Dear Dr. Marilaure Grégoire, Associate Editor

Thank you for your time and effort considering our manuscript for publication.

Please find below our responses to both reviewers' comments (starting on pages 2 and 8 of this pdf) followed by the revised manuscript (starting on page 14 of this pdf), which shows our tracked modifications.

Best regards,
Daniel Kaufman, Marjorie Friedrichs, John Hemmings, Walker Smith

**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 µGA 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.

**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 Anonymous Referee #2 that was received and published: 5 September 2017**

Dear Reviewer #2,

We greatly appreciate your time and effort spent reviewing our manuscript. According to your constructive feedback, we propose changes to clarify aspects of model setup, optimization method, and conclusions. Please find our responses (in blue) following each of your comments below. All line numbers refer to the original submitted manuscript. Thank you again for your review.

Sincerely,
Daniel Kaufman, Marjorie Friedrichs, John Hemmings, Walker Smith

Review comments for the manuscript: "Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea (bg-2017-258)" by Daniel E. Kaufman, Marjorie A. M. Friedrichs, John C. P. Hemmings, and Walker O. Smith Jr.

The authors present a data assimilation study that optimizes parameters in a one-dimensional biogeochemical model using glider observations in the southern Ross Sea. They show insensitivity of the result to the geographical location of observations, but the optimizing parameters is sensitive to the sampling frequency.

The paper is overall well-written, but I hope the reviewers be able to address comments that I have.

– The procedure can be clarified more. This study utilizes a one-dimensional model for 3D observations. Does the cost function use all the observations and estimate one set of the parameter? Or is there an optimized parameter set for each location? If the first approach is used, do you expect that the optimized parameter values represent the distribution of those obtained by the second approach?

As described on line 258 and in Table 2, the cost function uses all of the observations for experiment #1A. To make this clearer, we propose modifying the sentence in the abstract on line 16 to read "Assimilation of data from the entire glider track …"

We do find that the optimized parameter values from the first approach (using all observations) represent those obtained by the second approach (using observations from different locations), as described in the latitudinal (Expt. 1b) and longitudinal (Expt. 1c) experiments and shown in Table 4.

– The authors argue that the data assimilation performance is sensitive to the observation sampling frequency due to "mesoscale variability". Mesoscale variability also means the variation in space with the scale of O(100km). But it is odd to see that the geographical region does not show a big impact on the performance. Could the author comment on this?

Here we define mesoscale variability as “days-weeks, 1-10 km” (line 59). We had no a priori expectation that the geographical regions would show minor differences in model solutions, however we believe that the minor differences are reasonably explained in section 4.2, and especially in this section’s last paragraph. Although in situ observations from previous studies have shown spatial differences on these scales, it has been unclear whether the differences were due to temporal or spatial variations. The assimilation experiments in this study suggest that variability observed on the mesoscale in this geographical region may be more likely due to temporal patterns than spatial differences. Therefore, one could expect that assimilating these different locations would show a bigger impact if the observation times concurrently varied, such as is demonstrated in the cruise-based and satellite-based assimilation cases. On larger scales, however, it is likely that the importance of spatial variability would be greater. For instance, the distinct spatial differences observed by satellites are generally across scales larger than the 1-10 km discussed here.

– By construction, the role of advection is not considered in this study. Can authors comment on the role of advection in this region? Do authors think the insensitivity of the assimilation performance to the geographical location of observations is related to the omission of advection?

Previous studies have suggested that horizontal transport and eddies may be important near island land masses and the Ross Ice Shelf (Gerringa et al., 2015; Li et al., 2017). In this region of the Ross Sea particularly, moorings and modeling have indicated moderate westward currents close to the ice shelf (Keys et al., 1990); nevertheless, advection appears to be weaker as one moves farther from the shelf edge (Dinniman et al., 2003).

One cannot rule out the possibility that the sensitivity of the optimizations to the observations’ location could be affected by adding advection to the model. However, this would likely only be the case if there were, in reality, strong horizontal velocity gradients, i.e. differences in advection between the observation locations. A more thorough examination of the role of horizontal advection on modeled dynamics of the phytoplankton assemblage is beyond the scope of the current study, but would be benefitted greatly by contemporaneous and co-located mooring and/or ship-based current measurements.

– line 85. The effort on estimating biological state variables can be listed here. (e.g., Song, H., C. A. Edwards, A. M. Moore and J. Fietcher, 2016: Data assimilation in a coupled physical-biogeochemical model of the California Current System using an incremental lognormal 4-dimensional variational approach: Part 3, Assimilation in a realistic context using satellite and in situ observations. Ocean Model., 106, 159-172.)

It is a good idea to reference Song et al. 2016 here.

– section 2.1: What is the vertical resolution of the model?

