Ocean Sci. Discuss., 9, C636–C645, 2012
www.ocean-sci-discuss.net/9/C636/2012/

Ocean Science
Discussions

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.

Anonymous Referee #2

Received and published: 29 June 2012

General comments

The paper "Toward a multivariate reanalysis of the North Atlantic ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data" by Fontana et al. describes a methodology for assimilating SeaWiFS data into an ocean biogeochemical model, with the aim of creating a reanalysis of chlorophyll, nitrate and other variables. The method is based on the SEEK filter, and runs with and without anamorphic transformations are performed, as well as a free run. Assessment is performed against the assimilated chlorophyll data and in situ nitrate data, and the

assimilation is found to improve both chlorophyll and nitrate, particularly when anamorphic transformations are used. The manuscript is clear and well written, and presents useful results.

There are a few points, detailed below, which would merit further discussion in the paper. In particular, one of the stated aims is to "identify the best possible implementation of a multivariate, ocean color assimilative system based on state-of-the-art methods", but this aim is not really addressed. The conclusion is reached that the use of anamorphic transformations is better than the "linear" implementation of the SEEK filter. However there is no discussion of how upgrading the simplified SEEK filter implementation may improve results, or the advantages/disadvantages of this approach compared to other multivariate approaches. Some discussion of this should be added to the "Conclusions and perspectives" section. This can be put in the context of limitations of the method implemented here, such as the use of the free run variability to specify the error covariances. As acknowledged, this means that only very small increments are applied in regions where there is little model variability, even if the model-data mis-match is high. This way of specifying the background error seems similar to the "Quick Canadian" method (Polavarapu et al., 2004, Atmosphere-Ocean) in a univariate scheme, which is generally used to provide an initial estimate for the background errors, before another method is applied in order to refine them. The method used is fine for this paper, but given the long-term aims of the work, a more accurate specification will be required in future, and so there should be more discussion of how this issue might be addressed.

Another point is the data sets used for validation. No independent chlorophyll data are compared against, only the assimilated data (although comparison is made to forecasts as well as analyses). The only other variable considered is nitrate, with comparison made to WOA data (about which there may be quality control issues, and perhaps some dependency between model and data, see below). It is demonstrated that the assimilation, with the use of anamorphic transformations, improves model

skill compared to these data sets, which is an important result, especially given that nutrients are rarely shown to be improved by biogeochemical data assimilation. However much more validation is required before confidence can be placed in a "data-driven climatology" produced using this method. Given the aims of the paper, and the availability of data, this validation is not required here, but the point should be noted when discussing the future of this approach.

Specific comments

p1891 l2&6: Please add a brief definition (or reference) of sequential and variational in this context.

p1893 l16: Please expand on the implementation of the "buffer zones". Furthermore, please state how the boundary conditions are implemented for LOBSTER.

p1896 l25-29: I appreciate that anamorphic transformations are detailed elsewhere. However it would aid understanding to expand the description given here.

p1897 l17-21: A two-year spin-up seems rather short given that nitrate is initialised from the same data set that much of the validation is performed against. Please comment on this. Does model skill compared to the WOA nitrate data remain steady throughout the nine-year reanalysis period, or does skill change later in the period, suggesting that the model nitrate is still spinning-up? If so, is the assimilation able to correct for this?

p1897 l25: Which day of the eight-day binning period are the chlorophyll maps assimilated on? Are the maps at the model resolution, or higher resolution?

p1898 l18-21: Please expand the description of this step, since it is crucial to the performance of the assimilation.

p1898 l27: Is any consideration made of representativity error (the fact that the model cannot resolve high resolution processes which affect the observations), or the error introduced by using a modelled Chl/N ratio to convert chlorophyll to phytoplankton?

