

Papers published in *Ocean Science Discussions* are under open-access review for the journal *Ocean Science*

Controllability of mixing errors

D. Béal et al.

Controllability of mixing errors in a coupled physical biogeochemical model of the North Atlantic: a nonlinear study using anamorphosis

D. Béal¹, P. Brasseur¹, J.-M. Brankart¹, Y. Ourmières², and J. Verron¹

¹LEGI/CNRS, Université de Grenoble, CNRS, BP 53X, 38041 Grenoble, France

²LSEET, Université du Sud Toulon Var, 83957 La Garde Cedex, France

Received: 4 June 2009 – Accepted: 16 June 2009 – Published: 30 June 2009

Correspondence to: P. Brasseur (pierre.brasseur@hmg.inpg.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

In biogeochemical models coupled to ocean circulation models, vertical mixing is an important physical process which governs the nutrient supply and the plankton residence in the euphotic layer. However, mixing is often poorly represented in numerical simulations because of approximate parameterizations of sub-grid scale turbulence, wind forcing errors and other mis-represented processes such as restratification by mesoscale eddies. Getting a sufficient knowledge of the nature and structure of these error sources is necessary to implement appropriate data assimilation methods and to evaluate their controllability by a given observation system.

In this paper, Monte Carlo simulations are conducted to study mixing errors induced by approximate wind forcings in a three-dimensional coupled physical-biogeochemical model of the North Atlantic with a $1/4^\circ$ horizontal resolution. An ensemble forecast involving 200 members is performed during the 1998 spring bloom, by prescribing realistic wind perturbations to generate mixing errors. It is shown that the biogeochemical response can be rather complex because of nonlinearities and threshold effects in the coupled model. In particular, the response of the surface phytoplankton depends on the region of interest and is particularly sensitive to the local stratification. We examine the robustness of the statistical relationships computed between the various physical and biogeochemical variables, and we show that significant information on the ecosystem can be obtained from observations of chlorophyll concentration or sea surface temperature. In order to improve the analysis step of sequential assimilation schemes, we propose to perform a simple nonlinear change of variables that operates separately on each state variable, by mapping their ensemble percentiles on the Gaussian percentiles. It is shown that this method is able to substantially reduce the estimation error with respect to the linear estimates computed by the Kalman filter.

OSD

6, 1289–1332, 2009

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



1 Introduction

Our understanding of the ocean biogeochemistry and marine ecosystems has made significant progress during the past decade. Coupled physical-biogeochemical models are becoming a useful source of information for many practical applications of societal and environmental importance, such as the monitoring and forecasting of marine resources, water quality and the ocean carbon cycle. Biogeochemical modules are bound to be an essential component of the operational oceanographic systems that are being implemented, for instance, in the frame of the MERSEA and MyOcean European projects (Brasseur et al., 2009). In order to provide an accurate depiction of the essential biological variables, these models should be used in conjunction with global scale observation systems involving ocean colour satellites and profiling floats that, in the near future, will measure the sub-surface concentration of oxygen, chlorophyll and nutrients (e.g., Gruber et al., 2006). The optimal merging of these multiple types of information requires the development of purpose-built assimilation methods, taking into account the specificities of the coupled physical-biogeochemical models, and of the data available for assimilation.

In order to design appropriate assimilation methods and to evaluate the level of control that can be expected from a given observation system, it is necessary to explore the structure of the errors that affect the models and the observations. A standard way to explore the model errors is to perform Monte Carlo simulations (e.g., Evensen 1994). This requires making prior assumptions about the possible sources of errors, originating for instance in a set of model parameters or in forcing functions. One then postulates a prior probability distribution for these errors, from which a sample is drawn. Model integrations are then performed for each element of the sample, and the resulting ensemble simulation provides an image of the model error structure (a sample of its probability distribution). From this image, it is then possible to diagnose how the original errors cascade on the various model state variables, if the errors are correlated in space and time, if robust relationships exist between model and observation

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



errors, if these relationships are close to linearity, how a given observing system can be used to control these errors, etc. Ensemble statistics can also be used to determine to which extent the probability distribution functions (pdfs) are Gaussian, and from this, the theoretical properties of the assimilation methods required to control the errors. In the context of marine ecosystem modelling, it is useful for instance to understand the level of control that can be expected from ocean colour data.

In this paper, the Monte Carlo method is applied to the study of mixing errors in a coupled physical-biogeochemical model (CPBM) of the North Atlantic ocean (described in Sect. 2.1), with a specific focus on the analysis of the ecosystem response to these errors. It is well known indeed that a cautious control of the ocean stratification is crucial for consistent data assimilation in such coupled models (Berline et al., 2006), because it directly affects the diffusive flux of nutrient and plankton residence time in the euphotic layer through vertical mixing. Erroneous vertical mixing can be triggered by imperfections at different modelling stages, such as the wind forcing, the turbulent closure scheme or even the representation of mesoscale eddies through the restratification of the upper ocean (Oschlies, 2002).

Monte Carlo experiments are performed by prescribing perturbations in the wind forcing, which are assumed to be the main source of uncertainty in this study. Common knowledge suggests that wind errors will propagate into the system according to the scheme of Fig. 1. Wind perturbations first induce perturbations of the mixed layer dynamics which translate into modifications of the mixed layer depth (MLD) and sea-surface temperature (SST). Deepening or shallowing of the mixed layer then modifies the nutrient supply in the euphotic layer, and subsequently the phytoplankton production in the euphotic layer. The impact on the biogeochemical state can be measured by the surface nitrate (NO_3) and phytoplankton (PHY) concentration. The latter is directly related to surface chlorophyll concentration (CHL), a quantity that is well observed through ocean colour satellites. By following this conceptual causal chain in the ensemble, it will be possible to analyse the robustness of the statistical relationships between the successive model variables and the observed quantities, their dependence

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



on space and time, and eventually the possibility to inverse the observed information back to the model space and forcing functions. These questions will be examined in Sect. 3.

One of the results of the ensemble runs will be that even for short-term forecasts (1 day), the relationships between ecosystem variables and observations are not close to linear, so that they cannot be fully exploited by a linear estimation method. For such a system, nonlinear methods will be useful to improve the quality of the estimates. However, general nonlinear assimilation methods (e.g., particle filters as in Losa et al., 2003) which make no specific assumption about the shape of the prior pdf are too expensive for application to large size CPBM (16×10^6 state variables in our model), mainly because the identification of a general multivariate pdf with so many state variables would require too many ensemble members. Therefore, simplified solutions are needed to cope with real size problems.

A possible approach to nonlinear estimation problems is the use of anamorphosis transformations (i.e., Bertino et al., 2003), making nonlinear changes of variables to transform the forecast pdf (of arbitrary shape) into a Gaussian pdf. At first glance, this does not necessarily simplify the problem because identifying the change of variables requires a perfect knowledge of the original multivariate pdf, i.e. an ensemble as large as previously mentioned for particle filters. The simplified solution that we propose in this paper is to perform the change of variable separately for each state variable. It will be shown in Sect. 4 that, in this way, a moderate size ensemble is always sufficient to identify changes of variable to transform each marginal pdf to a nearly Gaussian pdf. This is obviously not sufficient to guarantee that the joint distribution becomes Gaussian. However, it is usually possible to diagnose from the ensemble the situations for which the approximation is accurate and the situations for which it is not. Both occur in our case study, so that we will be able to evaluate the relevance of the scheme in any situation. A quantitative evaluation of the improvement with respect to linear estimates is also attempted to conclude the study.

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2 Ocean model and wind forcing perturbations

2.1 The coupled physical-biogeochemical model

The CPBM used for the ensemble simulation was originally developed by Ourmières et al. (2009) for investigating the relative importance of nutrient vs. physical data to constrain the seasonal development of the phytoplankton bloom in the North Atlantic. The components of the coupled model include a NEMO/OPA9 circulation model of the North Atlantic basin at a $1/4^\circ$ horizontal resolution (see Sect. 2.1.1), and a biogeochemical model derived from the 6-compartment LOBSTER formulation (see Sect. 2.1.2). The central model simulation (without wind perturbation), that will serve as a reference for the Monte Carlo simulations, is described in Sect. 2.1.3.

2.1.1 The North Atlantic Ocean circulation model

The circulation model is a DRAKKAR configuration (The DRAKKAR Group, 2007) of the OPA9/NEMO model (Madec et al., 1998), which is a primitive equation model using a free surface formulation. The domain covered is the North Atlantic basin from 20° S to 80° N and from 98° W to 23° E. The horizontal grid has a so-called eddy-permitting resolution of $1/4^\circ$ (Barnier et al., 2006). The vertical discretization is done using 45 geopotential levels, with a grid spacing increasing from 6 m at the surface to 250 m at the bottom. Vertical mixing of momentum and tracers is modelled by the TKE turbulence closure scheme (Blanke and Delecluse 1993), and convection is parameterized with enhanced diffusivity and viscosity. Buffer zones are defined at the southern, northern and eastern (Mediterranean) boundaries with relaxation to Levitus temperature (TEM) and salinity (SAL) climatology (Levitus et al., 2001). The forcing fluxes are calculated using bulk formulations and the ERA40 atmospheric forcing fields (ECMWF 2002). The prognostic variables include the zonal and meridional velocity components (U and V), temperature, salinity and sea surface height (SSH).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2.1.2 The LOBSTER biogeochemical model

LOBSTER (LOcean Biogeochemical Simulation Tools for Ecosystem and Resources) is a nitrogen-based ecosystem model that includes 6 prognostics variables in the euphotic layer: nitrate (NO_3), ammonium (NH_4), phytoplankton (PHY), zooplankton (ZOO), detritus and semi-labile dissolved organic nitrogen (Levy et al., 2005a). The bottom of the euphotic layer is prescribed at a constant depth of 191 m. Below the euphotic layer, the model considers very simple parameterizations of decay to nitrate, detritus sedimentation and remineralization of zooplankton mortality. LOBSTER is coupled on-line to the circulation model without feedback of the biogeochemical variables on the physics. The coupling frequency is equal to the circulation model time-step (40 min). The on-line coupling as well as the maximum frequency is thought to allow accurate diagnostics of the ecosystem evolution without possible problems brought by the use of averaged physical fields as an off-line configuration would need. More detail about the model equations is available in Levy et al. (2005a and 2005b) and about the North Atlantic implementation in Ourmières et al. (2009).

