

**Response to ‘Interactive comment on “Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea” by Daniel E. Kaufman et al.’ by M.E. Gharamti that was received and published: 1 September 2017**

Dear M.E. Gharamti,

We greatly appreciate your valuable and constructive comments, which will help us improve and clarify the manuscript and presentation of the parameter selection and optimization procedures in particular. In response to your questions and suggestions, please find our answers and proposed changes (in blue) following each of your comments below. All line numbers refer to the original submitted manuscript.

We thank you again for your review.

Sincerely,

Daniel Kaufman, Marjorie Friedrichs, John Hemmings, Walker Smith

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.

Thank you for your positive comments.

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  $\sigma_{chl}^2$  and  $\sigma_{poc}^2$  (twin and real experiments)?

The procedure we used was simply to calculate the inverse of the standard deviation, similar to other assimilation efforts (e.g. in Experiment #1 in Hemmings and Challenor, 2012; Friedrichs et al., 2006; Xiao and Friedrichs, 2014b). This definition of sigma was used for both the twin experiments described in section 2.5.3 (where the sigma was

calculated from the synthetic dataset) as well as the real assimilation experiments described in section 2.6 (where sigma was calculated from the particular observation set assimilated in each case).

You may notice that the second reviewer raised a similar question, and we there also indicate our proposed text for section 2.3, referencing other assimilation studies that have used standard deviations to weight the misfit contributions. Specifically, we propose modifying the text on lines 126-128 as follows: “where  $N$  is the number of observation points,  $x_i$  is the simulated value of either chlorophyll or POC at the  $i$ th observation point and  $y_i$  is its observed value;  $\sigma$  is the standard deviation of the specific observation set assimilated in a particular experiment. Using the standard deviation of the observations to define a characteristic scale of variation for each variable is a technique used in previous studies (e.g. Friedrichs et al., 2006; Xiao and Friedrichs, 2014). It is designed to weight the relative misfit contribution of each variable appropriately when there are insufficient data to define a comprehensive error model. Such a model would require reliable information about the uncertainty associated with observation errors (instrument error and error of representativeness) and non-parametric errors in the simulation such as forcing errors (Schartau et al., 2017). The use of different cost function weighting schemes in plankton modelling including the characteristic scale technique is explored in more detail by Hemmings and Challenor (2012).”

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.

To facilitate comparison with other assimilation methods, including EnKF, we propose including the number of model evaluations (approximately 4000 - 5000) in the text as described below. The number of evaluations and therefore the computational cost for our method is typically higher than EnKF, because our method is designed for a more comprehensive investigation of the parameter space. We think including computational cost in terms of run times in the manuscript would be of limited value because it would then also be necessary to report all hardware specs, which may be too much info and regardless, the hardware will probably be out-of-date in just a few years.

You are correct that these optimization methods won't be as useful in large-scale models when applied directly. However, the parameters identified in a 1D model by these techniques can be used in larger 3D models, and this has been shown to improve those larger models (e.g. see the review in section 7.2 of Schartau et al., 2017). Furthermore, you are correct to point out that allowing parameters to vary spatially would increase the degrees of freedom in the system with further implications for the practicality of our method. However, the method is not intended for estimating spatially varying parameters but for estimating parameters that are spatially uniform over as large a domain as

possible. Given the increased model degrees of freedom associated with spatially varying parameters and the consequent increased risk of over-fitting, it is unclear to what extent allowing parameters to vary spatially would be useful. The issue is discussed in detail in Schartau et al (2017).

We propose adding a paragraph at the end of section 4.1 to make the above points clear to the reader: “The high number of model evaluations in each optimization case (roughly 4000 – 5000) makes such direct optimization impractical for large-scale models; however, the parameters identified in a 1D model by these techniques can be used in larger models, and indeed locally optimized parameters have been previously shown to improve the skill of 3D models in other regions (Oschlies and Schartau, 2005; Kane et al., 2011; McDonald et al., 2012; St-Laurent et al., 2017). It is expected that the optimized parameter values found in the one-dimensional assimilation experiments described here will be of value in a future 3D biogeochemical modeling analysis of the Ross Sea and, through model inter-comparisons, provide a basis for examining the dependence of these parameter values on model structure and level of complexity, as has been done elsewhere (Friedrichs et al., 2007; Bagniewski et al., 2011; Ward et al., 2013; Irby et al., 2016).”

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.

This is a good point, and yes, the selection of values from within the range for each parameter is discussed in section 2.5.1. Nevertheless, it is understandable for the reader to be wondering about it earlier in section 2.4. To help the reader, we propose adding a sentence at line 147 that provides a brief clarification and cross-references section 2.5.1: “The constituent parameter values are selected randomly from within a pre-determined range of allowable values (Sect. 2.5.1).”

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

In order to clarify the meaning of this sentence, we propose rephrasing it to: “Ideally, optimal values are identified for all parameters in a model, however, uncertainty in the parameter estimates from an algorithmic optimization increases as the number of parameters included in that optimization increases (Friedrichs et al., 2007; Ward et al., 2010).”