Underestimating the observation error could lead to giving too much weight to the observations.

p1899 l18-21. Independent data are definitely required. Whilst I think it is perhaps sufficient for this paper, I don't think using a single data set for a single variable is enough to be "totally conclusive" about method efficiency. This is especially relevant given that the model was initialised from WOA data, and the error covariances are based on the model free run, and so there may be some dependency. Please comment on this.

p1900 ll17-19: I assume this specifically refers to the area just north of the elongated structure? Generally, chlorophyll is higher in the free run than the observations above 45°N in May/June.

p1900 l25: It seems to me that the SeaWiFS data exhibit larger values beyond the summer season too.

p1901 l1-6: Most of the main features are described, but the model does not exhibit a fall bloom or increased chlorophyll along the North American coast (I appreciate that good performance is not expected in shelf seas). This leads to these features not really being captured by the assimilation (as discussed later).

p1901 l7-24: The reliance on the free run variability seems to be a weakness which will need to be overcome if the aim of producing "data-driven climatologies" is to be realised. More discussion should be given to possible ways this could be overcome (in the "Conclusions and perspectives" section), including the improvements that might be expected from using the full version of the scheme.

p1902 l1-3: Please expand on the reasons for this.

p1903 l6-29: Overall the assimilation is doing a good job of correcting chlorophyll magnitudes. However in the free run there is sometimes a bias in the timing of the spring bloom, especially in regions #1 and #3 (seen also in Fig. 2), which the assimilation

does not seem to be correcting. Please comment on this. Generally speaking the anamorphosis run follows a very similar temporal evolution to the free run, even when this differs from that of the SeaWiFS data. Is this due to the way the errors are specified? Or to the data coverage?

p1904 l1-7: In region #4, the free run sometimes overestimates the SeaWiFS data, whereas the assimilation runs underestimate it, which seems odd. Is this just down to the averaging and data coverage used in the comparison, or is it a result of how the assimilation works?

p1904 l18: Statistical significance has not been presented, so please use a different word than "significantly".

p1905 l1: "The free run performs well" - it looks to me like the free run is performing fairly poorly, particularly compared to some of the other regions.

p1905 l12-18: State here that the non-linear run performs better than the linear run.

p1905 l28: Are these exactly equivalent? If I understand correctly, the model forecast is valid for an exact time, whereas the observations are averaged over an eight-day period. This probably doesn't invalidate the conclusion, but may affect the results. Please comment on this.

p1906 l8: Statistical significance has not been presented, so please use a different word than "significantly".

p1907 l6: Why have data shallower than 10 m been excluded?

p1907 l10: How has the colocalization procedure been performed (nearest grid square, interpolation?)

p1907 l16: Expand "RMS". Moreover, what is this the RMS of? The difference between modelled and observed nitrate? Or the (log) ratio?

p1907 l23-24: "Overestimations remain more or less unchanged ... whilst underesti-

mations are significantly reduced". Unless the model and observations agree exactly more often (which you don't show), surely a reduction in underestimations must be balanced by an increase in overestimations?

p1907 l24: Statistical significance has not been presented, so please use a different word than "significantly".

p1908 l4: Garcia et al. (2006) describes WOA05. Should the reference be Garcia et al. (2010)?

p1908 l1-11: It may well be the case that comparing to better quality controlled data will show the model to match the observations better. However this is not guaranteed, and calls into question the robustness of the conclusions based on the comparison that has been performed. Can you use the fully quality controlled data to compare against?

p1908 l16: Please state at this point why it was chosen to exclude data lower than 1 mmol(NO_3) m^{-3} from this comparison. At the moment, following on from the previous paragraph, it reads as if this is being done in order to address the quality control issues just discussed. This is not the case, it seems to be done because the model performs poorly at the lowest concentrations. Also, please say how many observations are excluded.

p1909 l19-20: "Underestimated and overestimated by (resp.) the free (a) and the linear run (b)" - is this the wrong way round? The linear run seems to underestimate in frame 2 (red dots).

p1910 l5-6: Why was this section chosen? There appears (Fig. 6) to be nitrate data here. Is this for the same period? If so, can it be shown alongside? Can the equivalent nitrate climatology section be shown too?

p1911 l10-13: I'm not sure I fully understand the reasons for this conclusion, please explain more clearly.

p1912 l26: Statistical significance has not been presented, so please use a different

word than “significantly”.