2.1.3 Reference simulation of the coupled model

The reference simulation of the coupled model used in this study corresponds to year 1998 of the FREE simulation described in Ourmières et al. (2009) and performed without data assimilation. In this simulation, the U and V components of the velocity and the SSH variables are initialised to zero, while the TEM and SAL variables are interpolated using the December Levitus climatology (Levitus et al., 1998). Then, the physical model is run for 12 years from 1 January 1984 to 1 January 1996, providing a balanced physical ocean state to start the biogeochemical model spin-up. At that time, the nitrate field is initialized with the December Levitus climatology 2001 (Conkright et al., 2002) interpolated on the model grid. The other biogeochemical variables are set to constant values in the euphotic zone and to zero below: zooplankton is set to 0.01 mmol N/m^3 , phytoplankton to 0.1 mmol N/m^3 and ammonium, dissolved organic matter and detritus

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



to 0.001 mmol N/m³. The coupled model is then run for 2 years starting January, the 1st 1996 and using the physical ocean state obtained after 12 years of spin-up. Ourmières et al. (2009) analysed the convergence of the run and showed that the model was able to reproduce satisfying seasonal cycles of the biogeochemical variables. In this study, we will focus on the monthly period starting 15 April 1998, i.e. when the bloom event occurs.

2.2 Wind ensemble perturbations

In order to produce the Monte Carlo simulations, wind perturbations must be introduced in the model forcing. In this section, we describe the method used to generate perturbations of the wind components that are physically realistic. We proceed in two steps, assuming that the uncertainty in the wind can be estimated from the seasonal and interannual variability of ERA40 winds during 1985–2000: (i) the covariance of the wind variability is calculated using the ERA40 database, and (ii) the wind perturbations are randomly sampled from a Gaussian probability distribution function with zero mean and this pre-calculated covariance.

In practice, an ensemble composed of one wind field every 4 days is extracted from the 1985–2000 ERA40 winds during the 3 months period centred on 15 April. This ensemble contains 368 members representative of the season during which the Monte Carlo simulations are performed. An EOF analysis of this ensemble is performed, and the first 50 dominant EOFs (representing 80% of the wind variance) are used to generate the perturbations. Figure 2 (left panel) illustrates the first 50 eigenvalues in decreasing order, their corresponding percentage of explained variance and the cumulated percentage of explained variance. Figure 2 (right panel) also shows the wind variability standard deviation which is also the expected standard deviation of the wind perturbations. It is especially large over the subpolar gyre and over the Gulf Stream region. As mentioned above, wind perturbations will generate anomalies of the biogeochemical model variables. As a result, a more intense ecosystem response will be

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



expected in the subpolar and Gulf Stream regions. These regions are also the location where the intensity of the spring bloom is maximum in the reference simulation.

The Monte Carlo simulations are then performed using an ensemble of 200 time-varying perturbations of the wind forcing. Assuming that the typical decorrelation time scale of wind errors is about 4 days, we draw independent 200-member samples every 4 days (with the covariance defined above). These are then interpolated linearly in time to obtain perturbations every 6 h, which is the input frequency of forcing fields in the ocean model. In practice, this corresponds to sample coefficients for each EOF in $\mathcal{N}(0,1)$ every 4 days, interpolate them in time to obtain the perturbation amplitude $\alpha_i(t)$ for every EOF_{*i*}, *i* = 1...50, and then compute the perturbed wind using Eq. (1). It is worth noting that in Eq. (1), the normalized EOFs are multiplied by the squared root of the corresponding eigenvalue, so that each EOF is a column of the squared root of the perturbation covariance matrix.

$$\text{WND}_{\text{pert}}(t) = \text{WND}_{\text{reference}}(t) + \alpha_1(t)\text{EOF}_1 + \dots + \alpha_{50}(t)\text{EOF}_{50} \quad (1)$$

3 Study of the ensemble forecast

The objective of this section is to describe the ensemble response of the model to the wind perturbations described in section 2. This response will be analysed by studying the ensemble forecast at a dozen of locations in the North Atlantic (at BATS, INDIA and NABE biogeochemical stations, in the Gulf Stream – hereafter noted GS –, the Labrador Sea – hereafter noted LAB–, the subtropical gyre, the Mauritanian coast, the Azores and the Norwegian coast regions), and by studying how the correlations between model variables evolve with time. In addition, we will investigate the time scales over which the correlations between the wind and the observed quantities can be considered robust enough to be exploited by data assimilation systems.

More precisely, we will diagnose ensemble scatterplots of variables characterizing the relationships in the transfer function described in Fig. 1, i.e. WND/MLD (WND is

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the wind stress modulus expressed in N/m^2), MLD/TEM, MLD/ NO_3 , MLD/PHY and NO_3 /PHY. To interpret in more detail the mechanisms behind these relationships, we will also need to analyse the distribution of vertical TEM, NO_3 and PHY profiles obtained from the ensemble. In addition, the information extracted from the ensemble will be synthesized using simple statistics:

- the linear correlation coefficient (Pearson):

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (2)$$

where $x = (x_i)_{i=1}^n$ and $y = (y_i)_{i=1}^n$ are n-size samples of 2 random discrete variables and \bar{x} and \bar{y} are respectively the mean of the samples;

- the rank correlation (Spearman) that is identical to the linear correlation except that each value x_i (respectively y_i) is replaced by the value of its rank R_i (respectively S_i) among all other x_i 's (respectively y_i 's) in the sample (i.e. the index of x_i in the sorted sample). The sequence R_i contains thus all integers between 1 and n :

$$r_s = \frac{\sum_i (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_i (R_i - \bar{R})^2} \sqrt{\sum_i (S_i - \bar{S})^2}} \quad (3)$$

where \bar{R} and \bar{S} are respectively the mean of R and S . The rank correlation is useful to detect non linear relationships between variables; it is also more robust than the linear correlation coefficient to some defects in the data (see Press et al., 1992, chapter 14).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



3.1 The ensemble response at three locations

By looking at the ensemble forecast after only one day of run, we will see that mixing will be the dominant mechanism responsible for the propagation of wind forcing errors to the other state variables, in most locations. This is because the daily time scale is too short to trigger intense dynamical interactions between the biogeochemical variables of the LOBSTER model.

The ensemble response is analysed in details at three specific locations: at the BATS station (station 5 in Fig. 2; see Fig. 3), the GS station (station 9 in Fig. 2; see Fig. 4) and the INDIA station (station 11 in Fig. 2; see Fig. 5). The figures show the five scatterplots describing the transfer function of Fig. 1, as well as the ensemble vertical profiles of temperature, nitrate, phytoplankton and zooplankton. We will discuss in sequence the propagation of uncertainties from the wind forcing to the physical properties, and then to the biogeochemical properties of the mixed layer. The corresponding correlation statistics are given in Table 1 for all 12 stations shown in Fig. 2.

3.1.1 Relationships between wind forcing and physical properties of the mixed layer

As a first step, we analyze the cascade of errors from the wind forcing to the physical variables (first line in Fig. 1).

WND/MLD. Wind errors generate different types of response on the mixed layer depth (see WND/MLD scatterplots in Figs. 3, 4 and 5). As a general rule, the larger the wind, the deeper the mixed layer; however, there are significant differences between the 3 situations. At INDIA station, the relative modifications of the mixed layer depth around 400 m are significantly smaller than those observed at BATS and GS, for similar perturbations of the wind. Further, the spread around the linear regression is large for small wind anomalies, while such spread does not occur in the same way at the other stations. For large wind anomalies, one can observe a sort of saturation of mixed layer depth perturbations. The relationship between WND and MLD is obviously non-linear

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Controllability of mixing errors

D. Béal et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

at INDIA station. This can be explained by the very different mixed layer structures of the 3 reference states: at BATS, the mixed layer is very shallow and the turbulent energy brought by the wind immediately propagates down to the thermocline. The exactly opposite situation occurs at INDIA, where the water column of the reference run is well mixed down to around 400 m. As a result, the mixed layer depth is relatively insensitive to wind anomalies.

MLD/TEM. In general, the consequence of the mixed layer deepening when wind forcing increases is a cooling of the sea surface (see TEM/MLD plots in Figs. 3, 4 and 5). The mixing of warm surface water with cold water at depth results in a cooling of the mixed layer. The TEM/MLD relationships decrease monotonously, but not necessarily in a linear way. As an evidence, the shape of this relationship depends on the shape of the vertical TEM profile. Moreover, the statistics of Table 1 show very high rank correlations, meaning that a quite robust relationship may exist for this combination of variables.

3.1.2 Relationships between mixed layer and biogeochemical properties

As a second step, we analyze the cascade of errors from the mixed layer to biogeochemical variables (second line in Fig. 1).