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

This is a good idea. We propose adding an appendix titled “Appendix A: Latin hypercube sampling (Sect. 2.5.2)” and inserting a cross-reference to this appendix on line 199. The proposed appendix would have the following suggested text:

“Latin hypercube sampling (LHS) and Monte Carlo sampling are both techniques that can be used to randomly draw a finite number of samples from input distributions in order to approximate a full multidimensional distribution. The LHS incorporates stratified random sampling, i.e. in each dimension each sample is drawn randomly from within a different interval (also called a stratification or layer) of the distribution (McKay et al., 1979). Intervals are chosen with reference to the probability distribution such that each represents an equally probable range. In contrast, Monte Carlo sampling proceeds in each dimension with each sample drawn randomly from the entire distribution. Stratified random sampling with intervals of uniform probability ensures a good representation of the distribution, reducing the risk of samples being clustered in one or a small number of areas. In LHS sampling, if the sample size is  $n$ , each dimension is divided into  $n$  intervals such that in multi-dimensional space each interval of each dimension is sampled once and once only. This is based on the idea of a Latin square in which an individual symbol appears once in each row and each column. It ensures a good representation of the distribution is achieved for all dimensions.”

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.

The primary reason for unsuccessfully optimizing more parameters is that many model parameters are correlated. Previous studies have also followed a procedure to reduce the set of optimizable parameters, particularly to avoid optimizing highly correlated parameters. For examples of these procedures, see Xiao and Friedrichs, 2014b, Friedrichs et al. 2006 (or 2007).

To help any reader wondering similarly about limiting the number of optimized parameters and to clarify that this is not a drawback of this specific optimization algorithm, we propose adding the following paragraph in section 2.5.3 (line 222): “There is a limit to the number of parameters that can be independently constrained by the available observations because varying different parameters can often have similar effects on the cost function. Optimizing a larger set increases the potential for correlation between the effects of different parameters, reducing the algorithm's effectiveness in identifying unique optimal parameter sets. This, combined with the increased potential for over-fitting associated with the greater model degrees of freedom, can reduce the ability of an optimized model to reproduce independent data sets (Matear et al., 1995; Friedrichs et al., 2007; Xiao and Friedrichs et al., 2014b). The limitation on the number of optimizable parameters applies to both  $\mu$ GA and variational adjoint optimizations (Ward et al., 2010). In fact, rather than being a function of the optimization algorithm, it is dependent on the available data and the design of the cost function. A larger or richer observation set can help to constrain more parameters. The impact of cost function design is more complicated because an improved cost function may allow for greater uncertainty

in the observations and/or non-parametric uncertainty in the simulation, leading to weaker but more realistic constraints on the parameters (Hemmings & Challenor, 2012).”

In addition, we propose modifying the first sentence of section 2.5.3 (line 213) to read: “After selecting the 21 potentially optimizable parameters, Numerical Twin Experiments (NTEs) were conducted to identify an optimizable subset by evaluating...”

Additional literature cited in responses:

Bagniewski, W., Fennel, K., Perry, M. J., and D’Asaro, E. A.: Optimizing models of the North Atlantic spring bloom using physical, chemical and bio-optical observations from a Lagrangian float, *Biogeosciences*, 8(5), 1291–1307, doi:10.5194/bg-8-1291-2011, 2011.

Irby, I. D., Friedrichs, M. A. M., Friedrichs, C. T., Bever, A. J., Hood, R. R., Lanerolle, L. W. J., Li, M., Linker, L., Scully, M. E., Sellner, K., Shen, J., Testa, J., Wang, H., Wang, P., Xia, M.: Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison, *Biogeosciences*, 13(7), 2011–2028, doi:10.5194/bg-13-2011-2016, 2016.

Kane, A., Moulin, C., Thiria, S., Bopp, L., Berrada, M., Tagliabue, A., Crépon, M., Aumont, O., and Badran, F.: Improving the parameters of a global ocean biogeochemical model via variational assimilation of in situ data at five time series stations, *J. Geophys. Res. Ocean.*, 116(6), 1–14, doi:10.1029/2009JC006005, 2011.

McDonald, C. P., Bennington, V., Urban, N. R., and McKinley, G. A.: 1-D test-bed calibration of a 3-D Lake Superior biogeochemical model, *Ecol. Modell.*, 225, 115–126, doi:10.1016/j.ecolmodel.2011.11.021, 2012.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Value of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21(2), 239–245, 1979.

Oschlies, A., and Schartau, M.: Basin-scale performance of a locally optimized marine ecosystem model, *J. Mar. Res.*, 63(2), 335–358, doi:10.1357/0022240053693680, 2005.

St-Laurent, P., Friedrichs, M.A.M., Najjar, R.G., Martins, D.K., Herrmann, M., Miller, S.K., and Wilkin, J.: Impacts of atmospheric nitrogen deposition on surface waters of the western North Atlantic mitigated by multiple feedbacks. *J. Geophys. Res. Ocean.*, in press September 2017.

Thomalla, S. J., Racault, M., Swart, S., and Monteiro, P. M. S.: High-resolution view of the spring bloom initiation and net community production in the Subantarctic Southern Ocean using glider data, *ICES J. Mar. Sci. J. du Cons.*, 72(6), 1999–2020, doi:10.1093/icesjms/fsv105, 2015.