Ocean Sci. Discuss., 6, C305–C309, 2009
www.ocean-sci-discuss.net/6/C305/2009/

Ocean Science Discussions

Interactive comment on "Controllability of mixing errors in a coupled physical biogeochemical model of the North Atlantic: a nonlinear study using anamorphosis" *by* D. Béal et al.

Anonymous Referee #1

Received and published: 3 August 2009

Controllability of mixing errors in a coupled physical biogeochemical model of the North Atlantic: a nonlinear study using anamorphosis by D. Béal, P. Brasseur, J.-M. Brankart, Y. Ourmières, and J. Verron

The manuscript discusses the problem of controllability and data assimilation of non-gaussian distributed data. A biogeochemical model coupled to a physical model is used to study the relationship between variables such as temperature, mixed layer depth, phytoplankton, nutriments and zooplankton. The limitations of a linear update and how an analysis with an anamorphosis transform can improve the results are shown. The

manuscript is in general well written and clear. It also provides a good illustrations of the data assimilation problem with non-gaussian distributed variables.

However it represents only a small step compared to the general problem since only model results (without errors) were assimilated and, more importantly, the model is not restarted from the assimilated solution. Also the results depend strongly on the way the perturbations are chosen (which is indeed acknowledged by the authors). One crucial parameter is the decorrelation time-scale of the wind perturbations which is chosen to be 4 days. This begs the question if the time scale discussed in the conclusion over which various parameters tend to decorrelate is not simply a product of the time scale chosen for the wind stress and how robust the results are in general.

But I realize that it is complex problem. The manuscript is a timely contribution to this issue as coupled models with data assimilation start to emerge. After major revision, I believe that the manuscript can become suitable for publication in Ocean Science.

Major comments:

* section 2.2: EOFs are used to generate an ensemble of wind perturbations. It appears that the authors used the time-variance of the wind stress as the error variance of the wind field which seems excessive to me. This is equivalent to assume that the averaged wind is as accurate as instantaneous wind fields. Since the main objective of this paper is to assess the non-linear and the non-gaussian model response to wind errors, it is important to choose wind perturbations with a realistic amplitude. One could anticipate that non-linear saturation effects would be less of a problem if smaller and more realistic wind perturbations are used.

Another issue with the perturbation scheme is the chosen time-scale of 4 days. The first EOFs represent generally slow and large-scale processes. The seasonal cycle is for example often the most dominant EOF. However, the EOFs corresponding to the seasonal cycle is perturbed also with a time-scale of 4 days which effectively means that the model solution can switch from a summer to a wind regime over only this time scale. Later in the manuscript, the authors were able to provide a clear explanation of the model error propagation over 1 day, but noticed that over longer time scales (2 to 15) the variables tend to decorrelate. It seems to me that this negative resolution is simply a result of the short decorrelation time scale of the wind forcing and its large amplitude. Due to the highly variable error forcing, the ensemble members are quite far away from an equilibrium dynamics (and tend to move further away the equilibrium as the simulation goes on). I'm wondering how robust the presented results are relative to the time-scale of the wind perturbation (and their amplitude).

* The model state can be viewed as a time integrated response on the wind stress (and other fluxes). If the wind field suddenly changes, the instantaneous wind might indeed not be related to other instantaneous variables (which includes the effect of past winds). It might thus be interesting to look not only the relationship between instantaneous wind and the model variable but also to some integrated quantities such as the time-integrated surface turbulent flux. This can still be useful in a data assimilation context because an error in the time integrated turbulent flux can be corrected in a approach similar to the incremental analysis update approach (Bloom et al. 1996, MWR).

* The authors test their scheme in a configuration which is similar to a twin experiment: they assimilate observations extracted from an another model run to assess the impact on the model solution. However, here the observations are taken from the unperturbed reference simulation which is to me a questionable choice. The ensemble tends to be centered around the observations and per construction the ensemble average would be close to the "true" solution. It would be better to choose one ensemble member as the truth (and not using it to derive the statistics such as covariance and the anamorphosis transform).

* section 4.3: The authors assess the benefit of the transformation compared to the linear approach by calculating the standard deviation of the updated ensemble. The ensemble spread of the updated ensemble is indeed a measure of uncertainty but it is in this case too closely related to the update approach which the authors want to

validate. A method which over-estimates (in a unrealistic way) the relationship between observed and non-observed variables would lead to a lower spread than a method which uses a more realistic assumption. An extreme case to clarify this point would be an "assimilation scheme" which updates the MLD by MLD = PHY/2. This scheme would have no spread at all since PHY observations are assumed to be perfect. But this is clearly a very bad assimilation scheme. Why the authors do not look to the ensemble RMS error between the MLD and the "true" MLD corresponding to the observed PHY? Idem for nitrate and zooplankton.

Minor comments.

* section 2.1.2: what is the barotropic time step?

* section 2.2: From the manuscript it is not clear if the authors combined the u- and v- components in the EOF calculation and perturbation scheme or if the components were used independently.

* Please define acronyms such as BATS and INDIA

* It would be useful to remind the reader from time to time that wind refers to wind stress and not wind speed.

* section 3.3: You describe how an update would affect successively PHY -> MLD -> Temperature. Is this chain of interaction only a conceptual view or do you propose actually to perform the assimilation in sub-steps? The standard scheme would correct directly temperature from PHY (using their covariance and possibly involving an anamorphosis). In the case of a gaussian distributed variable, an assimilation with sub-steps would lead to an suboptimal analysis since some processes link PHY directly to temperature without involving MLD (e.g. mortality).

* section 4.2.1: the inverse anamorphosis transform is defined over [y_1, y_p]. However, the linear analysis update may produce a value outside of this range (in some extreme cases). What would be the strategy to transform this value back to a physical

one?

* section 4.2.2: The back-transformed linear regression of the anamorphosed variable is referred to as the "true nonlinear regression". Unless this is a mathematical term already used previously, I would avoid calling it this way. There is not guarantee that the anamorphosed variables can be linked by a linear regression when the anamorphosis is performed on the variables independently (or if just one variable is transformed).

* Figure 10: It would be clearer if the anamorphosis in figure 10 would be derived with the variable PHY to complement figure 9 showing the relationship between MLD and PHY.

* page 1309: "... the proposed solution is in this case very close to optimality": As the authors know, different optimality criteria are used in data assimilation (leading to different solutions for non-gaussian distributed variables). To which definition of optimality the authors refer here?

* page 1309: " the regression line is nonlinear and non-monotonous (... Figure 5)". This is a bit confusing at first since figure 5 does not show a nonlinear and non-monotonous regression function (only a linear one). It would be clearer if the authors refer to the general tendency in those scatter plots which have those characteristics. Also "a *non-linear* regression *line*" is a contradiction. Consider to change to wording here.

* conclusions: "Before general conclusions can be reached about the controllability of the system or about the least cost effective algorithm, ...": should that read the "least costly algorithm" or the "most cost effective algorithm" ?

Interactive comment on Ocean Sci. Discuss., 6, 1289, 2009.