Ocean Sci. Discuss., 9, C775–C785, 2012
www.ocean-sci-discuss.net/9/C775/2012/
© Author(s) 2012. This work is distributed under
the Creative Commons Attribute 3.0 License.

Ocean Science Discussions

OSD

9, C775–C785, 2012

Interactive Comment

Interactive comment on "Toward a multivariate reanalysis of the North Atlantic ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data" *by* C. Fontana et al.

C. Fontana et al.

clement.fontana@obs-vlfr.fr

Received and published: 27 July 2012

1 Major comments

A Time-averaging of the simulations

When an assimilation step is performed, the analysis becomes the first time step from which the simulation restarts. Then the model evolves freely until the next assimilation step. The time-averaging consists in computing the mean of every time steps for the period considered. For example, we have 60 time-steps per day, thus 60*60=3600

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

time-steps for a period of 60 days. In the time-averaging procedure, there are 7 analysis, and 3593 time-steps of free evolving simulation.

Text was added by the end of paragraph 3.1

B Description of the Gaussian anamorphosis

Yes, the reviewer is right. The description of the anamorphic transformations was not specific enough in section 3.2. We have added the following sentences at the end of the 2nd paragraph to provide the missing information:

"The same ensemble of 210 members is also used in the non-linear run to build the anamorphosis transformation locally in space and time. More precisely, each time observations are assimilated, a specific non-linear transformation is computed for each model grid cell and for each model variables from the histogram of 210 values that are associated to this day of the year. Moreover, as in Doron et al. (2011), the tails of the distribution are defined by assuming zero probability to values out of the range of the sample."

Second, it is also true that the parameterization of the observation error in the nonlinear run was missing in section 3.2. We have added the following sentences at the end of the 3rd paragraph to provide the missing information:

"In the non-linear run (with anamorphic transformations), uncertainties in the observations cannot be specified exactly in the same way since they must be assumed Gaussian for the non-linearly transformed variables rather than for the original variables. Nevertheless, to give a similar importance to each observation in the two assimilation runs, we compute the observation error standard deviation for the transformed variable by multiplying the original observation error standard deviation (i.e. the 30% of the observed chla concentration) by the local slope of the non-linear transformation (which we approximate in practice by a finite difference over one standard deviation)."

C Tails of the anamorphosis functions Yes, the reviewer is right, in this study (as in

OSD

9, C775–C785, 2012

Interactive Comment

Doron et al., 2011), the tails of the anamorphic functions are defined by truncating the values out of the range of the ensemble. This is a very simple choice, which avoids producing estimated values outside of the range effectively explored by the model simulation. The direct consequence is indeed that, wherever the model simulation is biased as compared to the observations, the observations may fall outside of this range. Then, the observation (more precisely, the expected value of the observation) is viewed as impossible by the forecast probability distribution, and the estimation of the observed variable can only stay away from the observation inside the range defined by the ensemble. We thus agree with the reviewer that, in this case, with a parameterization of the tails of the anamorphic transformations, the estimation can be brought closer to the observations. But in our view, relying on an arbitrary parameterization of the tails of the univariate distributions is not a good way of dealing with model biases. What about the corrections applied to the unobserved variables? How confident can we be on the multivariate correlation structure in the peripheric regions of the state space that have not been explored by the ensemble? Either the observations are very often outside the range of the ensemble, and something else must be done to correct for the bias, because it is inefficient to produce a large ensemble which usually misses the observations, and then to rely on Gaussian tails to compute the corrections (the tails of the distributions must correspond to improbable values). Or it does not happen very often, and it may be safer to avoid any kind of extrapolation outside the range of the ensemble. In our application, we are clearly in the second case: the dispersion of the ensemble is most often large enough to include the assimilated observations (except in some regions of the subtropical gyre). This is because the ensemble is built using the seasonal and interannual variability of a free model simulation (which is large in ecosystem models, especially during the spring bloom), so that we have been able to tune the ensemble dispersion using the time window of the seasonal variability. With a two-month period around the assimilation date, the ensemble dispersion was most often large enough to include the assimilated observations. On the other hand, we agree that we do not assimilate the same observational information in the linear and in the

OSD
9, C775–C785, 2012

CC BY

nonlinear runs, but this is intrinsic to using anamorphosis transformations. This does not only occur if the expected value of an observation falls in a zero-probability range. In the linear run, the observation error probability distribution is assumed Gaussian for the original variable, and in the non-linear run, it is assumed Gaussian for the transformed variable. There is no way of making these two kind of observational information equivalent. For instance, if we assume that we use an exponential parameterization of the tails of the distribution: exp[-a(x-xmax)], then the transformation becomes perfectly bijective whatever a>0. But, as the parameter a becomes larger and larger, the observation probability distribution for an observation y>xmax (i.e. the backward transform of thr Gaussian distribution that must be assumed for the transformed variable) becomes more and more distorted, and the resulting estimation of x becomes closer and closer to xmax. Our parameterization of the tail of the distribution corresponds to the limit a->infinity (non-bijective for x>xmax), and the estimated value cannot be larger than xmax anymore. But this is just a limit case, there is no discontinuity in the behaviour (i.e. a sudden change in the assimilated observation) as soon as the transformation becomes non-bijective.