MLD/NO₃. Deepening of the mixed layer is expected to bring nitrate to the surface by mixing nutrient-rich deep water with nutrient-depleted surface water. This is exactly what happens at BATS and GS stations, where a non-linear increase of NO₃ concentration is observed when the mixed layer deepens. From the scatterplot of the Gulf Stream station, one can however notice the existence of a plateau around the reference NO₃ concentration of 1.5 mmol m⁻³: the perturbations of the wind below some threshold is unable to propagate anomalies down to the nutricline depth. By contrast, the wind reduction yields restratification of the water column, which favours the consumption of NO₃ by phytoplankton. At INDIA station, we observe the same phenomenology as for MLD: the wind perturbations are not strong enough to significantly modify the NO₃ concentration over the whole 400 m mixed layer.

**Controllability of
mixing errors**

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



MLD/PHY. As phytoplankton concentration typically dominates in the euphotic zone and weakens at depth, phytoplankton is expelled from surface layers by mixing and the MLD and PHY variables are negatively correlated. A non-linear decrease of PHY concentration is thus observed when the mixed layer deepens. Compared to nitrate at BATS and GS stations, this is an exactly opposite behaviour and again, mixing seems to be the dominant effect. The INDIA station still shows a complex response which is difficult to interpret by simple mechanisms.

Finally, we analyze the scatterplots between the NO_3 and PHY biogeochemical variables.

NO_3 /PHY. Surface phytoplankton generally decreases when nitrate concentration increases, as a results of the inverse distribution of these two quantities over the water column. The scatterplots can be characterized by well-defined relationships with pretty high correlations, sometimes altered by threshold effects as illustrated for the Gulf Stream Station. In the LOBSTER model, the phytoplankton growth is made possible by 2 different pathways: the new production sustained by nitrate, and the regenerated production sustained by ammonium. A cluster of high phytoplankton concentrations can be observed at BATS station for poor nitrate values, which might be explained by the regenerated phytoplankton production associated to very thin MLD. This is an example where a biogeochemical mechanism, different than mixing, transforms the error propagation in the coupled model.

In summary, the results discussed here above indicate that the propagation of wind errors after a one-day forecast is strongly dependent on the local stratification of the ocean. Mixing seems to be the dominant mechanism explaining the behaviour of the ensemble. In a first approximation, the state variables (TEM, NO_3 , PHY) can be considered as passive tracers as long as the lead time remains small (one day). Further, the relationships between variables are generally loosing their robustness when the mixed layer deepens. The response of the CPBM after one day can be very complex, demonstrating non linear relationships between state variables with sometimes threshold effects. In the following section, we will focus on the evolution of the ensemble

spread and the corresponding correlations with time.

3.2 Temporal evolution of the ensemble response

The objective of this section is to analyse the stability of these statistical relationships over a 2 week period after the application of wind perturbations. Figures 6 (BATS station) and 7 (Gulf Stream) show the scatterplots after 1, 2, 4, 8 and 15 days of run, illustrating the temporal evolution of relationships between variables.

The spread of the ensemble with time is the first general trend clearly illustrated by these 2 figures. The more the experiment lasts, the larger the dispersion (following each line from left to right), and the variables tend to decorrelate with time. This is particularly visible for MLD/TEM, PHY/TEM and MLD/PHY relationships, leading for instance to an almost complete decorrelation after 8 or 15 days between WND and MLD, or between PHY and TEM, at BATS station. Note that sometimes one can observe a decorrelation during the first days of run followed by the recorrelation of the ensemble, as for example on the MLD/PHY scatterplots at Gulf Stream station before and after the 4th day of run.

The shape of the relationships may also change with time. For instance, the nonlinear TEM/MLD relationship at BATS station is getting almost linear after the 8th day of run (except for small MLD values). More than that, initially well-defined relationships such as TEM/MLD and PHY/MLD at Gulf Stream station are becoming fuzzy after 4 days of run, and recover some structure after 8 or 15 days, but with a different shape. Finally, scatterplots could also disperse in such a way that no relationship exists anymore (e.g., PHY/TEM scatterplots on Fig. 6 after 8 days).

As a conclusion, the ensemble response of the CPBM at lead times greater than one day is quite complex, with often enhanced dispersion and structural modification of the relationships. The temporal evolution of the scatterplots shows that reasonable relationships are sometimes preserved after 4 days of wind perturbation (e.g., at BATS station), sometimes not (e.g., at Gulf Stream Station). In particular, relationships at BATS station obtained after a 4-day forecast could be used to determine the cascade

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



of errors from WND to MLD, from MLD to TEM, and finally from TEM to PHY. In the following section, we will discuss the potential utilization of observed chlorophyll data by inverting such relationships to control the state variables of the CPBM. In particular, the existence of robust relationships over 4 to 6 days temporal windows will be examined in a data assimilation perspective.

3.3 Observability of physical and biogeochemical variables using chlorophyll data

The objective of this section is to determine what is controllable in the CPBM using surface chlorophyll measurements over typical data assimilation time scales of 4 to 6 days. We propose to use the example of BATS station after 4 days of wind perturbations to illustrate how to use the chain of errors with a linear observational update.

The scatterplots of Fig. 8 can be used to describe the backward propagation of information from ocean colour observations to the model variables, consistently with the scheme of Fig. 1. Indeed, chlorophyll measurements provide direct access to phytoplankton concentration information since the 2 quantities are linearly linked in the CPBM. The linear regression line of the ensemble shown on the first scatterplot of Fig. 8 allows an estimation of nitrate concentration from phytoplankton. A similar regression line on the second scatterplot then provides the mixed layer depth, which could also be retrieved using the PHY/MLD scatterplot of the third scatterplot. MLD information then provides estimates of temperature (fourth scatterplot), possibly in conjunction with directly observed temperature data. MLD can also be used to retrieve the wind tension (fifth scatterplot). For each scatterplot, the observational updates (blue dots) correspond to the projection of the ensemble (red dots) among the linear regression line (green line). The distance to the reference value (big blue dot) is an indication of the estimation error.

A limitation of the previous method is that the quality of the linear update requires linear relationships with sufficiently low dispersion to compute accurate inverse estimates. Linear updates could be used at stations such as BATS, but Gulf Stream or

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



INDIA stations might ideally benefit from non linear relationships. In the next section, we will demonstrate how linear updating methods can be modified to take into account such non-linear relationships. Examples of application to the North Atlantic ocean will be discussed to quantify the improvement.

4 Toward data assimilation: inference method using anamorphosis

The diagnostics of the ensemble forecasts presented in the previous section show the omnipresence of non-Gaussian behaviours as well as nonlinear relationships between state variables. As a consequence, a linear observational update cannot be optimal. The purpose of this section is to propose a method to do better than the linear estimate. In the first subsection (Sect. 4.1), we demonstrate the problems that occur if a linear observational update is used. This will be done at the surface of the ocean using the reference phytoplankton as observation, and each member of the ensemble as background state. In a second stage (Sect. 4.2), a simple nonlinear transformation of the variables (anamorphosis) is proposed to execute the observational update. And finally, in Sect. 4.3, we illustrate how this anamorphic transformation can improve the quality of the estimation. The gain obtained in our specific case study is quantified for the whole North Atlantic domain.

4.1 Problems with linear observational update

The linear observational update, that is used in conventional Kalman filters, is computed using the formula:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K} (\mathbf{y} - \mathbf{H}\mathbf{x}^f) \quad (4)$$

where \mathbf{x}^f is the forecast (or background) state, \mathbf{y} is the observation vector, \mathbf{H} is the observation operator and \mathbf{K} is the gain. It minimizes the estimation error variance (and

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



thus corresponds to the best linear unbiased estimate) if the gain is computed by:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (5)$$

where \mathbf{P}^f is the forecast (or background) error covariance matrix and \mathbf{R} is the observation error covariance matrix. It can be demonstrated that the gain (4) provides the absolute minimum error variance estimate (not only the best linear one) providing that the probability distributions are Gaussian. In this case, it also corresponds to the maximum likelihood estimate. If the pdf are not Gaussian, it is possible that better estimates are found.

In this paper, we restrict ourselves to the problem of estimating one state variable from the perfect observation of another state variable. For instance, in Fig. 9, we estimate the mixed layer depth from one phytoplankton observation. We use the reference simulation (large blue dot) as observation, and in order to get a solution that is statistically valid, we use successively each member of the ensemble as background (red dots). The solution will be deduced from the distribution of the updated values (small blue dots). In Sect. 4.1, we focus on the left panel of Fig. 9 which illustrates the linear observational update. For that specific example, formula (6) rewrites

$$\text{MLD}^a = \text{MLD}^f + \gamma \frac{\sigma_{\text{MLD}}}{\sigma_{\text{PHY}}} (\text{PHY}^o - \text{PHY}^f) \quad (6)$$

where $(\text{PHY}^f, \text{MLD}^f)$ are the background values (red points), PHY^o is the observed value (abscissa of the large blue dot), $(\sigma_{\text{PHY}}, \sigma_{\text{MLD}})$ are the ensemble standard deviation for PHY and MLD, and γ is the linear correlation coefficient between PHY and MLD. Since the observation is perfect, all updated values $(\text{PHY}^o, \text{MLD}^a)$ (blue dots) are aligned vertically on the PHY^o value.

