease. The analysis steps are less efficient and may damage the solution in the ANA configuration, leading to a slightly lower RMS error in the ECO experiment. This is explained by the presence of observations out of the range of the model data set used to build the anamorphosis functions. It may lead to unlikely high values for the transformed observation if the right tail of the anamorphosis function is not defined carefully, leading to locally biased analysis. The addition of more extreme events and observations in the anamorphosis function data set can efficiently remedy for this model bias. Finally the third phase corresponds to the end of the bloom. The RMS error and the standard deviation decrease slowly to reach their initial values. Furthermore the lack of observations in shallow waters leads to some difficulties in correcting the solution in several areas (cf. Sect. 4.5).

Finally the truncation due to the post-processing step in the ECO experiment affects a very few number of state variables (not shown) thanks to the local specification of the observation error as a percentage of the value of the observation: by reducing the frequency of appearance of negative perturbed observations during the cold period comparing to an observation error defined uniformly from an average error value, it prevents the appearance of negative values in the analysis ensemble.

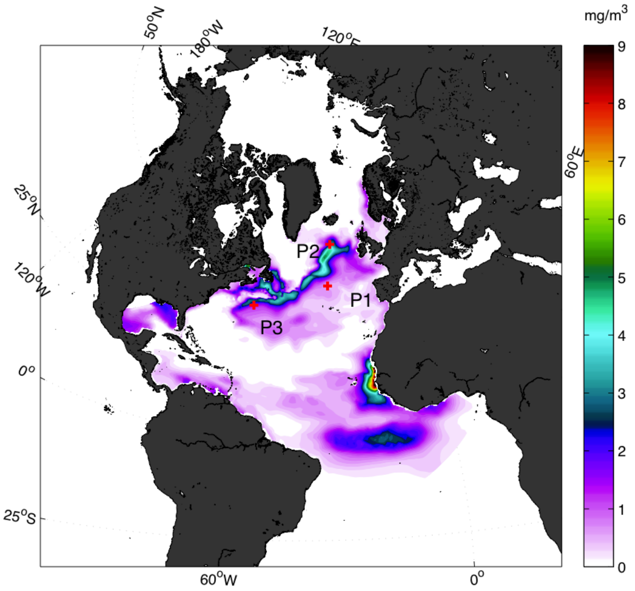


Fig. 8. chlorophyll-a concentration (mg/m^3): the top layer on 23 April 1998. The points P_1, P_2 and P_3 are localized by a red cross.

4.3 Local evolution of the ensemble

We are interested in the evolution with time of the mean and standard deviations of the ensembles and observations as well as the true state at different grid points localized in the vicinity of the Gulf Stream (Fig. 9). Our aim is to study the local effects of the linear analysis on the observed variable for both systems in order to highlight assimilation biases that could have been hidden in the previous diagnostic due to the spatial averaging. This area is characterized by strong dynamics in both components of the coupled model (strong spring bloom in area of the Gulf Stream). The investigated points P_1 and P_2 are localized by red crosses on Fig. 8. Since we are interested in the behavior of the analysis, the several diagnostics are computed in the Gaussian space for the ANA configuration.

First, we note that both assimilating systems are efficient: the mean of the ensemble is very close to the true state despite the presence of observations with significant errors. Nevertheless, some assimilation biases appear. For the case of the ANA configuration, we note an increase of the standard deviation of the ensemble at the beginning of January in both locations. At this time, few outliers with very low values appear in the forecast ensemble (not shown). These values being unlikely when considering the data set used to build the anamorphosis function, this results in the presence of few outliers with high negative values in the transformed forecast ensemble, hence an artificial increase of the transformed forecast error estimate in the filter. This leads to few corrections towards erroneous transformed observations. Spatial refinements of the anamorphosis function have to be investigated to reduce the transfer of local bias from the model to the anamorphosis function and to improve the local distribution of the transformed variables. In the case of the ECO configuration, the observation error defined by a percentage of the value of the observation leads to a decrease (resp. an increase) of the confidence in observations with high values (resp. low values). It can be useful when the observation error increases the value of the observation comparing to the true state, as noted at the point P_2 in July 2008 (Fig. 9). On the other hand, it can induce an underestimation of the error for observations lower than the true state or with low values, leading to too strong corrections towards erroneous observations as noted at the point P_1 in May 2008 (Fig. 9).