To clarify this, we propose adding details of the model setup to section 2.1, on line 108: "The model is configured to focus on dynamics within the euphotic zone with a vertical resolution of 5 m from the ocean surface to 200 m."

– line 114: The full name of BCO-DMO can be given.

Absolutely. We will expand the acronym to the full name.

– line 115: 5-m vertical binning is done using averages? or weighted average?

Vertical binning of the glider data was accomplished using averages, and to clarify this we propose modifying the sentence to read: "Data spanning the upper 200 m of the water column were binned by means into hourly, 5-m vertical bins.

– Equation for the cost function shows that the observational error covariance is estimated using the standard deviation of the observations. Is this right? I think using standard deviation may overestimate the observational error if the blooms dominate the chloro- phyll variability. If the error levels of the instruments are known, why not use these values?

The misfit contributions are weighted by using 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, 2014). If the aim were to estimate observational error, then the increase in variance due to the bloom would indeed likely lead to over-estimation. However the aim here is to weight the misfit contributions of chlorophyll and POC, and there is less impact of the bloom on these relative weights. Generally, a more sophisticated treatment of uncertainty in both the observations and the model is desirable as indicated by Hemmings & Challenor (2012), but such a treatment is beyond the scope of the present study and may not be practical with the available data. It makes sense therefore to initially employ a simple well-established method as we have done, but we recognize that it does have its limitations.

You may notice that the first reviewer raised a similar question as well, 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)."

– section 2.4: Personally, it is not easy to digest this method. Maybe a diagram can help me and readers to understand the assimilation procedure better.

We appreciate the difficulty in understanding this section without a high level overview. Although we do not believe a full diagram is necessary, we propose two changes to this section to offer the reader a broader view of the method, rather than its current focus on technical details.

First, to clarify what is being done, rather than how, we propose changing the title of this section from "Implementation of micro-genetic algorithm and direction set algorithm" to "Cost function minimization."

Second, we propose adding a paragraph to the beginning of this section that summarizes the role of the two algorithms:
"Model parameters were optimized in MarMOT by finding the minimum of the cost function (Sect. 2.3) through a combination of the micro-genetic algorithm (µGA) and Powell's non-gradient direction set algorithm. The µGA runs first and identifies sets of parameter values that produce low cost values; this is achieved by "evolving" a population of various parameter sets over successive iterations, called generations. The low-cost parameter sets identified by the µGA are then used as starting points for the direction set method, which performs successive linear searches to identify nearby lower cost solutions."

– lines 244–245: Can you provide the number for the difference? If these two cases (50 m vs 200 m) are not significantly different, I would rather present the one with 200 m. Is it because of the computational time? (Also I hope the authors say something about the speed of this data assimilation calculation).

There is a relatively minor (~14%) difference between the results of the assimilation down to 50 m compared to 200 m. The trends and major conclusions of the study are likely not strongly affected by this choice. Conducting the assimilations for the upper 50 m avoided issues related to assimilating many low values of chlorophyll and POC, and also enabled a direct comparison of these results with the results of Kaufman et al. (2017) who similarly focused on the upper 50m concentrations. Computational time did not play a role in the decision to present results for the upper 50 m.

In further response to your question about computational cost, along with reviewer #1, we propose adding the number of model evaluations conducted for the assimilation experiments to the end of section 4.2, with the text: "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]."

– section 2.6.2: Are there any changes in spatial coverage between "glider", "cruise" and "satellite" data cases? If they have the same spatial coverages, naming this way may confuse readers because it is obvious that their spatial coverages are significantly different.

As mentioned (on line 418), these cases alter both spatial and temporal resolution, and therefore they don't have identical spatial coverage. As such, we feel these names are appropriate.

– lines 474–476: Do authors have any ideas why satellite-derived data underestimates carbon export?

This is addressed earlier in the manuscript on line 424: "The lower estimates of carbon export occurred because the optimal diatom fraction for fast-sinking detritus obtained via the assimilation of surface-only data (0.62 ± 0.14) was significantly lower than that obtained via the assimilation of data throughout the upper 50 m (Expt. 2a: 0.86 ± 0.05; Expt. 2b: 0.86 ± 0.11)."

– lines 480–483: I think the phrases after "and it is" are not necessary. Please consider to remove them.

Excellent idea. We agree and will take out the phrase starting with "and it is", and we will also remove the unnecessary "Ross Sea" on line 478.

Additional literature cited in responses:

Dinniman, M. S., Klinck, J. M., and Smith, W. O.: Cross-shelf exchange in a model of the Ross Sea circulation and biogeochemistry, *Deep-Sea Res. II*, 50(22–26), 3103–3120, doi:10.1016/j.dsr2.2003.07.011, 2003.