From the previous equation, it is apparent that the observational update (from the red point to the blue point) is done along a straight line with the given slope $\gamma \frac{\sigma_{\text{MLD}}}{\sigma_{\text{PHY}}}$, which is the slope of the linear regression line (in green on the figure) passing through

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the ensemble mean (green square). Hence, in this simple example, the ensemble observational update can be viewed as drawing from each red point a parallel to the green line and find the updated value at the intersection of this line with the vertical $\text{PHY}=\text{PHY}^o$.

5 But, from the ensemble displayed in Fig. 9 (red points), it is quite clear that the pdf is far from being Gaussian. For example, the quartiles of the marginal distributions (thin dashed lines) are not symmetric around the median (thick dashed line). On the other hand, in a general two-dimensional pdf, the regression line (for instance for MLD) is defined as the line with maximum MLD probability density for each value of the other variable (PHY). If a pdf is Gaussian, the regression line is a straight line and corresponds to the linear regression line defined above (and drawn in green in the figure). Obviously, in our example, the maximum MLD probability for each PHY value is usually well above or well below the linear regression line, indicating again a non-Gaussian behaviour. Hence performing the observational update by following the linear regression line without exploiting the real shape of the distribution necessarily leads to suboptimal estimates, with significantly larger estimation errors. Moreover, we observe in Fig. 9 that the true regression line has a general positive curvature, so that the linear estimate is almost systematically above the true MLD value.

4.2 Nonlinear observational update using anamorphosis

4.2.1 Description of the anamorphosis transformation

We propose here a simplified nonlinear method with the general idea of transforming each marginal pdf to a pdf that is close to Gaussian. This is achieved by performing a change of variables (anamorphosis) separately for each single variable of the state vector. For instance, Fig. 10 (left panel) shows the ensemble distribution of surface nitrate at the BATS station. Again, the pdf is obviously far from Gaussian. Let us denote by x the original random variable, and by $y=f(x)$, the transformed random variable. The objective is to find the function f defining a change of variables (anamorphosis)

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



[Title Page](#)
[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

such that the random variable y is as close as possible to the Gaussian pdf $\mathcal{N}(0, 1)$. Moreover, we want to infer f from the ensemble description of the pdf of x .

In order to reach that objective, the idea is to use the piecewise linear change of variable f to remap as set of percentiles of the pdf of x to the same percentiles of $\mathcal{N}(0, 1)$.

- 5 For instance, if $x_k, k=1, \dots, p$ are the p percentiles of x (such that $p(x < x_k) = r_k$, for a given set of values $r_k, k=1, \dots, p, 0 < r_k < 1, r_k < r_{k+1}$), and y_k are the corresponding percentiles of $\mathcal{N}(0, 1)$, the function $f(x)$ writes:

$$f(x) = \begin{cases} y_1 & \text{for } x < x_1 \\ y_k + \frac{y_{k+1} - y_k}{x_{k+1} - x_k} (x - x_k) & \text{for } x \in [x_k, x_{k+1}] \\ y_p & \text{for } x > x_p \end{cases} \quad (7)$$

- 10 This change of variables is only uniequivocal on the range $[x_1, x_p]$ so that the reciprocal function is only defined on the range $[y_1, y_p]$. In order to go back to the original space, we will use the transformation $x=g(y)$ defined by

$$g(y) = \begin{cases} x_1 & \text{for } y < y_1 \\ x_k + \frac{x_{k+1} - x_k}{y_{k+1} - y_k} (y - y_k) & \text{for } y \in [y_k, y_{k+1}] \\ x_p & \text{for } y > y_p \end{cases} \quad (8)$$

- 15 In order to reduce as much as possible the region of the state space out of the interval $[x_1, x_p]$, a good idea is certainly to include in the list of percentiles, the minimum of the ensemble as x_1 (as percentile $r_1 = 1/2n$ if n is the size of the ensemble) and the maximum of the ensemble as x_p (as percentile $r_p = \frac{2n-1}{2n}$). In that way, all estimates will always be in the range described by the original ensemble, and no extrapolation is possible.

- 20 Figure 10 (middle panel) shows the transformation that is obtained for the surface nitrate concentration at BATS station, using $p=20$ equidistant percentiles (dividing the pdf into 20 equidistant intervals), and Fig. 10 (right panel) shows the resulting distribution in the transformed space. By construction, this distribution has the same 20

percentiles as $\mathcal{N}(0, 1)$ and is thus close to Gaussian. The quality of the transformation relies on one subjective choice, which is the set of percentiles $r_k, k=1, \dots, p$. The larger p , the more complex is the change of variables that it is possible to represent. But a complex transformation needs a large ensemble to be properly identified. It is certainly a good policy to keep p small with respect to the size of the ensemble ($p \ll n$), and to distribute the percentiles as regularly as possible, for instance (with p odd): $r_1 = \frac{1}{2n}, r_k = \frac{k-1}{p-1}, 2 \leq k \leq p-1, r_p = \frac{2n-1}{2n}$.

Note that our approach is quite different from the Gaussian anamorphosis algorithm proposed by Simon and Bertino (2009) to assimilate ocean colour data in a North Atlantic model using the EnKF. In their study indeed, each model variables is transformed using the same univariate anamorphosis function at all grid points of the model. In the present implementation, the transformation is adjusted locally using the ensemble statistics obtained at each particular grid point.

4.2.2 Observational update in the transformed space

We now apply this idea to the example presented in Sect. 4.1 (Fig. 9). We thus transform *separately* the MLD and PHY variables according to Eq. (7), using their respective percentiles corresponding to: $r_k = 0.0025, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.9975, k=1, \dots, 11$. The transformed scatterplot is shown in Fig. 10 (middle panel). The dotted line corresponds to the Gaussian percentiles $y_k \sim -1.28, -0.84, -0.52, -0.25, 0, 0.25, 0.52, 0.84, 1.28, k=2, \dots, 10$ (the two extreme ones are not drawn). By construction, each marginal pdf (for MLD and PHY) has got the same percentiles y_k as a Gaussian pdf. More remarkably, the mean of the transformed ensemble is close to the origin of the axes, and the linear regression line (green line) is always close to the true regression line (corresponding to maximum MLD probability for each PHY value): these are two features that are not guaranteed by the method and that depend on the shape of the initial ensemble distribution. Moreover, due to the transformation the linear correlation coefficient between MLD and PHY has increased from 0.85 to 0.97. We thus observe that in this particular case, it is more appropriate to perform

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the ensemble observational update in this transformed space (blue dots) since moving parallelly to the linear regression line (in green) is certainly here the right thing to do (even if there are still a few members that are significantly above the linear regression line).

5 After that, we transform the solution back into the original space using Eq. (8) (Fig. 9, right panel). As expected, the ensemble of updated values (blue dots) is closer to the true state (large blue dot). The updated ensemble error variance is thus much smaller than it was using directly the linear observational update (compare to Fig. 9, left panel). If we also transform back the linear regression line from the transformed space (the
10 green straight line in the middle panel of Fig. 9), we obtain the thick green line of the right panel. We observe that it is very close to the true nonlinear regression line (maximum MLD probability for each PHY value). Performing the observational update in the transformed space is more or less like moving along this nonlinear regression line, which leads obviously to a smaller resulting error variance.

15 In order to analyse the situations in which the method is likely to work correctly, we now redo mentally the same exercise for some of the example scatterplots presented in Sect. 3. Four kind of situations may be distinguished. (i) The data are well correlated and the regression line is linear (as for instance, in Fig. 3: WND/MLD, PHY/NO₃, in Fig. 4: WND/MLD or in Fig. 5: MLD/TEM, NO₃/PHY). In this situation, the linear
20 observational update already exploits quite correctly the information contained in the observed variable, and only little improvement can be expected from the transformation. (ii) The data are well correlated, the regression line is nonlinear and monotonuous (as for instance MLD/TEM, MLD/NO₃, MLD/PHY in Fig. 3 and MLD/TEM, MLD/NO₃, MLD/PHY, NO₃/PHY in Fig. 4). In this situation, performing a linear observational update (following the linear regression line in green) is not a good solution, and making the
25 simple anamorphosis described above always leads to a significant improvement. This is the typical case for which it is designed, and the proposed solution is in this case very close to optimality. (iii) The data are well correlated (nonlinearly), the regression line is nonlinear and non-monotonuous (as for instance WND/MLD, MLD/NO₃ or MLD/PHY in

Controllability of mixing errors

D. Béal et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Controllability of
mixing errors**

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Fig. 5). In this situation, our simplified method does not fully solve the problems of the linear observational update, and remains quite suboptimal. No separate transformation of the two variables can transform the non-monotonuous regression line into a straight line; a joint two-dimensional nonlinear transformation (or another method) would be needed here. However, the method that we propose is not likely to be worse than the linear observational update. (iv) The data are poorly correlated (as can happen after a longer forecast in Figs. 6, 7 or 8). In this situation, transforming the variables does not help a lot: not much can anyway be expected from the multivariate observational update.

Up to now, the method has only been applied to a state vector made of 2 variables and with a perfect observation of one of the variable. However, the method is general and can be applied for any number of state variables and observations. One only needs to transform every state variables and observations separately and perform the standard multivariate observational update in the transformed state space. (If the observation operator is complex, transforming the corresponding observation requires computing the model equivalent to that observation for each member of the ensemble and find the function f given by Eq. (7) from this ensemble of value.) If the observations are not perfect, we need also to transform the observation error standard deviation. This can be done approximately by using the local slope of the transformation f as a scale factor. The additional cost of these operations with respect to the linear observational update is very small so that the method can easily be applied to large size systems (see Sect. 4.3).