Technical corrections

p1888 l8: Change “North-Atlantic” to “North Atlantic”.

p1888 l12: Change “experiences” to “experiments”.

p1888 l17: Change “surface chlorophyll concentrations analysis and forecast” to “analysis and forecast surface chlorophyll concentrations” (or similar).

p1888 l20: Change “litterature” to “literature”.

p1888 l22: Change “assessement” to “assessment”.

p1889 l13: Reword “ocean color sensors from space”.

p1889 l15: Throughout the manuscript, both “water leaving radiances” and “water-leaving radiances” are used. Please be consistent.

p1889 l19: Expand the acronyms “MERIS” and “MODIS”.

p1890 l24: Add “e.g.” before “altimetry, sea surface temperature” - sea ice variables are also routinely assimilated operationally. These models also routinely assimilate in situ temperature and salinity data.

p1892 l1: Change “need” to “needs”.

p1892 l6: Add a reference for MyOcean.

p1892 l8: Change “performances” to “performance”.

p1892 l12-13: Please rephrase to make it clear that you mean both chlorophyll at places and times when observations are not available, and other variables (nitrate). “Non-observed variables” should probably be used only once in this sentence.

p1892 l22: Change “nonlinear” to “non-linear” and “non gaussian” to “non-Gaussian”.

p1892 l24: Add a reference for WOA09.

p1893 l11: Expand “NEMO” acronym.

p1893 l15: “Three-dimensional” has been written as “3-D” earlier in the manuscript (although not expanded).

p1893 l20: Add a reference for ERA-INTERIM and expand “ECMWF” acronym.

p1894 l4: Expand “Chl/N”.

p1894 l9: Change “may be be due”.

p1894 l15: Add a comma after “furthermore”.

p1894 l19: Throughout the manuscript, both “spin up” and “spin-up” are used. Please be consistent.

p1896 l5: Expand “SESAM” acronym.

p1896 l11: Expand “EnKF” acronym.

p1896 l22: Change “transormation”.

p1896 l27 & p1897 l3: Change “gaussian” to “Gaussian”.

p1897 l12: Throughout the manuscript, both “set up” and “setup” are used. Please be consistent.

p1898 l9: Change “CPBMs requires” to “CPBMs require”.

p1899 l3: Change “month” to “months”.

p1899 l9-10: “This value was chosen as it is in order of magnitude of meso-scale typical length feature for mid-latitude regions” - please rephrase.

p1899 l23: Expand “GODAR” and “WOD”, and change “project” to “projects”.

p1899 l28: “Allows performing” - please rephrase.

p1900 l27: Change “too late, chlorophyll” to “too late and chlorophyll”.

p1901 l21: Change “states” to “state”.

p1901 l22: Change “state” to “states”.

p1901 l26: Change “differs” to “differ”.

p1902 l13: Delete either “CPBM” or “model”.

p1904 l10: Rephrase “Gulf Stream anomalous pattern”.

p1904 l20: Change “others” to “other”.

p1906 l8: Change “diagnostics” to “diagnostic”.

p1906 l23: Rephrase “such as”. Only nitrate has been compared to.

p1909 l7: Rephrase “preventing the assimilation to perform correctly”.

p1909 l18: Change “between the each” to “between each”.

p1909 l25: Change “were” to “where”.

p1910 l9: Change “2006), as” to “2006). As”.

p1910 l17: Change “as it was” to “as was”

p1911 l21: Change “such those” to “such as those”.

p1912 l9: Change “yield” to “yields”.

p1912 l10: Change “as key” to “as a key”.

p1912 l14: Rephrase “on 8-days period”.

p1912 l18: Change “years” to “year”.

p1912 l24: Change “sub-tropcial” to “sub-tropical”.

p1913 l2: Change “prototpype” to “prototype”.

p1913 l11: Rephrase “weightly-integrated”.

p1913 l18: Rephrase “as various of” and “biogeochemical concentrations vertical distribution”.

p1913 l22: Delete “more” - currently no information is assimilated explicitly at depth.

p1913 l23: Change “explicitely” to “explicitly”.

p1913 l24: Change “measuring systematically” to “systematically measuring”.

p1914 l1: Change “avaiblable” to “available”.

Fig. 2&3: I think it would help understanding if the run names were explicitly labelled across the top (or bottom) of these figures, and the averaging period labelled down the side.

Fig. 5: It would be useful to plot a vertical line at zero.

Fig. 9: Please label which plot corresponds to which run, and plot the frames on all the maps.

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