In summary, since the main purpose of this paper is not to discuss assimilation schemes, and even less anamorphosis transformations (we just rely on existing schemes), we tried to summarizethis point of view on the parameterization of the tails of the probability distributions in a new paragraph added in section 3.2:

"Concerning the parameterization of the tails of the anamorphic transformations (outside the range of the available ensemble), we make the simple assumption of a zero forecast probability in these regions of the state space. The direct consequence is that, even if an observation falls in these peripheric regions, our estimation of the observed variable cannot get close to the observation, because it is bounded to stay inside the range defined by the ensemble. However, in our application, this does not occur very often, because the ensemble is built using the seasonal and interannual variability of the free model simulation (which is large during the bloom event), so that the dispersion of the ensemble could easily be tuned to be large enough to include most of the assimilated observations (except in some regions of the subtropical gyre). We thus preferred safety by avoiding any kind of extrapolation outside the range of values effectively explored by the model simulation. More sophisticated assumptions about the tails of the distribution (e.g. Gaussian tails) can be found in the works of Bertino et al. (2003) and Simon and Bertino (2009)."

Specific comments

p. 1891, l. 12-14 "this issue partly explains why most of the pioneer studies dealing with ocean color data assimilation were first carried out using pseudo-data rather than with real data": Twin experiments are not necessarily performed due to the poor quality of the data. They offer a perfect framework for the assessment of methodological developments: the true state and most of the error sources are known which is not the case when assimilating real observations. The sentence could be rewritten.

We agree with this comment and this is why the word "partly" was used.

p. 1891, l. 19 "making the traditional assimilation framework inappropriate to develop this applications (Doron et al., 2011)": A reference to Bertino et al. (2003) might be relevant as well.

Reference was added

p. 1892, l. 6-8 "(i) to identify the best possible implementation of a multivariate ocean color assimilative system based on state-of-the-art methods": I am not really sure that this point is addressed in the manuscript. For practical reasons, the authors use a simplified version of the SEEK filter and assess the performances of two different analysis schemes. But, they do not consider ensemble methods that could constitute "better implementations of a multivariate ocean color data assimilative system".

We agree that this point is not addressed in the manuscript. Sentence is changed: Âă(i) to implement a multivariate, ocean color assimilative system based on state-of-the-art

OSD
9, C775–C785, 2012

methods ”

p. 1895, l. 4-5 "and composite data would make the diagnostics of the assimilation experiments rather complex": I am not rally sure to understand what are the issues arising from the use of merged data. Composite data sets are expected to lead to a better spatio-temporal coverage of the area, so why is it important to use one sensor only?

Between 1998 and 2002, only SeaWiFS was operational. Then MODIS and MERIS were launched by 2002. Using a composite data set would change the spatial coverage of the data set between periods 1998-2002 and 2002-2006. Thus the assimilation “impact” on model trajectory would be different for these 2 periods. An good example is shown in Mélin et al. 2003 (Figure 1).

F. Mélin, G. Zibordi, S. Djavidnia. Merged series of normalized water leaving radiances obtained from multiple satellite missions for the Mediterranean Sea. Advances in Space Research, 43 (2009), pp. 423–437

p. 1894-1895 §2.2 "The ocean color data set and associated errors": The binning period of the observations in not specified. As the assimilation is performed every 8 days, I presume the authors use a time-averaged 8-days product. Is it correct?

Yes this is correct

If yes, are these observations assimilated at a date corresponding to the center of the 8-days time window? This information should be specified in the manuscript. No, it’s not centered. The observation for a date t is defined as the binning of data between t-7 days and t. While we know it may induced a temporal shift in the assimilation procedure, the idea was to stay in an operational framework where future observations are not known.

Sentence was added.

p. 1895 l.25 - p. 1896, l. 3 "The sequential approach is consistent [..] rather than assuming that one particular aspect of the model [..] is responsible for the model/data misfits": I am not sure to understand the meaning of this sentence. The authors might consider to clarify this point.

Sentence removed.

p. 1896, l. 13-14 " the upgrade of the assimilation scheme toward a fully explicit ensemble scheme will be straightforward in forthcoming applications": I presume that the upgrade of the assimilation system towards an ensemble scheme might require to work on the anamorphic transformations as well. A first issue concerns the choice of the samples used to build the empirical anamorphosis functions (from the forecast ensemble, from previous simulations as done in this study). A second issue is related to the tails of the transformations as defined in this study (see major comment [C]). Because the transformations truncate values out of range of the local samples, we might observe a depletion of the ensemble for some variables in different grid cells either when transforming the forecast en- semble (not in agreement with the historical simulation) or transforming back the analysis (not in agreement with the forecast ensemble or the historical simulation) and a divergence of the filter. A discussion on the strategy to define and apply the anamorphic transformation in the framework of explicit ensemble-based Kalman filtering could be included.

Answered in “Major Comments”' section

p. 1896, l.18-19 "in the first version, the analysis is performed using the original state variables": I presume that negative values produced by the analysis steps are processed before the propagation steps. What is the strategy adopted? A simple increase of these values to zero?