It is also worth noting that the method also solves the problem of inequality constraints that can exist on the value of some state variables, for instance $a \leq x \leq b$. The linear observational update (assuming Gaussianity) can indeed often violate such constraints, thus leading to inappropriate estimates. With anamorphosis 7 and 8, it is sufficient to choose $x_1 \geq a$ and $x_p \leq b$ for the final estimate to satisfy the inequality constraints. This can be compared to the truncated Gaussian filter proposed by Lauvernet et al. (2009) to solve the problem. By contrast to their approach, the method described

here can only deal with inequality constraints that apply separately on each state variable. Moreover, a larger size ensemble is required to identify the anamorphosis than to identify a truncated Gaussian pdf. The truncated Gaussian filter is thus cheaper, it can deal with more general inequality constraints, but the shape of the prior pdfs is less general (truncated Gaussian pdfs are assumed).

4.3 Application of the non linear update over the North Atlantic

The results detailed in the previous section for the BATS station are here generalized to the whole North Atlantic domain, i.e. the same exercise is repeated at every model grid point (no horizontal correlations are taken into account here). The surface phytoplankton of the reference simulation is considered to be the observation (still assumed perfect), and the 1-day ensemble forecast at surface is used the same way to compute the observational update (i) in the regular state space and (ii) in the anamorphosed state space. The benefit of the transformation is characterized by the standard deviation of the updated ensemble.

Figure 11 shows the standard deviation of the 1-day forecast ensemble for the mixed layer depth, nitrate and zooplankton concentration before the observational update. It represents the standard deviation of the error that we want to reduce using the phytoplankton observations. The maps show that the largest MLD errors (left panel) are located in the Northern part of the domain that corresponds to large wind standard deviations (see Sect. 2). Large MLD errors usually yield large NO_3 errors (middle panel), as can be expected from the scheme in Fig. 1. In times, this leads to errors in the primary and secondary productions, that are nevertheless confined here to the Gulf Stream region (see ZOO errors standard deviations range, in the right panel of Fig. 11), because spring bloom starts in that area at the time of this experiment (15 April).

Figure 12 shows the error standard deviation reduction that is obtained with the linear observational update, i.e. the ratio of the updated ensemble standard deviation to the forecast ensemble standard deviation (that is shown in the Fig. 11), and Fig. 13 shows the same result obtained using the anamorphosis scheme. These results can

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Controllability of mixing errors

D. Béal et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

be analysed using the classification given in Sect. 4.2.2. (i) There are regions and variables for which the linear observational update is already very good and not much can be expected from anamorphosis to significantly improve the solution (it can even degrade it). In these regions, the variable (MLD, NO_3 or ZOO) is well correlated to PHY and the regression line is linear. (ii) There are also many regions where the error standard deviation can be substantially reduced by anamorphosis. In these regions, the variables are well correlated to PHY but along a nonlinear regression line, so they are controllable through PHY observations but not with a linear analysis scheme. (iii) Finally, there are regions where nor the linear observational update, nor anamorphosis can reduce the forecast error that was induced by the wind perturbations. These errors cannot be controllable by PHY observations only. Direct observations would be necessary. Fortunately, they mostly corresponds to regions where the forecast ensemble error is small (see Fig. 11). Here, the wind is thus not likely to be one of the dominant sources of errors, so that no conclusion of practical consequence can be derived from this simplified study involving wind errors.

Finally, in order to investigate the performance of the method for longer lead times, the same experiment has been repeated for the ensemble forecast at days 2, 4, 8 and 15. In order to summarize the results, Fig. 14 shows for each case study, the fraction of the domain (X-axis) for which the error reduction factor by the ensemble observational update (fully illustrated at day 1 by the maps in Figs. 12 and 13) is lower than a given value (Y-axis). Thus, the lower the curve, the largest fraction of the domain below a given reduction factor. For instance, at day 1 (thickest curves), the nonlinear observational update (with anamorphosis) is always better than the linear observational update (as already diagnosed from Figs. 12 and 13). As the length of the ensemble forecast increases (from day 1 to day 15, from thick curves to thin curves), all three variables tend to decorrelate from phytoplankton observations (see Sect. 3.2), so that the accuracy of the estimation is deteriorating with time whatever the analysis scheme. We observe however that the nonlinear scheme remains most often significantly better from day 1 to day 15 (except for zooplankton at day 15), which means that there are

still regions where nonlinear correlations can be exploited to improve the observational update. This figure can also be viewed as a synthetic (linear vs nonlinear) measure of the controllability of these errors by phytoplankton observations, indicating that, in this case study, controllability is decreasing with time.

5 Conclusions and perspectives

The Monte Carlo experiments that were designed to study mixing errors in a coupled physical biogeochemical model of the North Atlantic yield a number of conclusions in the perspective of ocean colour data assimilation. As a general rule, the results of the ensemble forecasts validate the conceptual transfer function proposed in the introduction (Fig. 1): the first order causal relationships summarized in the figure lead to tight correlations. However, the response is rather complex, depending in particular on the local stratification, in such a way that even the general features of the probability distributions can change radically in space and time (e.g. sign and strength of the correlation, shape of the regression lines, asymmetry between positive and negative anomalies, presence of thresholds, . . .). More embarrassing, the tight correlations (in a nonlinear sense) observed for short term forecasts (1 day) decrease quickly with time, and thereby reduce the level of control that can be expected from a partial observing system like surface temperature and surface chlorophyll. Nevertheless, our results suggest that, in many regions, a significant error variance reduction (on all variables shown in Fig. 1) can be obtained from these observations if the forecast does not exceed a few days (2 to 7 days as a function of the region), and providing that the nonlinear relationships between the variables are appropriately exploited.

Nonlinearities in the model indeed lead to many kinds of non-Gaussian behaviours, that cannot be properly handled by classical linear assimilation methods. In order to tackle this problem at moderate cost (i.e. in a way that is compatible with large size data assimilation problems), a simplified approximate nonlinear scheme has been proposed in this paper. The idea is to perform a nonlinear change of variables (anamor-

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



phosis) separately for each element of the control vector, by remapping the ensemble percentiles of their marginal distribution to Gaussian percentiles. In that way, the additional cost of the observational update to make it nonlinear is negligible; the main cost is in the computation of an ensemble forecast that is sufficient in size to identify properly the transformation functions. The method has been evaluated using idealized inference experiments, in which several control variables (MLD, NO_3 , ZOO) are estimated from a perfect and local chlorophyll observation. The results show that our simplified scheme is often sufficient to detect and to exploit the nonlinear relationships between observations and estimated variables, thus restoring the control of the system in situations for which linear estimation fails. In many regions of the North Atlantic, a very substantial reduction of error variance has been obtained.

However, these results have been produced for wind errors only, while many other error sources exist in basin scale CPBMs. Before general conclusions can be reached about the controllability of the system or about the least cost effective algorithm, it is necessary that similar studies be attempted for other important sources of errors, like the parameters governing the ecosystem processes, the light forcing, the vertical advection or the horizontal advection and diffusion. In addition, in following this research scenario, one should be aware of possible nonlinear interactions between the error sources: conclusions obtained by considering them separately may no more be valid if they are present altogether.

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Acknowledgements. This work is a contribution to the GMMC Mercator Vert project and the MERSEA project supported by the European Commission under contract SIP3-CT-2003-502885. Calculations were conducted with the support of IDRIS/CNRS.



5

The publication of this article is financed by CNRS-INSU.

References

Barnier, B., Madec, G., Penduff, T., Molines, J. M., Tréguier, A. M., Beckmann, A., Biastoch, A., Boning, C., Dengg, J., Gulev, S., Le Sommer, J., Rémy, E., Talandier, C., Theetten, S., Maltrud, M., and Mc Lean, J.: Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy permitting resolution, *Ocean Dynam.*, 56(5–6), 543–567, 2006.

10

Berline, L., Brankart, J.-M., Brasseur, P., Ourmières, Y., and Verron, J.: Improving the physics of a coupled physical-biogeochemical model of the North Atlantic through data assimilation: impact on the ecosystem, *J. Mar. Syst.*, 64(1–4), 153–172, 2007.

15

Bertino, L., Evensen, G., and Wackernagel, H.: Sequential Data Assimilation Techniques in Oceanography, *International Statistical Reviews*, 71, 223–241, 2003.

Blanke, B. and Delecluse, P.: Variability of the tropical Atlantic ocean simulated by a general circulation model with two different mixed layer physics, *J. Phys. Oceanogr.*, 23, 1363–1388, 1993.

20

Brasseur, P., Gruber, N., Barciela, R., Brander, K., Doron, M., El Moussaoui, A., Hobday, A., Huret, M., Kremer, A.-S., Lehodey, P., Moulin, C., Murtugudde, R., Senina, I., Svendsen, E., and Matear, R.: Integrating biogeochemistry and ecology into ocean data assimilation systems, *Oceanography*, in press, 2009.

25

Conkright, M. E., Antonov, J. I., Baranova, O., Boyer, T. P., Garcia, H. E., Gelfeld, R., Johnson,

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



D., Locarnini, R. A., Murphy, P. P., O'Brien, T. D., Smolyar, I., and Stephens, C.: NOAA Atlas NESDIS 42, WORLD OCEAN DATABASE 2001 Volume 1: Introduction, US Gov. Printing Office, Wash., D.C., 160 pp., 2002.

European Centre for Medium-Range Weather Forecasts, 2002. The ERA-40 Archive.

5 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.