4.4 Errors in the sub-surface

In order to explore the multivariate aspect of the data assimilation, we focus on the evolution of the RMS error and the standard deviation, computed on only one grid point (58.8° S, 38.7° E) in the area of the Gulf Stream, for the diatoms and the silicate. This point, called P_3 and localized by a red cross on Fig. 8, is in the 8th layer (waters between 30 m and 38 m) of the model, the deepest one locally before vanishing of the diatoms. As the concentrations of diatoms at this point can change quickly with time, it is a good indicator of the front of structures.

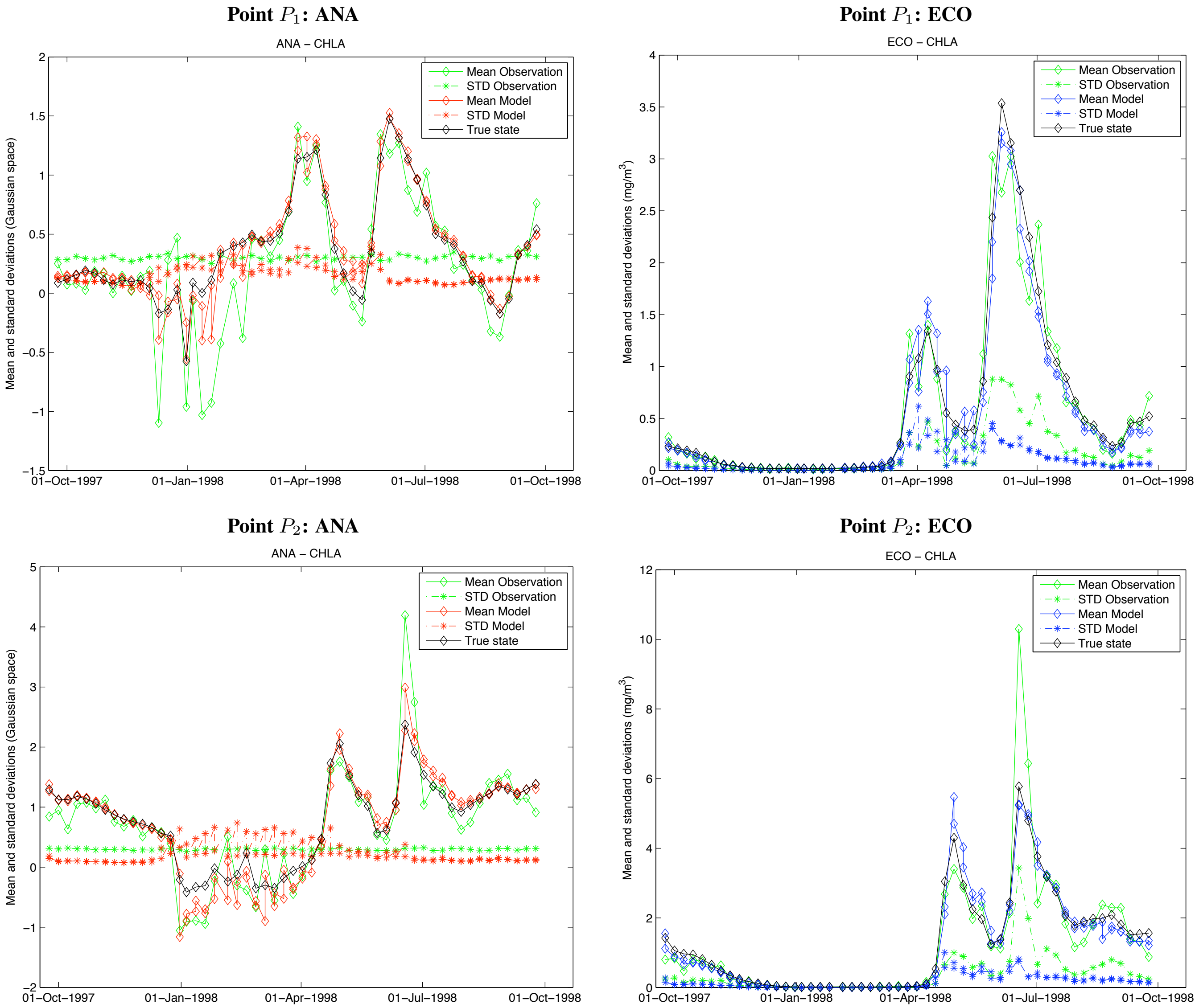
Once again we do not note significant differences between the two systems (not shown). The RMS error and the standard deviations remain low: the RMS error reaches a maximum of $4\,\mathrm{mg\,m}^3$ for the diatoms and $20\,\mathrm{mg\,m}^3$ for the silicate. Furthermore, both assimilating systems overestimate the error.

