Ocean Sci. Discuss., 6, C151–C161, 2009
www.ocean-sci-discuss.net/6/C151/2009/

Ocean Science Discussions

Interactive comment on "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" *by* E. Simon and L. Bertino

E. Simon

ehouarn.simon@nersc.no

Received and published: 15 June 2009

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

E. Simon and L. Bertino

Answers to P. Brasseur (Referee)

1- General comments

1 *The review of Kalman-based assimilation studies into coupled physical-biological ocean models presented in Section 1 (p 620) is interesting although probably not exhaustive. Instead of just enumerating the applications published in the literature, the authors should provide a brief discussion about the main lessons learned from these previous studies, and how the proposed anamorphosis approach articulates with them (especially Natvik et al. who considered a very similar modelling framework). In addition, alternative methods such as particle filters (SIR, Losa et al., 2003) exist for non-linear data assimilation: how would they compare regarding, e.g., computational efficiency ?*

The review has been rewritten in that way. The introduction has been heavily modified.

As stated by the referee, our modelling framework is very close to the one of Natvik and Evensen. We both apply an EnKF to assimilate surface chlorophyll data in a North Atlantic configuration. Nevertheless there are important differences. First Natvik and Evensen assimilated SeaWiFS satellite data while we performed twin experiments. They recognized that they were not able to check the improvement/degradation of non-observed components of the solution, as they did not access to the true solution. In this present work, we realized twin experiments to check the performances of the algorithm on observed and non-observed variables. We think that it has to be done before dealing with real observations. Secondly, their experiments were too short (only 2 months during the Spring bloom) and did "not allow for investigating any long term trends in the model". We fill this lack by performing one year experiments. These works allowed to highlight several assimilation biases due to the Kalman filter method-

OSD

6, C151–C161, 2009

Interactive Comment

ology which would have not been detected with experiments realized only during two months of the Spring bloom.

Finally concerning the non-linear data assimilation methods, some comments have been added. From a theoretical point of view, the particle filters seem to be very interesting to assimilate data in ecosystem models (parameters optimization, state estimation) as they are built for any probability density function. As stated by Losa et al, the SIR filter appears to be a variance minimizing scheme for any probability density function, which is not the case for the EnKF applied to non-linear models (due to the Gaussian assumption made on the distribution of the variables). Nevertheless the application of the SIR filter in operational framework seems to be unrealistic. Experiments in Losa et al used a 1D dynamical system of 9 variables (ecosystem model) and required 1000 members for estimating relevant parameters and the state vector (several ten of variables). Realistic experiments requires to estimate millions of variables which does not allowed to use (yet) such big ensembles.

2 *The description of the procedure for constructing the anamorphosis functions (sections 2.2 and 3.3) remains a little obscure and incomplete (especially for step 1, page 625). Some more practical details should be given with a view to make the procedure reproducible in similar setups, and to facilitate the interpretation of results (especially the spectacular improvement obtained with the anamorphosis method in ANA experiment). One could wonder what are the key mechanisms responsible for this improvement, and the reader is left a bit hanging on that point. In ECO experiment, it might be interesting to quantify how frequently the statistical analysis generates negative (or unrealistically high) concentration values.*

An appendix describing the construction of the empirical anamorphosis function (step 1) has been added. Comments on the interpolation step (step 2) have also been added in the section 3.3 ("Construction of the monovariate anamorphosis functions").

The spectacular improvements due to the introduction of anamorphosis functions lie in the combination of the respect of assumptions done in the theory of the estimation by Kalman filtering (which produces better analysis) and the strong non-linear behavior of the model which does not allow to compensate the assimilation bias (the model seems rather to amplify them).

We quantified the frequency of appearance of negative values.Evolution with time of the rate of null values in the post-processed analysis ensemble for silicate (SIL) and diatoms (DIA) are shown on figure **??**.

We note that the truncation affects strongly the concentration of SIL and DIA. From the beginning of the Fall to January the rate of truncated values increases from 2% to 17% for DIA. During this period, the concentration of DIA is low and decreases towards 0, leading to an increase of the number of analyzed negative values. At the end of the winter, the concentration of DIA increases. The number of analyzed negative values decreases, the mean of the forecast ensemble moving away from 0.