Gruber, N., Doney, S. C., Emerson, S. R., Gilbert, D., Kobayashi, T., Körtzinger, A., Johnson, G. C., Johnson, K. S., Riser, S. C., and Ulloa, O.: The Argo-Oxygen program: a white paper to promote the addition of oxygen sensors to the international Argo float program, http://www-argo.ucsd.edu/o2_white_paper_web.pdf, 2007.

10 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 in ocean models, *Ocean Modell.*, 27, 1–17, doi:10.1016/j.ocemod.2008.10.007, 2009.

Levitus, S., Antonov, J. I., Boyer, T. P., and Stephens, C.: World Ocean Database 1998, National Oceanographic Data Center, Silver Spring, MD, 2001..

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

20 Lévy, M., Krémur, A. S., and Mémery, L.: Description of the LOBSTER biogeochemical model implemented in the OPA system, Internal report IPSL/LOCEAN, 2005b.

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.

25 Madec, G., Delecluse, P., Imbard, M., and Lévy, C.: OPA 8.1 Ocean General Circulation Model reference manual, Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No. 11, 91 pp., 1998.

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

30 Oschlies, A.: Improved Representation of Upper-Ocean Dynamics and Mixed Layer Depths in a Model of the North Atlantic on Switching from Eddy-Permitting to Eddy-Resolving Grid Resolution, *J. Phys. Oceanogr.*, 32, 2277–2298, 2002..

OSD

6, 1289–1332, 2009

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: Numerical Recipes in FORTRAN 77, Cambridge University Press, 992 pp., 1992.

Simon, E. and Bertino, L.: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment,

5 Ocean Sci. Discuss., 6, 617–652, 2009,
<http://www.ocean-sci-discuss.net/6/617/2009/>.

The DRAKKAR Group: Eddy-permitting Ocean Circulation Hindcasts of past decades, CLIVAR Exchanges 42, 12, 8–10, 2007.

OSD

6, 1289–1332, 2009

Controllability of mixing errors

D. Béal et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Controllability of mixing errors

D. Béal et al.

Table 1. Linear vs rank correlation coefficients between variables at 14 stations of the North Atlantic domain, after 1-day ensemble simulations.

Stations	WND/MLD	MLD/TEM	MLD/NO ₃	TEM/PHY	SAL/N03	NO ₃ /PHY
Mauritania (1)	0.83/0.94	-0.88/-0.97	0.87/0.97	0.85/0.96	-0.93/-0.93	-0.81/-0.94
Norway (2)	0.85/0.06	0.98/0.91	0.75/0.37	-0.48/-0.01	0.97/0.93	-0.95/-0.67
New Foundland (3)	0.91/0.63	0.95/0.93	0.80/0.75	-0.79/-0.59	0.99/1.00	-0.96/-0.91
Acores (4)	0.87/0.87	-0.95/-1.00	0.95/0.99	0.99/0.98	-0.97/-0.98	-0.98/-0.99
BATS (5)	0.88/0.91	-0.78/-1.00	0.85/0.97	0.99/0.98	-0.99/-0.99	-0.99/-0.95
Labrador 1 (6)	0.89/0.81	0.79/0.94	0.88/0.99	-0.87/-0.84	0.98/0.98	-0.99/-0.97
Subtropical Gyre (7)	0.72/0.92	-0.83/-0.93	0.32/0.31	0.78/0.62	-0.45/-0.69	-0.84/-0.61
Labrador (8)	0.76/0.69	0.29/0.37	0.89/0.94	-0.31/-0.33	0.98/0.84	-1.00/-0.99
Gulf Stream (9)	0.90/0.91	-0.92/-0.98	0.89/0.4	0.91/0.99	-0.32/-0.17	0.85/0.39
Pomme (10)	0.87/0.93	-0.99/-1.00	0.96/0.99	0.99/1.00	0.05/0.41	-0.93/-0.98
INDIA (11)	0.31/0.48	-1.00/-0.98	0.45/0.41	0.53/0.48	0.93/0.97	-0.99/-0.98
NABE (12)	0.09/0.22	-0.97/-0.94	0.51/0.46	0.34/0.29	0.80/0.83	-0.90/-0.86
Gulf Stream 1 (13)	0.37/0.07	-0.67/-0.64	0.70/0.32	-0.10/-0.20	0.72/0.82	-0.73/-0.65
Gulf Stream 2 (14)	0.22/0.13	-0.98/-1.00	0.97/0.95	0.93/0.96	-0.73/-0.72	-0.97/-0.99

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

I◀

▶I

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Controllability of
mixing errors**

D. Béal et al.

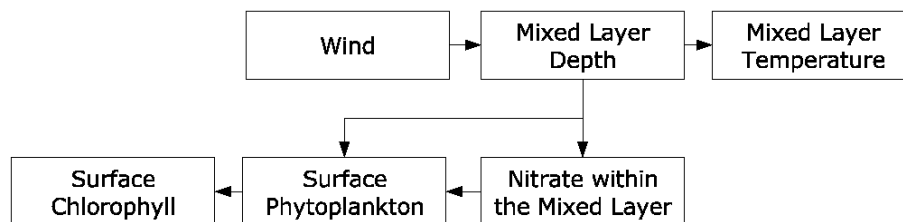


Fig. 1. Illustration of the conceptual transfer function between wind errors and the variables of a coupled physical-biogeochemical model. The arrows show the dominant effect that can be intuitively expected from ocean mixed layer and ecosystem dynamics.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Controllability of
mixing errors

D. Béal et al.

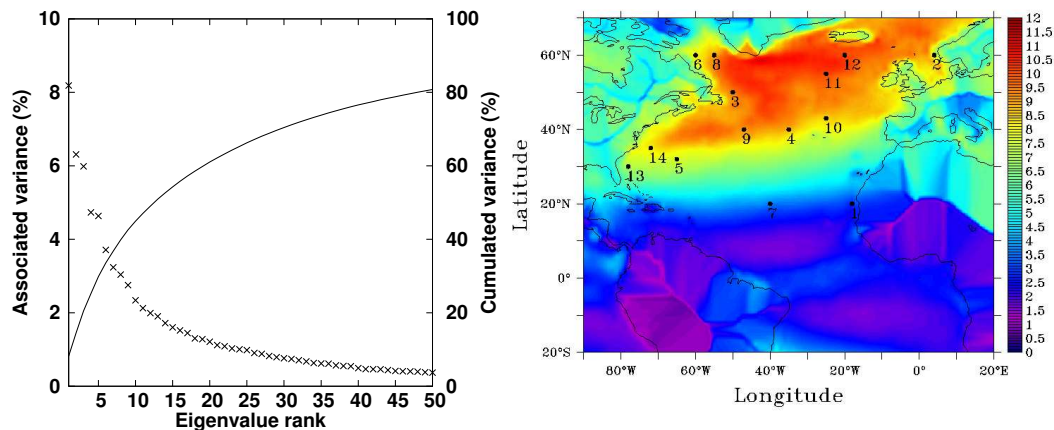


Fig. 2. (left) Percentage of explained variance (left axis) and cumulated variance (right axis) for the first 50 EOFs computed from the variability of the ERA40 1985–2000 wind archives. (right) Wind standard deviation (in N/m^2) calculated over the used archives.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Controllability of mixing errors

D. Béal et al.

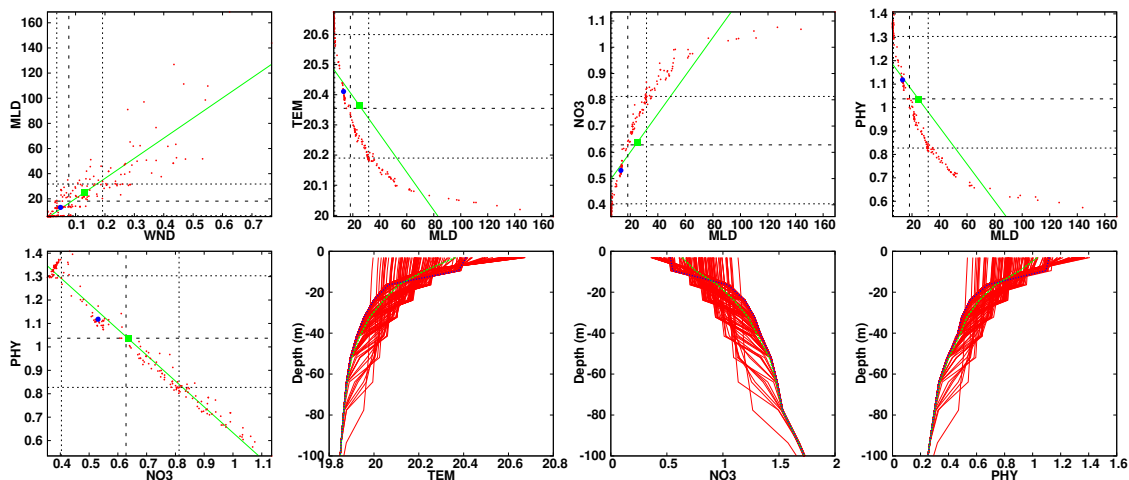


Fig. 3. Scatterplots of 1-day ensemble forecasts at BATS (65° W/ 32° N) station: the red points correspond to the 200 ensemble members; the blue point corresponds to the reference (unperturbed) run; the green square is the ensemble mean; the green line represents the linear regression of the ensemble; the black dotted lines indicate the quartiles of the distribution. Vertical profiles: the red lines correspond to the 200 ensemble members, the black line is the profile of the reference run, the green line is the mean profile over the members, blue - mean plus or minus the standard deviation over the member.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Controllability of mixing errors