4.5 Regional distribution of the errors

We examine the spatial localization of the error on the surface chlorophyll-a before, during and after the main bloom. Figures 10, 11 and 12 represent the maps of the surface chlorophyll-a component of $\bar{\mathbf{x}}^a - \mathbf{x}^t$ on 31 December 1997, 14 May 1998 and 3 September 1998. As stated previously, the observations present in the southern boundary area are not assimilated, due to this, important errors remain in this part of the domain. The maps of RMS error focus only on the regions of interest (North Atlantic and Arctic regions).

On 31 December, we note that the error is mainly localized in the south of the domain where the concentration of chlorophyll-a is highest. Slight differences appear in the distribution of the errors. For the ANA configuration, the mean of the analyzed ensemble tends to be higher than the true state while the error is better balanced in the ECO configuration. The observation error being overestimated in the ANA configuration, it leads to weaker corrections by the filter in area of high chlorophyll-a production.

On 14 May, during the spring bloom, we note an increase of the error comparing to winter. The mean solution of the


Point P_1: ANA **Point P_1: ECO**

Point P_2: ANA **Point P_2: ECO**

Fig. 9. Surface chlorophyll-*a*: one year evolution of the mean and the standard deviations of the ensembles, the observation and the true state at the points P_1 and P_2. The variables are represented in the Gaussian space for the ANA configuration.

ensemble is slightly better in the ANA configuration. Nevertheless, the overestimation of the observation error in the transformed space does not allow the EnKF to efficiently reduce the error issued from a too strong spring bloom in the forecast ensemble. In the ECO configuration, the bloom is too weak in the domain from the North American coast to Europa. This negative error is an inherited consequence of the underestimation of the observation error at the beginning of the spring bloom (April–May) that generates important local analysis step in direction of erroneous low observation. Furthermore, the lack of observations on the European North West Shelf leads to important persistent errors in the North Sea (between UK and Norway) for both configurations. This bias is a nonlinear response to the perturbations of atmospheric forcings (likely more resuspension in average for example).

After the spring bloom, on 3 September, we observe errors in a chlorophyll-*a* structure localized south of Greenland for both configurations. However, the solutions present significant differences in this area: the concentration of chlorophyll-*a* is underestimated in the ECO configuration while this one is overestimated in the ANA configuration. These are apparently inherited from the previous biases observed during the spring bloom. We note also significant errors in the North Sea and the Barents Sea where no observations are present.