The case of SIL is more complicated. From Fall to Winter, the remineralization of DIA leads to an increase of the concentration of silicate. So according to the previous explanation, we should not observe such synchronization between the curves of SIL and DIA, it means an increase of the number of 0 in the analyzed ensemble during the cold period. Following the notations of Evensen (2006), the analysis equation of the EnKF can read:

$$\begin{aligned} A^a &= A^f \left[I + (HA'^f)^T((HA'^f)(HA'^f)^T + (N-1)R)^{-1}(D - (HA^f)) \right] \\ A^a &= A^f X_H \end{aligned}$$

with D the perturbed observations, A the ensemble matrix (forecast or analysis) and $A' = [(A_i - \bar{A})_{i=1:N}]$ the ensemble anomalies with $\bar{A}$ the mean of the ensemble. We note that the corrections to the forecast ensemble brought by the assimilation X_H depend only on the variables of the ensemble related to the observation by H (and the observation errors). In our particular case, the corrections depend only of concentrations of DIA and FLA. The combination of the presence of numerous low values (close to 0) of DIA in the the forecast ensemble (increasing the weight in the correction of the few high values of DIA) and the negative correlations between DIA and SIL (and more generally nutrients and phytoplankton) in cold period leads to this generation of negative values in the SIL analyzed ensemble. The increase of negative values in the phytoplankton ensemble due to the dynamics of the model in winter leads to an increase of negative values in the nutrients analyzed ensemble.

A sentence has been added at the end of the section 4.1.

3 *The perturbed atmospheric fields used to generate the initial ensemble are primarily affecting the physical state of the coupled system. The perturbations then cascade into the biological components. However the assimilation scheme does only correct the biological state of the system, leaving the circulation component unchanged. Why not simply consider a fully coupled physical/biological state vector, or alternatively perturb the biological model components without perturbing the physics? The strategy chosen in this study should be better justified.*

As the system is intended to be the ecosystem module of the TOPAZ system, the strategy to generate the ensemble is the same that the one adopted in TOPAZ. The perturbations in the atmospheric fields will lead to perturbations on the biological components through the generation of perturbed physical states. Assimilating data only in the ecosystem component of the coupled model does not allow correction in the physical component, and so does not allow corrections in the full error subspace of the coupled model.

Nevertheless we think that is not an important issue as far as we are dealing with twin experiments in a coarse resolution model. In such context, the bias due to the errors in the physical component is relatively low: the main structures (the

OSD

6, C151–C161, 2009

Interactive Comment

Gulf Stream for example) are similar in the ensemble and in the reference simulation. It allows us to focus only on the improvement of the ecosystem component of the coupled system, free of biases.

This is not longer true when working in a realistic framework. One strategy consists to improve the physics by assimilating classical physical data (SST, SSH, etc..) in the physical model, and then assimilating chlorophyll data in the ecosystem model. Such strategy is easy to process when having already a physical operational assimilating system. This is the case with the TOPAZ system in NERSC, justifying the strategy adopted in our twin experiments. An improvement of this strategy is the direct perturbations of the ecosystem component. The addition of these perturbations is considered for the future development of TOPAZ-ECO.

Finally considering a fully coupled physical/biological state vector is a challenging open issue. Apart from the fact that the time scales of the physical and biological dynamics can be strongly different (what will be the information contained in the full coupled error covariance matrix?), it will need also major improvements in the modelling of the coupled system, notably in the parameterizations of phenomenon coupling the physics and the biology. Ecosystem models are too simple in comparison with physical models and present high bias and uncertainties. Correcting the physics by observing the biology can lead to a transfer of these bias and uncertainties (which can be amplified by the assimilation if the method used is not suitable for the properties of the model) to the physical component of the coupled model. It is probably too ambitious at present time. Furthermore coupled physical/ecosystem models are expensive. The high resolution required to be able to consider a multivariate biophysical analysis is not easily tractable for operational systems covering large areas (Arctic and North Atlantic oceans in Topaz).

