Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2018-153-RC2, 2019

Ocean Science
Discussions
Open Access

Interactive comment on "An ensemble probabilisitic approach to reconstruct the biogeochemical state of the North Atlantic Ocean using ocean colour images" *by* Florent Garnier et al.

Anonymous Referee #2

Received and published: 27 February 2019

The manuscript examines the assimilation of satellite chlorophyll data into an ocean-biogeochemical model with stochastic parameterization. In particular the study uses the outputs of an ensemble run with stochastic parameterization to conduct single-step assimilations over one year. Thus, the assimilation is not cycled in this case, but each assimilation is applied to the ensemble state from the previous ensemble simulation. In addition, the ensemble forecast over one month initialized from the analysis ensemble on April 5, 2005 is used to assess the influence of the assimilation on the model forecast skill. The authors conclude that the assimilation improves the chlorophyll field

Printer-friendly version

Discussion paper

estimate and the ensemble. Further the method updates some nutrients and subsurface chlorophyll in a stable way. For the forecast case, a stable positive influence of the assimilation is found for one day.

When one reads the manuscript and is aware of the current state of research in the assimilation of satellite chlorophyll data into a biogeochemical model, it is difficult to determine what is actually new in this work. There are already several studies that assimilate satellite chlorophyll data and that actually do perturb parameters of the biogeochemical model (Hu et al. 2012, Doron et al. 2013, Ciavatta et al. 2016, Jones et al. 2016, Pradhan et al. 2019). However, the exact way how the perturbations are applied are different from that used here. To this end the manuscript shows that the stochastic perturbation approach of Garnier et al. (2016) leads to ensemble covariances that are usable for data assimilation (at least for a single step). On the other hand, the data assimilation setups of other studies include cycled data assimilation and are hence far more advanced than the single-step data assimilation methodology applied here. From my experience, one cannot deduce a stable successful data assimilation process from a single analysis step. Thus, the data assimilation experiments of the manuscript can only indicate that the real cycled assimilation might be successful (because the single analysis step looks promising). With regard to the forecasting experiment, the finding that the assimilation leads to better forecast over one day is far shorter than what other studies find. This might indicate that the perturbation method applied here is in fact not usable for operational biogeochemical data assimilation, in contrast to the claim at the end of the abstract. Over all, I cannot recommend the manuscript for publication in the present form. Below I provide some recommendations, which might enhance the study sufficiently to be relevant enough for publication.

Recommendations:

I find it essential that the authors perform cycled data assimilation. A single analysis step (or many single analysis steps) do not allow to conclude that a cycled assimilation process (and this is required to get significant improvements of the model state over time) will be successful.

With the focus of the authors to ’assess the benefits of using error covariances based on an explicit simulation of model uncertainties’ and ’The relevance of this approach is evaluated’ (lines 8-9) it will be relevant to compare the proposed stochastic perturbation scheme with standard schemes, e.g. that applied by Doron et al (2011), of which some authors are the same as in the manuscript. A ’benefit’ or ’relevance’ can only be determined in comparison to other methods, which is completely missing the manuscript.

The authors also need to perform a verification of their results with independent data. Actually, the statement ’... SeaWIFS ocean colour data are used to update the ensemble...’ (line 13) and that MERIS satellite data is used for comparison (line15), suggests an independence which is actually not present. According to the manuscript, OC-CCI satellite data obtained from the Copernicus web site is used for the assimilation. The OC-CCI data is actually a combination of data from several satellites (SeaWiFS, MERIS and MODIS in 2005). Thus MERIS data is actually part of the OC-CCI data set and hence the authors verify their assimilation influence with regard to a part of the assimilated data. (This fact is actually not too obvious from the Copernicus web site, but the original OC-CCI data files (www.oceancolour.org) contain the information which satellites contribute to the data)

With regard to the description of the methodology it is irritating that the authors explain the full ETKF method in Sec. 3.1.1 but avoid any details on the anamorphosis transformation. Actually, the ETKF is a widely-used standard method has already been explained in many papers. In contrast, the anamorphosis methodology is much less common. To this end it would be much more relevant to explain the details of the anamorphosis method instead of showing the equations of the ETKF. This similarly holds for the decomposition of the CRPS score in Sec. 3.3.2. Here, the manuscript provides the definition of the CRPS, but only mentions that the CRPS can be decomposed into the terms ensemble reliability and potential resolution. Then, the authors then show only the two decomposition terms. However, without the definition of the two decomposition terms and their proper explanation, it is impossible for a reader to actually interpret Fig. 7, which shows the effect of the assimilation on these terms.