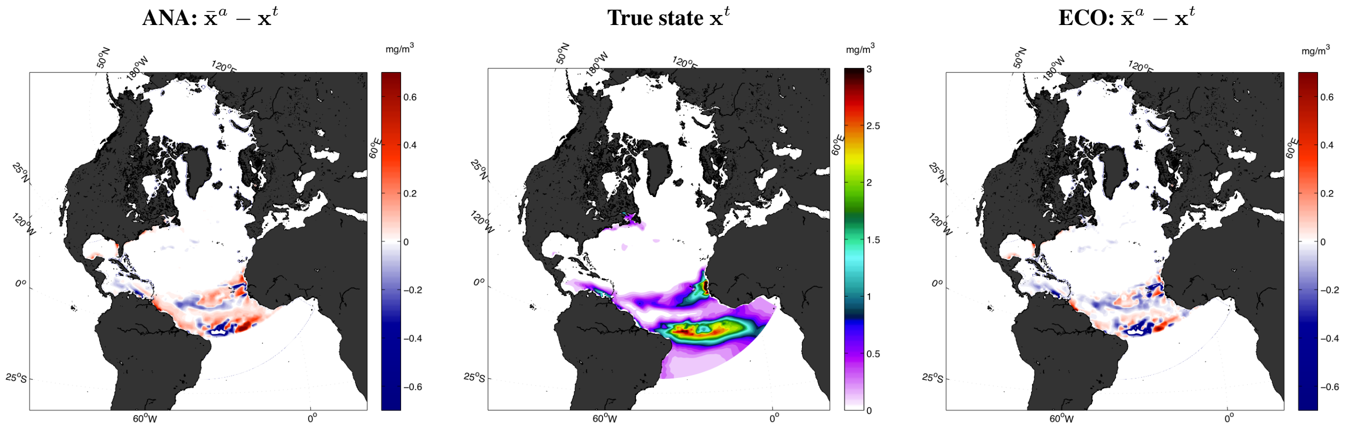


Fig. 10. $\bar{x}^a - x^t$: surface chlorophyll-a component (mg/m^3) on 31 December 1997. Errors in the equatorial Atlantic Ocean are not plotted.

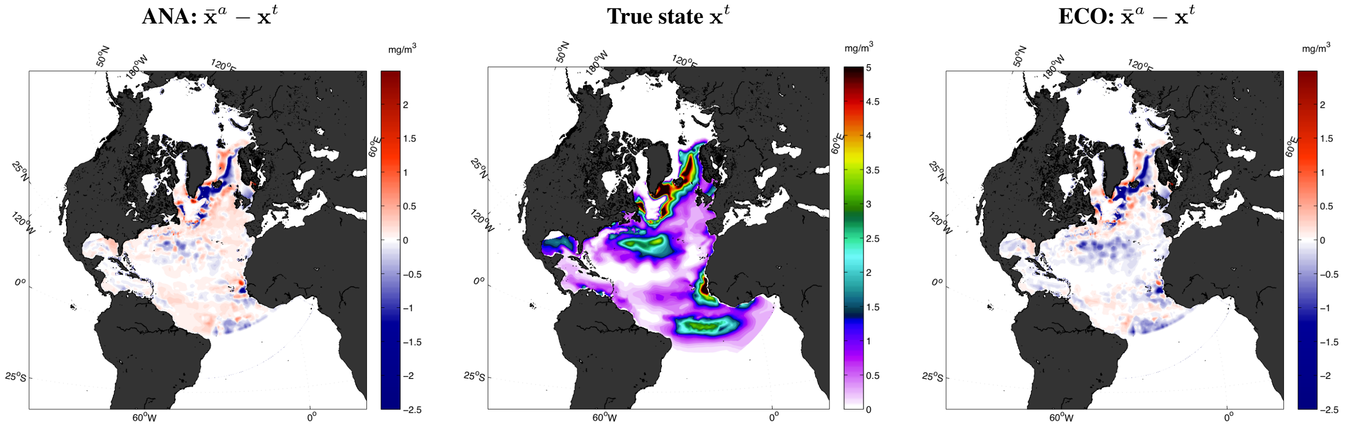


Fig. 11. $\bar{x}^a - x^t$: surface chlorophyll-a component (mg/m^3) on 14 May 1998. Errors in the equatorial Atlantic Ocean are not plotted.