Justifications of our strategy have been added in the revised manuscript section 3.2. Furthermore we would like to borrow the word "cascade" from the reviewer in the revised manuscript.

4 *The observation operator H is not precisely defined. Please explain how the surface chlorophyll concentration is diagnosed by the model, and clarify the relationship between observations and the two phytoplankton compartments. It might be interesting to compare the anamorphosis functions constructed for diatoms, flagellates and chlorophyll.*.

The observation operator H has been defined in section 3.2. The surface chlorophyll concentration is diagnosed by a linear combination of the diatoms and flagellates concentrations (equation (9) in the revised manuscript).

Anamorphosis functions for chlorophyll, diatoms, flagellates and silicate (nutrient) have been plotted and briefly discussed in the revised manuscript.

5 *The word "limitations" is repeatedly used in the text, with apparently different meanings from place to place and sometimes misleading interpretations. For instance, p 621 "Such models present important practical and theoretical limitations for the application of data assimilation methods . . .": in this case, the "limitations" are probably more in the assimilation methods than in the models! Similarly, "physical bounds" (e.g., legend of Table 1, p 622 etc.) is misleading as long as biological quantities are concerned. I suggest to adopt less sloppy terminologies and check the language throughout the text.*

Done.

6 *The coupled model simulations are unrealistic in the equatorial Atlantic Ocean. Since the assimilation procedure is not activated along the southern boundary, the figures showing horizontal distributions (e.g. Figs. 9, 10, 11) could be redrawn to focus on the regions (North Atlantic and Arctic regions) where the assimilation results are significant.*

OSD
6, C151–C161, 2009

Interactive Comment

Done: the 2D plots of the RMS errors do not show anymore the equatorial Atlantic Ocean.

2- Minor comments

- *p 619, l.2: "in term of" → "in terms of"*

  Done.

- *p 619, l.8: "ecosystem" → "ecosystems"*

  Done.

- *p 919, l.21: the error on ocean colour (30%) seems to be excessively optimistic w.r.t. estimates from the literature (e.g. Ballabrera et al., 2003)*

  We agree that the value of 30% for the error can appear to be optimistic. Nevertheless it is closed to the error estimate for the SeaWiFS chlorophyll data. The comparison by Gregg and Casey (2004) of SeaWiFS chlorophyll data with independent in-situ data led to a global mean error of 31%, with important variations depending of the area. We highlighted this average value keeping in mind that the error can be much more important in some area. The sentence has been rewritten.

  As we have not introduced the data assimilation concept before this paragraph and we do not explain the representativity error, we prefer to avoid to refer to Ballabrera-Poy et al, their study dealing with the quantification of this error. We agree that the problem of scale representativeness between the model and the observations is a challenging issue, but not really in the scope of this present paper.

  Gregg W.W. and Casey N.W. : Global and regional evaluation of the SeaWiFS chlorophyll data set, Remote Sensing of Environment, 93 ,463-479, 2004.

OSD

6, C151–C161, 2009

Interactive Comment

- *p 619, l.27 and l.28 are unclear. Please rephrase*

  Done.

- *p 620, l.10 : "on ecosystem model mainly this last decade" → "to ecosystem models mainly during this last decade" ?*

  Done.

- *p 621, l.18 : "develops" → "developed"*

  Done.

- *p 622, l.7 : "conclusion" → "conclusions"*

  Done.

- *p 623, l.9 : "variables" → "variable"*

  Done.

- *p 623, eq.4 : the product should be defined*

  Done.

- *p 623, eq 5: it might be useful to specify which forecast error covariance matrix (in the transformed space ?) is used to compute the "transformed" Kalman gain*

  Done.

- *p 624, l.7 : "in general case" → "in general"*

  Done.

- *p 624, l.20 : "check" → "satisfy" ?*

  No. We say that the size of the vectors are too large to allow us to check practically their multi-Gaussianity (χ_2-test).