Yes, but actually the minimum concentration is set to 1e-6 mmol.m-3. Sentence was added by the end of paragraph 2.3

p. 1896, l. 29 - p. 1897, l.1 "The parametrization of the anamorphic transformation is

OSD

9, C775–C785, 2012

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

equivalent to the one in Doron et al. (2011)": More details would be helpful (see major comment [B]).

Answered in “Major Comments”' section

Introduction and §2.3 "Assimilation method": According to the bibliography and the way the references are included in both sections, it seems that only people from LEGI have worked on Gaussian anamorphosis extension of Kalman filters since the original works of Bertino et al. (2003). It could have been mentioned that anamorphosis functions were also used in Simon and Bertino (2009). In the framework of the ocean biogeochemistry, the works of Simon and Bertino (2012) investigating the strategy to build the anamorphic transformations, or Ciavatta et al. (2011) using logarithmic transformations to handle the positiveness of the variables could have been cited as well.

While it is true that a reference Simon and Bertino (2012) can be included, Bertino et al. (2003) and Simon and Bertino (2009) are already cited in the manuscript as references for a detailed anamorphosis description. A reference to Simon and Bertino 2012 was added.

p. 1898, l. 26-27 "The error associated with each distinct observation pixel is set to 30% of the considered data": I am wondering what is the strategy to specify the error of the transformed observations (see major comment [B]). What does "30% of the considered data" mean for the transformed observations?

Answered in “Major Comments”' section

p. 1899, l. 10-12 "It is noteworthy that Hu et al. (2011) recently proposed equivalent parametrization [..] experiment": The reference is 2012.

Changed

p. 1899, l. 25 "Figure 6 provides": it seems to be more natural to label this figure "Figure 2" . Furthermore, few words on the temporal coverage of the data set would be helpful (basically, how many data are available every year).

OSD

9, C775–C785, 2012

Interactive Comment

A sentence was added, with the number of data per years

p. 1900, l. 9 "over successive 60-day periods": Does it correspond to an average of 60 daily output or 7-8 analysis outputs (see major comment [A])?

Answered in “Major Comments”' section

p. 1901, l. 4-5 "this is a crucial point since the free run is actually sampled to compute EOF basis used in the assimilation scheme": It could be specified that this is also crucial for the nonlinear analysis scheme since the local transformations are built from this simulation, aren't they?

Sentence changed.

p. 1901, l. 7-13 It gives the feeling that the bloom has already started in the observations and not in the model simulations. In the same way, we note large chlorophyll concentrations in the Subpolar Gyre in the observations during the period July-August (fourth row) that are not present in the model simulations.

We agree with this comment. It might be link to the variability contains in the EOF set during these periods. Because the bloom starts too late in the free simulation, it is already started in this observation while the running EOF basis presents a low variability at this time.

p. 1901, l. 18-20 "This issue is related to the fixed-based variant of the SEEK filter chosen to assimilate data [..] increasing the temporal window during which the EOF are computed": It could also suggest to propagate the error covariance matrix during the estimation process.

Yes this is true, but the idea of the sentence is not to define what should be the best assimilation method but to suggest an evolution of the current one

p. 1902, l. 9-11 Could it be related to the fact that the anamorphic transformations do not allow values outside the range of values of the local sample defined from the free

OSD

9, C775–C785, 2012

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

run simulation? The transformation back and forth of one variable is equivalent to apply a truncation of values below and above given thresholds (the bounds of the sample). It means that the nitrate concentrations are constrained to belong to a range of values defined by the free run every 8 days. It can prevent a drift of the simulation towards large values, however it might not be the most elegant way to do it.. I am wondering why would be the impact of specifying minimum and maximum values (based on the free run) for the variables in the model on the simulation with the linear update. Would it prevent the occurrence of these localized spots in the phytoplankton concentration?

Maybe, it is difficult for us to tell weither this truncation plays a major role or not. We guess that it is not the case everywhere while it can prevent a systematic drift in concentration for some specific regions

p. 1902, l. 16-17 "these specific processes are most relevant in the context of coastal ocean color data assimilation experiments (Fontana et al., 2009, 2010; Hu et al., 2011)": The works of Ciavatta et al. (2011) assimilating ocean color data in the English Channel could be relevant as well.

Reference was added

p. 1903, l. 17-19 Again, could it be due to the fact that the anamorphic transformations constrain the observations and the solution to be inside the range of values of the local samples defined from the free run simulation (see major comment [C])? What would have been the result if "extrapolations" were added?

Same answer as p. 1902, l. 9-11 comment

§4.3 "Surface chlorophyll concentration forecast": The value of the mode of the three distributions could be specified.

We think that the figure is clear enough.

p. 1909, l. 18 "between the each experiment": between each experiment?

OSD
9, C775–C785, 2012

CC BY

Corrected

p. 1912, l. 24 "sub-tropcial"

Corrected

p. 1928 Figure 9: It would be helpful to specify what are the experiments (a-d) in the legend.

Added in the legend

Interactive comment on Ocean Sci. Discuss., 9, 1887, 2012.

OSD

9, C775–C785, 2012