5 Conclusions

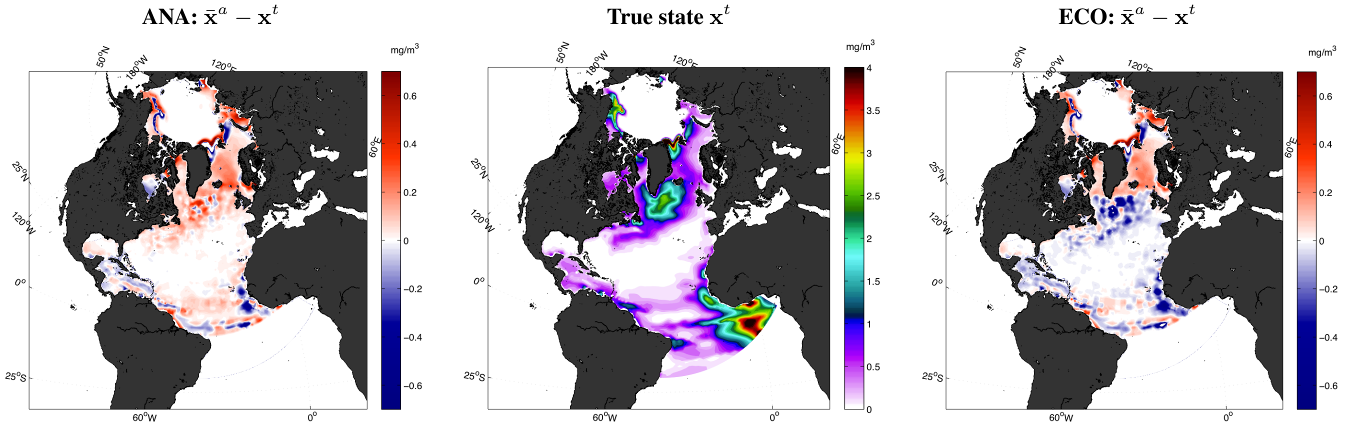
A twin experiment has been conducted with a realistic coupled physical-ecosystem model of the North Atlantic and Arctic Oceans, assimilating simulated surface chlorophyll-a with an EnKF, with and without Gaussian anamorphosis.

The study reveals that applying the plain EnKF with a simple post-processing of negative values or the EnKF with Gaussian anamorphosis leads to similar results. Both systems present low RMS errors as well as an overestimation of the error from the ensemble statistics. However, when considering that the observation error was clearly overestimated in the EnKF with Gaussian anamorphosis (between 5 and 10 percentage points), the anamorphosis seems to have an advantage in efficiency. The advantage should become clearer when using more accurate observations, would they become available in the future.

The introduction of Gaussian anamorphosis in the EnKF does not present any drawbacks. Furthermore, its computational overload is almost null comparing to the cost of the Forecast step of the EnKF that requires to run a large number of simulations. It is an easy and elegant solution to perform Kalman filter estimation in an extended framework of variables with non-Gaussian distributions. We thus encourage users of data assimilation to consider the pdfs of the state variables and observations before setting up the data assimilation experiment.

The Gaussian anamorphosis is by no means reserved to the EnKF but is naturally applied there because of Monte-Carlo formalism. It could be applied in a non-Monte-Carlo method provided that a random sampling is performed before the analysis step.

The assimilation of real satellite data with the EnKF with Gaussian anamorphosis has now to be investigated. It raises the challenging problem of model bias, well known in the data assimilation community, and particularly crucial for the use of anamorphosis functions built on the empirical marginal distributions of model variables. Furthermore two limits of the algorithm have been reached during these experiments: the first one concerns the assumption on an identical spatial distribution of the variables in the construction of the anamorphosis functions and the second one concerns the


ANA: $\bar{\mathbf{x}}^a - \mathbf{x}^t$

True state $\mathbf{x}^t$

ECO: $\bar{\mathbf{x}}^a - \mathbf{x}^t$

Fig. 12. $\bar{\mathbf{x}}^a - \mathbf{x}^t$: surface chlorophyll-$a$ component (mg/m^3) on 3 September 1998. Errors in the equatorial Atlantic Ocean are not plotted.

monovariate aspect of the algorithm. Works on the refinements in space of the anamorphosis functions or on multivariate transformations would allow a practical improvement of the algorithm. The statistical classification tools appear to be an interesting approach for the local refinement in space of the anamorphosis functions.