OSD

6, C151–C161, 2009

Interactive Comment

- *p 626, l.3 : what do you mean exactly by continuous distributions ?*

  A probability distribution function is continuous if its cumulative distribution function is continuous.

- *p 627, l.18 : "diatom" → "diatoms*

  Done.

- *p 628, l.5 : "all the component" →"all the components"*

  Done.

- *p 628, l.29 : "Eq. 8" → "Eq. 3" ?*

  The reference deals with the way to build the synthetic observations from the true state (reference simulation) as stated in Eq. 8. Eq. 3 refers to the construction of the transformed observation (built in our experimental case following Eq. 8) by the anamorphosis function.

- *p 632, l.17: "longer" → "higher" ?*

  Done.

- *p 633, l.6 : "on one only point" → "on only one grid point" ?*

  Done.

- *p 633, l.17 : unclear sentence, please rephrase*

  Done.

Interactive comment on Ocean Sci. Discuss., 6, 617, 2009.

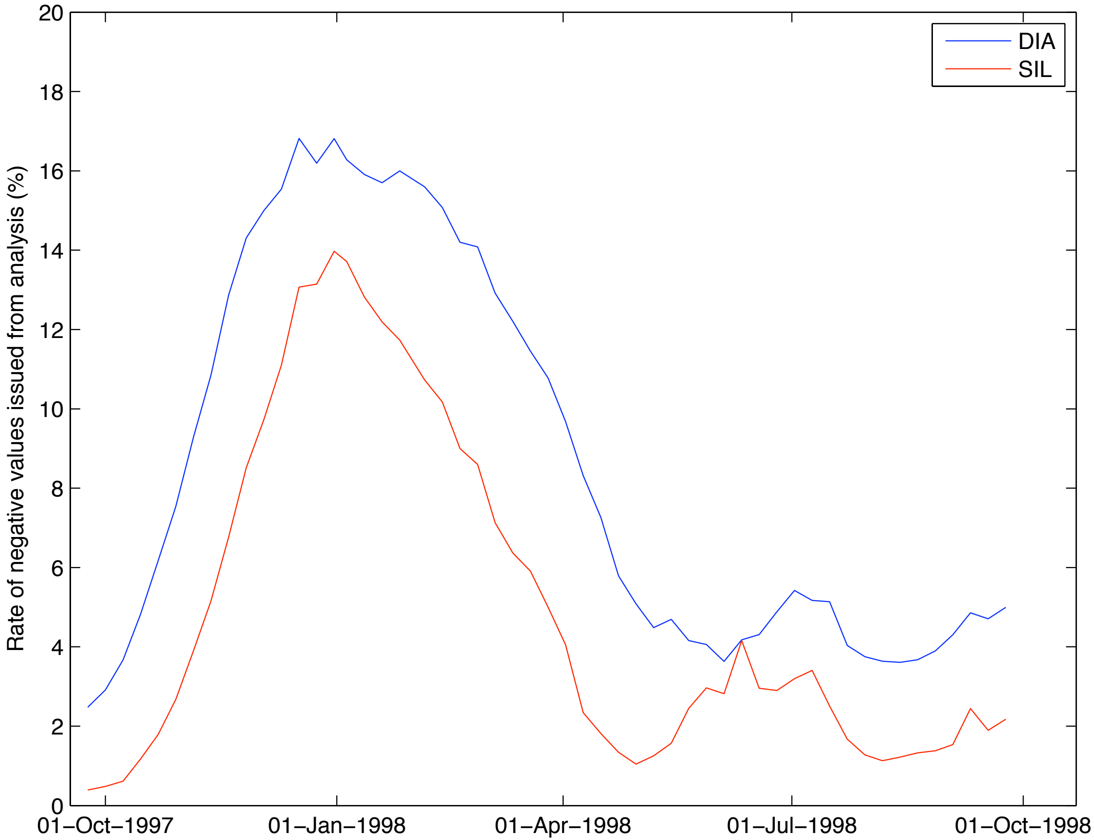

Fig. 1.

OSD

6, C151–C161, 2009

Interactive Comment