D. Béal et al.

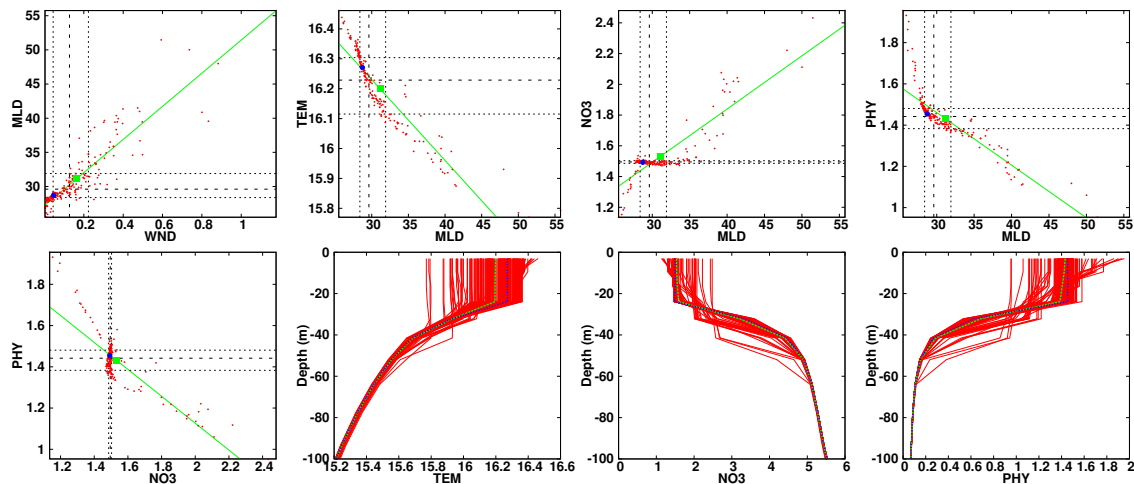


Fig. 4. Same as Fig. 3 but for the so-called Gulf Stream (47° W/ 40° N) station.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Controllability of mixing errors

D. Béal et al.

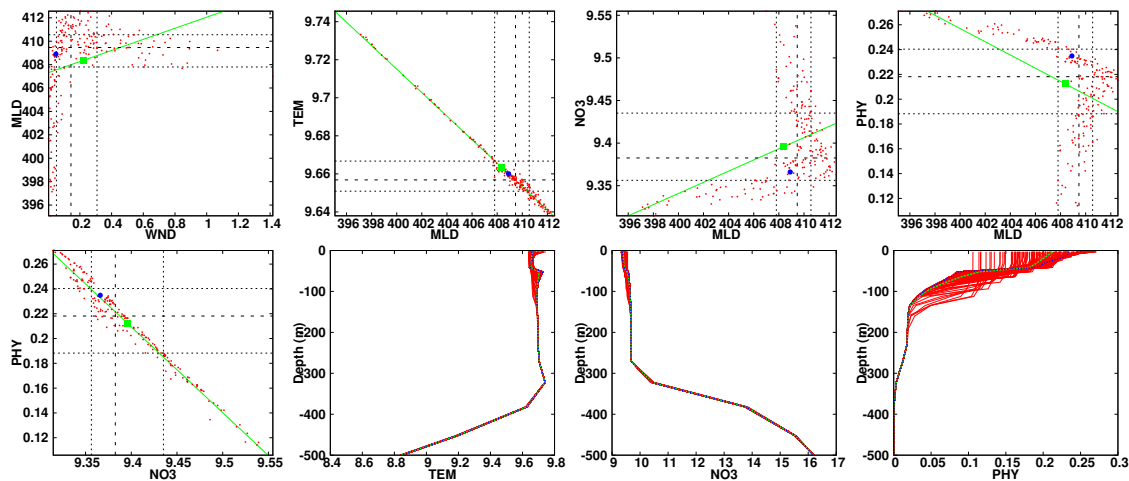


Fig. 5. Same as Fig. 3 but for INDIA (25° W/ 55° N) station

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

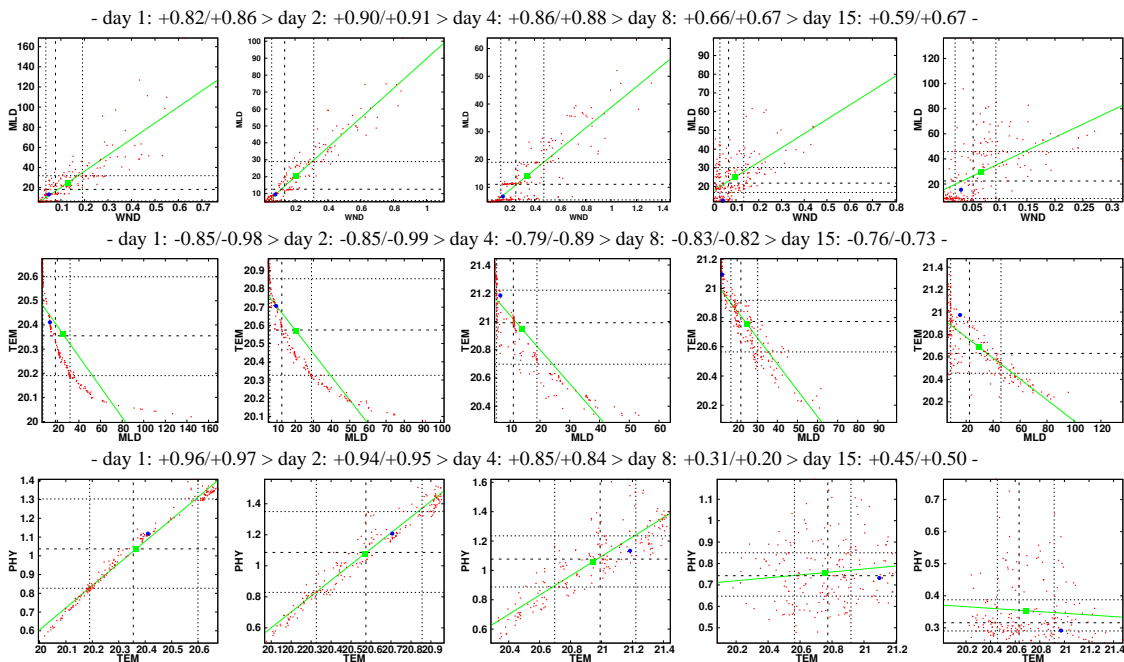


Fig. 6. Scatterplots of ensemble forecasts at BATS station ($65^{\circ}\text{W}/32^{\circ}\text{N}$) after 1, 2, 4, 8 and 15 days (from left to right): WND/MLD (top line), MLD/TEM (central line) and TEM/PHY (bottom line) relationships. Similar colour code as in Fig. 3.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

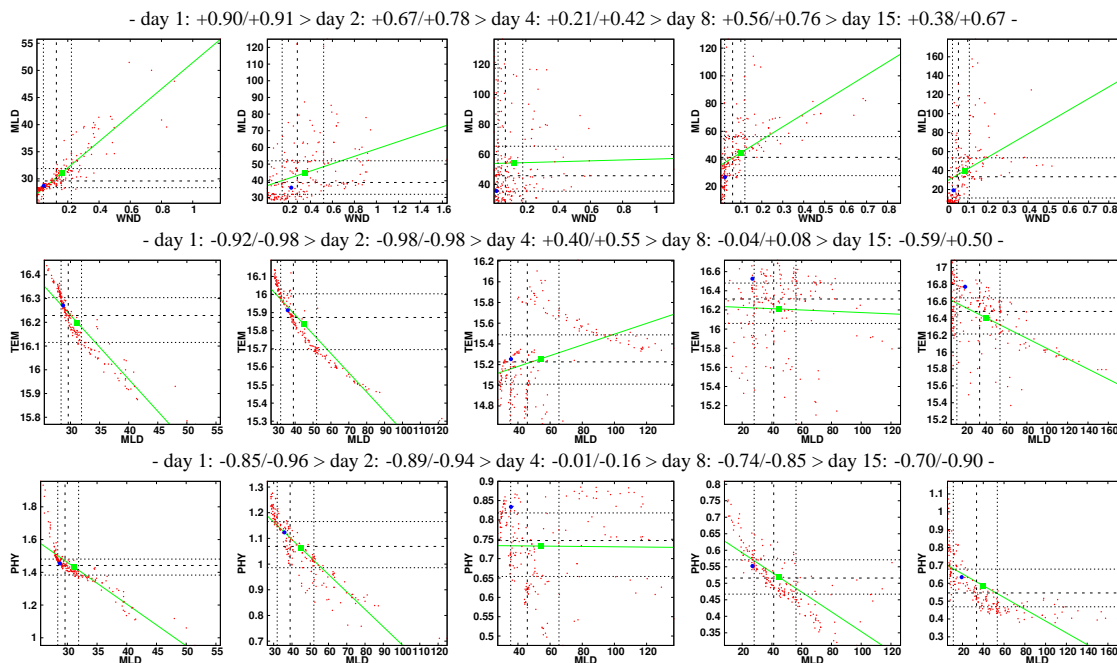



Fig. 7. Scatterplots of ensemble forecasts at Gulf Stream station (47° W/ 40° N) after 1, 2, 4, 8 and 15 days (from left to right): WND/MLD (top line), MLD/TEM (central line) and MLD/PHY (bottom line) relationships. Similar colour code as in Fig. 3.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[◀](#)
[▶](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Controllability of mixing errors