Appendix A

Construction of the empirical anamorphosis function based on the empirical marginal distribution

Given $\mathbf{Z}(x)$ the spatially distributed variable of interest. We assume that we do not know the marginal distribution of $\mathbf{Z}(x)$, but we have access to an approximation via a sample $(z_i)_{i=1:N}$ of this variable. The aim is to build a step function ψ such that

$$\mathbf{Z}(x) = \psi(\mathbf{Y}(x)) \tag{A1}$$

with $\mathbf{Y}(x)$ following a predefined marginal distribution. Here $\mathbf{Y}(x)$ is assumed to have a normal distribution $\mathcal{N}(0,1)$. The practical implementation of the anamorphosis follows:

- Sort the data $(z_i)_{i=1,N}$ by ascending values.

$$z_1 < z_2 < \ldots < z_{N-1} < z_N \tag{A2}$$

- Compute a sample $(y_i)_{i=1:N}$ of the Gaussian variable $\mathbf{Y}(x)$.

$$\forall i = 1:N, \quad y_i = G^{-1}\left(\frac{i}{N}\right) \tag{A3}$$

with G the cumulative distribution function of $\mathbf{Y}(x)$:

$$\begin{aligned} G(t) &= P(\mathbf{Y}(x) < t) \\ &= \int_{-\infty}^{t} f(y)dy \\ &= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} dy \end{aligned} \tag{A4}$$

- Define the empirical anamorphosis ψ.

$$\psi(y) = \sum_{i=1}^{N} z_i 1_{[y_{i-1},y_i[}(y) \tag{A5}$$

The empirical anamorphosis function ψ being non-bijective, one has to interpolate it. This is the aim of the last two steps of the algorithm (the interpolation of the empirical anamorphosis function and the definition of the tails).

Acknowledgements. The authors wish to thank the two referees M. Bocquet and P. Brasseur for their constructive comments. This study has been funded by the eVITA-EnKF project from the Research Council of Norway and the MERSEA IP from the European Commission's 7th FP. A grant of CPU time from the Norwegian Supercomputing Project (NOTUR) has been used. The authors are thankful to C. Hansen and K. A. Lisæter for providing the HYCOM-NORWECOM system and to B. Backeberg for the language corrections.

Edited by: B. Barnier

References

Aksnes, D., Ulvestad, K., Baliño, B., Berntsen, J., and Svendsen, E.: Ecological modelling in coastal waters: towards predictive physical-chemical-biological simulation models, Ophelia, 41, 5–36, 1995.

Allen, J. I., Eknes, M., and Evensen, G.: An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea, Ann. Geophys., 21, 399–411, 2003, http://www.ann-geophys.net/21/399/2003/.

Allen, J. I., Smyth, T. J., Siddorn, J. R., and Holt, J. T.: How well can we forecast high biomass algal bloom events in a eutrophic coastal sea?, Harmful Algae, 8(1), 70–76, 2008.

Bentsen, M., Evensen, G., Drange, H., and Jenkins, A. D.: Coordinate transformation on a sphere using conformal mapping, Mon. Weather Rev., 127, 2733–2740, 1999.

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, Int. Statist. Rev., 71, 223–241, 2003.

Bertino, L. and Lisæter, K. A.: The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans, J. Operational Oceanogr., 1(2), 15–19, 2008.

Bleck, R.: An oceanic general circulation model framed in isopycnic-cartesian coordinates, Ocean Model., 4, 55–88, 2002.

Campbell, J. W.: The lognormal distribution as a model for biooptical variability in the sea, J. Geophys. Res., 100(C7), 13237–13254, 1995.

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-biogeochemical model of the North Atlantic ocean, Ocean Model., 3, 167–192, 2001.

Chilès, J.-P. and Delfiner, P.: Geostatistics: Modeling Spatial Uncertainty, Wiley, New York, 1999.

Courtier, P., Thépaut, J. N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteorol. Soc., 120, 1367–1387, 1994.

