Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-258-RC1, 2017

Biogeosciences Discussions

Interactive comment on "Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea" *by* Daniel E. Kaufman et al.

M.E. Gharamti (Referee)

gharamti@ucar.edu

Received and published: 1 September 2017

The article presents a DA study in the Ross Sea, a region of the southern ocean. The authors use bio-optical glider data to reduce the model-data misfit of Chl concentration and POC. In the process, 8 different uncertain parameters are identified and optimized with incoming observations. The authors provide a thorough assessment of the DA system by changing the spatial and temporal resolution of the observations. This is performed in an effort to understand the impact of the number and type of observations (e.g., cruise and satellite) on the resulting biogeochemical modeling skill.

I think the paper is well-written, clear and nicely organized. The authors tackle an interesting problem that researchers within the DA-marine ecology community have

been investigating for a while. Although the results from such a small domain and a 1D model can not be generalized for large-scale problems (the authors recognize this), the article presents novel research points especially those of the parameter optimization. I have few minor comments (below), otherwise I don't see any reason for not publishing this article. It would be good to address the comments below in the manuscript before publishing.

1- Section 2.2: I would like to see how the observational error variance is parameterized. I believe the observational mapping operator is quite nonlinear. So, what procedure did the authors follow to find both \sigma_chlˆ2 and \sigma_pocˆ2 (twin and real experiments)?

2- Maybe I missed it but it would be good to provide a discussion on the computational cost of the genetic algorithm. Obviously, the authors are using some kind of hybrid algorithm (genetic + Powell) but I'm pretty sure these (non-gradient based) won't be as useful in large scale models. For instance, if the biogeochemical parameters are spatially varying then the degrees of freedom in the system will significantly increase. I'm not so much familiar with the algorithm the authors are using, so it would be good to see how does it compare computationally to an EnKF for example.

3- Section 2.4: I know it's mentioned somewhere, but it would be good to state that the algorithm selects random parameters within a range. After all, the chosen parameters need to be physically meaningful.

4- Lines 168-170: I am not sure what the authors mean by this sentence. Consider rephrasing.

5- Section 2.5.2: I think adding a small appendix section summarizing the differences between a MC and a Latin Hypercuble sampling would be useful for the reader.

6- Section 2.5.3: Why optimizing more parameters (>8) was not successful? Any reason for this, statistical one perhaps? Is it because the parameters maybe spatially

varying and this assumption is relaxed in the objective function? Or could it be due to the choice of the observational error variance? On another note, how to make sure it's not a drawback from the optimization algorithm itself? A paragraph addressing this is needed here. I could not find an explanation for such a behavior myself.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-258, 2017.