D. Béal et al.

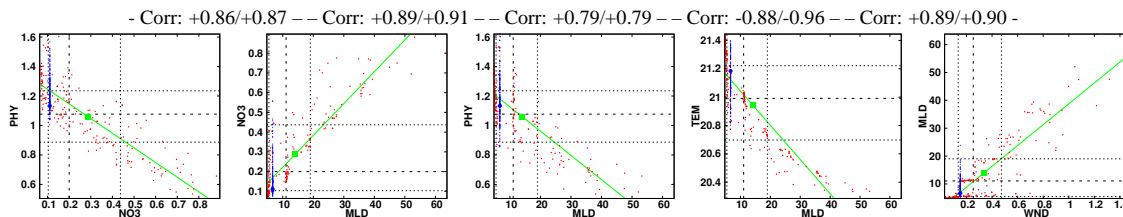


Fig. 8. Scatterplots of ensemble forecasts at BATS (65° W/ 32° N) station after 4 days: the red points correspond to the 200 ensemble members; the big blue point corresponds to the reference (unperturbed) run; the green square is the ensemble mean; the green line represents the linear regression of the ensemble; the blue points indicate the linear update of members following the green regression line.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Controllability of mixing errors

D. Béal et al.

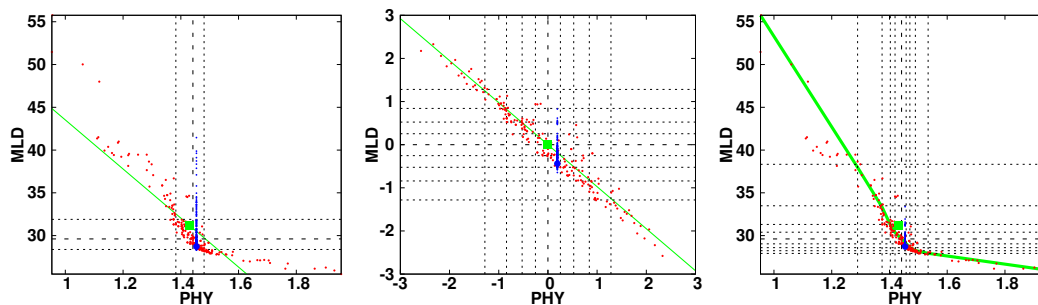


Fig. 9. Observational update at BATS station (65° W/ 32° N) using a perfect phytoplankton observation. The figure shows the 1-day forecast ensemble (red dots), with mean (green square) and linear regression line (thin green line), the reference simulation (large blue dot) that gives the PHY observation and the update ensemble (blue dots). The left panel illustrates a linear observational update performed in the original state space. In the middle panel the linear observational update is performed in a transformed state space (by anamorphosis). In the right panel the solution showed in the middle panel is transformed back into the original state space. The linear regression line of the middle panel (thin green line) transforms into the thick green line of the right panel. Dashed lines are medians, and dotted lines are percentiles (quartiles in the left panel and deciles in the other panels).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



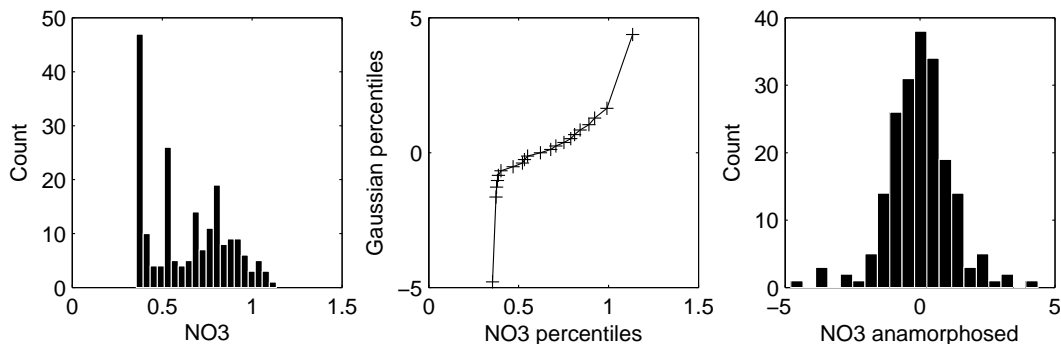


Fig. 10. Illustration of the anamorphosis transformation for nitrate at BATS station. The left panel shows the histogram of nitrate values in the 200-members 1-day forecast ensemble; the middle panel shows the piecewise linear change of variable mapping the nitrate percentiles to the $\mathcal{N}(0, 1)$ percentiles and the right panel shows the histogram of the transformed variable.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Controllability of mixing errors

D. Béal et al.

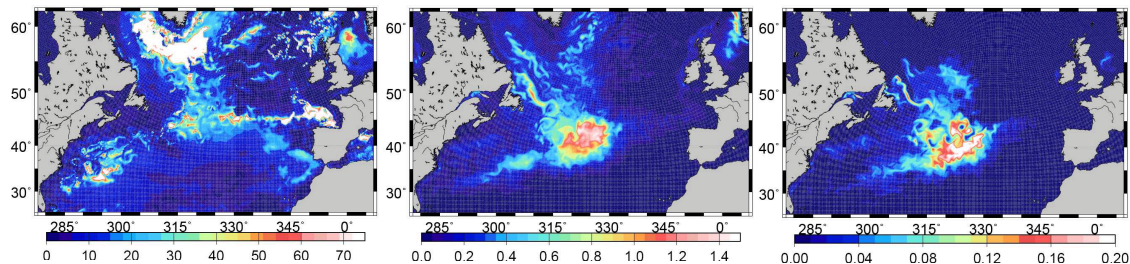


Fig. 11. Standard deviation of the 1-day forecast ensemble for the mixed layer depth (left panel), nitrate (middle panel) and zooplankton (right panel) concentrations.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Controllability of
mixing errors**

D. Béal et al.

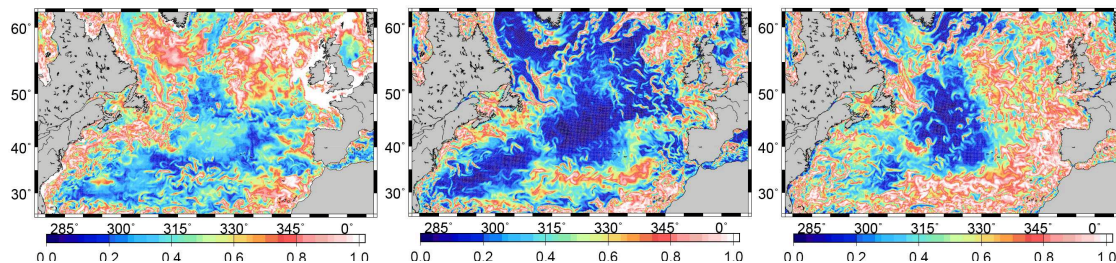


Fig. 12. Ratio of the updated ensemble standard deviation to the forecast ensemble standard deviation (shown in Fig. 11), as obtained using the linear observational update.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Controllability of
mixing errors**

D. Béal et al.

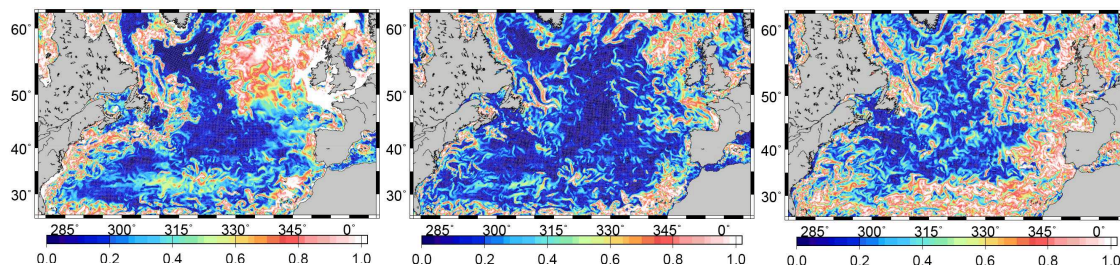


Fig. 13. Ratio of the updated ensemble standard deviation to the forecast ensemble standard deviation (shown in Fig. 11), as obtained using the nonlinear observational update (linear observational update in the transformed space).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**Controllability of
mixing errors**

D. Béal et al.

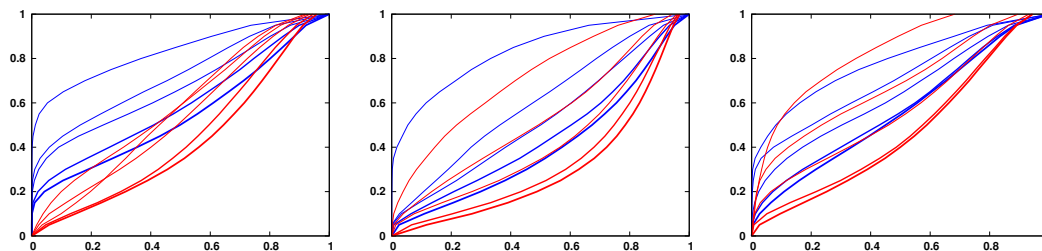


Fig. 14. Fraction of the domain (X-axis) for which the error reduction factor by the ensemble observational update (as illustrated for instance in Figs. 12 and 13 for day 1) is lower than a given value (Y-axis). The result is shown for the linear observational update (blue curves) and for the anamorphosis nonlinear observational update (red curves), at day 1, 2, 4, 8 and 15 (from thick curves to thin curves). The estimated variable is mixed layer depth (left panel), nitrate concentration (middle panel) or zooplankton concentration (right panel).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)