Doucet, A., de Freitas, N., and Gordon, N.: Sequential Monte Carlo methods in practice, New York, Springer, 2001.

Drange, H. and Simonsen, K.: Formulation of air-sea fluxes in the ESOP2 version of MICOM, NERSC Report 125, Nansen Environmental and Remote Sensing Center, Norway, 1996.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143–10182, 1994.

Evensen, G.: The Ensemble Kalman filter: theorotical formulation and practical implementation, Ocean Dynam., 53, 343–367, 2003.

Evensen, G.: Data Assimilation, The Ensemble Kalman Filter, Springer, 2006.

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

Gregg, W. W. and Casey, N. W.: Global and regional evaluation of the SeaWiFS chlorophyll data set, Rem. Sens. Environ., 93, 463–479, 2004.

Gregg, W. W., Friedrichs, M. A. M., 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(1–2), 16–33, 2009.

Hansen, C. and Samuelsen, A.: Influence of horizontal model grid resolution on the simulated primary production in an embedded primary production model in the Norwegian Sea, J. Mar. Syst., 75(1–2), 236–244, 2009.

Hunke, E. and Dukowicz, J.: An elastic-viscous-plastic model for sea-ice dynamics, J. Geophys. Res., 27, 1849–1867, 1999.

Johannessen, J. A., Hackett, B., Svendsen, E., Søiland, H., Røed, L. P., Winther, N., Albretsen, J., Danielsen, D., Pettersson, L., Skogen, M., and Bertino, L.: The Norwegian Coastal Current – oceanography and climate, Chapter 11, edited by: Sætre, R., Tapir Academic Press, 2007.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Trans. ASME Ser. D, J. Basic Eng., 82D, 35–45,1960.

Lauvernet, C., Brankart, J.-M., Castruccio, F., Broquet, G., Brasseur, P., and Verron, J.: A truncated Gaussian filter for data assimilation with inequality constraints: application to the hydrostatic stability condition in ocean models, Ocean Model., 27, 1–17, 2009.

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

Lions, J. L.: Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles, Dunod, Paris, 1968.

Losa, S. N., Kivman, G. A., Schröter, J., and Wenzel, M.: Sequential weak constraint parameter estimation in an ecosystem model, J. Mar. Syst., 43, 31–49, 2003.

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, 2003.

Natvik, L. J. and Evensen, G.: Assimilation of ocean colour data into a biochemical model of the North Atlantic. Part 2. Statistical analysis, J. Mar. Syst., 40–41, 155–169, 2003.

Nerger, L. and Gregg, W. W.: Assimilation of SeaWiFS ocean chlorophyll data into a global ocean model using a local SEIK filter, J. Mar. Syst., 68, 237–254, 2007.

Pham, D. T., Verron, J., and Roubaud, M.-C.: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Mar. Syst., 16(3–4), 323–340, 1998.

Pham, D. T.: Stochastic methods for sequential data assimilation in strongly nonlinear systems, Mon. Weather Rev., 129(5), 1194–1207, 2001.

Sasaki, Y.: A fundamental study of the numerical prediction based on the variational principle, J. Meteorol. Soc. Jpn., 33, 262–265, 1955.

Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, Nonlin. Processes Geophys., 15, 761–772, 2008, http://www.nonlin-processes-geophys.net/15/761/2008/.

Skogen, M. and Søiland, H.: A user's guide to NORWECOM v2.0. The NORWegian Ecological Model system, Technical Report Fisken og Havet 18, Institute of Marine Research, Norway,1998.

Thacker, W. C.: Data assimilation with inequality constraints, Ocean Model., 16, 264–276, 2007.

Torres, R., Allen, J. I., and Figueiras, F. G.: Sequential data assimilation in an upwelling influenced estuary, J. Mar. Syst., 60, 317–329, 2006.

Triantafyllou, G., Hoteit, I., and Petihakis, G.: A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-D complex physical-biogeochemical model of the Cretan Sea, J. Mar. Syst., 40–41, 213–231, 2003.