In the experimental setup, I find the handling of the observation error very unusual. First, a observation error of 0.3 is used (line 227). This value is never motivated. It was the error target for SeaWiFS, but actually, the OC-CCI data include spatially varying error information. I'm wondering why this is not used in the experiment. Further, the authors describe that to account for possible error correlations in the observations, the observation error variances are inflated, but the observation error covariance matrix is kept diagonal. While I have seen such argumentation before, the authors miss to provide any reference for this choice. As such, the approach looks fully ad hoc here without theoretical foundation of the method. The inflation of the observation error variance is then performed with a factor of 55.2 (2.3 x 24), such that the assimilation uses an observation error variance of 4.968. I cannot remember any study assimilating chlorophyll data with such an extreme value, and I cannot follow the argumentation of the authors to perform this inflation (Ciavatta et al. (2016) describe the handling of the observational errors, based on OC-CCI product user guide, when averaging them onto a coarser resolution grid. However, this does not account for correlated errors) Actually, the extreme inflation of the observational error variance should lead to a very small influence of the assimilation. However, from the description of the results this is not evident. This might actually imply that the ensemble is too disperse, because only a large ensemble variance can counter a large observation error variance in the assimilation. Fig. 4 indicates that this might be the case. In particular at STAT_B (described cursorily as 'in the equatorial region' in line 283) we see minimum ensemble values very close to zero (seemingly below 0.1 mg/mˆ3) while at the same time values up to about 2.5 mg/mˆ3 at reached by other ensemble members. Given that this happens at the same location at the same time, the ensemble appears to represent so vastly different condition that it doesn't seem to have a reasonable skill here (which seems to contradict the histogram in Fig. 6, which shows on under-dispersive ensemble on average over the whole model domain). For STAT_E, also shown in Fig. 4, the ensemble spread is lower, but still the values seem to vary by a factor of 8 (between 0.1 and 0.8 mg/mˆ3), which is also large. The authors should clearly give the theoretical basis for the inflation and should more clearly describe how the observations are used (I have the impression that the used observation operator does interpolate from the 1/4deg model grid to the observation resolution of 1/24deg (OC-CCI is usually provided at 4 km resolution) without any super-obbing).

I further recommend that the authors perform a more thorough literature research. The manuscript does not leave the impression that the authors are really aware of the current status of the research on the assimilation of satellite chlorophyll data into ocean-biogeochemical models with ensemble data assimilation methods. E.g. Nerger and Gregg (2007, 2008) have already shown successful chlorophyll assimilation with an ensemble Kalman filter. Hu et al. (2012) demonstrate a successful assimilation with a rather highly-resolved regional model. Further, any reference to papers by Ford et al. (2012, 2017), current papers by Ciavatta et al. (2018) and Skakala et al (2018) are missing. Also the study by Pradhan et al (2019) should be relevant. It is also essential that the author relate their results in the subsections of Sec. 3 to already published findings. This is completely missing at the moment even more leaving the impression that the authors are not aware of the current status of this research. This even concerns studies like Doron et al. (2011, 2013) which are referred to in the methodological part, but their findings are ignored in Sec. 3. Relating the results to other studies does e.g. immediately show that the improvement of forecasts over just one day is extremely short (e.g. Pradhan et al. show that 5-day forecasts are still much better than the free running ensemble, but actually for a coarse global model. However, Ciavatta et al. use a much higher resolved model, but obtain longer forecast influence, and in some cases even assimilate only once per month.).

Finally, I like to recommend that the authors are more careful in their formulations. E.g. 'using ocean colour images' used in the title is incorrect. Used are neither 'images' nor

Printer-friendly version

Discussion paper

'ocean colour', but concentration maps of chlorophyll, which are derived from ocean color. Also 'we present first results illustrating the potential of our approach for biogeochemical forecasts.' (line 20) is misleading. Indeed, these are the 'first results' in which the chlorophyll assimilation is used with the perturbation method of Garnier et al. (2016). However, this ignores all other published studies, as the results described in the manuscript are not at all the first results in which chlorophyll is assimilated to improve the forecasts of biogeochemical model fields and of course not the first results pointing to the potential operational use of biogeochemical data assimilation. The authors leave a very odd impression of being 'first', while the assimilation methodology presented in the manuscript is in fact behind the state-of-the-art.

Given these significant shortcomings of the manuscript, I omit minor comments, like grammatical errors or unclear sentences. Anyway, I have the impression that the perturbation method used in the study can have potential, but this would need to be carefully shown.

References: Ciavatta et al. JGR Oceans 123 (2018) 834-854 Ford et al, Remote sensing of Environment 203(2017) 40-54 Ford et al., Ocean Sci. 8 (2012) 751-771 Hu et al., J. Mar Syst. 94 (2012) 145-156 Jones et al., Biogeosciences 13 (2016) 6441-6469 Nerger & Gregg, J. Mar. Syst. 68 (2007) 237-254 Nerger & Gregg, J. Mar. Syst. 73 (2008) 87-102 Pradhan et al., JGR Oceans 124 (2019) 470-490 Skakala et al., JGR Oceans 123 (2018) 5230-5247

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-153, 2019.