Gerringa, L. J. A., Laan, P., van Dijken, G. L., van Haren, H., De Baar, H. J. W., Arrigo, K. R., and Alderkamp, A.-C.: Sources of iron in the Ross Sea Polynya in early summer, *Mar. Chem.*, *177*, 447–459, doi:10.1016/j.marchem.2015.06.002, 2015.

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.

Keys, H. (J. R.), Jacobs, S. S., and Barnett, D.: The calving and drift of iceberg B-9 in the Ross Sea, Antarctica, *Antarct. Sci.*, *2*(3), 243–257, doi:10.1017/S0954102090000335, 1990.

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.

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.

# Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea

Daniel E. Kaufman[1], Marjorie A. M. Friedrichs[1], John C. P. Hemmings[2,3], Walker O. Smith Jr.[1]

[1]Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, USA
[2]Wessex Environmental Associates, Salisbury, UK
[3]now at Met Office, Exeter, UK

*Correspondence to:* Daniel E. Kaufman (dkauf42@gmail.com) and Marjorie A. M. Friedrichs (marjy@vims.edu)

Author
**Deleted:** dekaufman@vims.edu

**Abstract.** The Ross Sea is a region characterized by high primary productivity in comparison to other Antarctic coastal regions, and its productivity is marked by considerable variability both spatially (1-50 km) and temporally (days to weeks). This variability presents a challenge for inferring phytoplankton dynamics from observations that are limited in time or space, which is often the case due to logistical limitations of sampling. To better understand the spatiotemporal variability of Ross Sea phytoplankton dynamics and determine how restricted sampling may skew dynamical interpretations, high-resolution bio-optical glider measurements were assimilated into a one-dimensional biogeochemical model adapted for the Ross Sea. Assimilation of data from the entire glider track using the micro-genetic and local search algorithms in the Marine Model Optimization Testbed improves model-data fit by ~50%, generating rates of integrated primary production of 104 g C $m^{2}$ $y^{-1}$ and export at 200 m of 27 g C $m^{-2}$ $y^{-1}$. Assimilating glider data from three different latitudinal bands and three different longitudinal bands results in minimal changes to the simulations, improves model-data fit with respect to unassimilated data by ~35%, and confirms that analyzing these glider observations as a time series via a one-dimensional model is reasonable on these scales. Whereas assimilating the full glider data set produces well-constrained simulations, assimilating subsampled glider data at a frequency consistent with cruise-based sampling, results in a wide range of primary production and export estimates. These estimates depend strongly on timing of the assimilated observations, due to the presence of high mesoscale variability in this region. Assimilating surface glider data subsampled at a frequency consistent with available satellite-derived data results in 40% lower carbon export, primarily resulting from optimized rates generating more slowly sinking diatoms. This analysis highlights the need for strategic consideration of impacts of data frequency,

Author
**Deleted:** glider

duration, and coverage when combining observations with biogeochemical modeling in regions with strong mesoscale variability.

## 1 Introduction

Phytoplankton blooms in the Ross Sea are responsible for some of the highest rates of productivity in the Southern Ocean (Arrigo et al., 2008), and yet the phytoplankton assemblage exhibits considerable spatiotemporal variability (DiTullio and Smith, 1996; Hales and Takahasi, 2004; Smith et al., 2010). This heterogeneity, and the spatial/temporal limitations of observations due to logistical challenges of sampling, may affect the inferred phytoplankton dynamics and produce biases in productivity or export estimates. The magnitude of the underlying ecosystem variability that contributes to these potential biases is not well understood, nor is it well known how the use of different observational platforms in the Ross Sea might affect the inferred dynamics. Acquiring data with an appropriate resolution is important for assessing phytoplankton variability in the Ross Sea (Hales and Takahashi, 2004).

Over the past several decades, biogeochemistry in the Ross Sea has been observed by ship and satellite, providing data at different temporal and spatial resolutions. Since Ross Sea phytoplankton became a focus of scientific research in the late 1970s, water column measurements have primarily come from research vessels (e.g., El-Sayed et al., 1978; Smith and Nelson, 1985; Vaillancourt et al., 2003). Typically, sampling stations are separated by tens of kilometers (Hales and Takahashi, 2004), and although vessels may return to resample a station, they typically do not return more than once or twice in a single year. During the 1990s, the use of remote sensing was expanded to look more closely at the Ross Sea bloom (Arrigo and McClain, 1994), and satellite retrievals have continued to provide valuable insights into characteristics of the phytoplankton assemblage (Arrigo et al., 1998; Arrigo and van Dijken, 2004; Peloquin and Smith, 2007; Schine et al., 2015). Satellite observations offer a synoptic view of spatial regions at frequencies that are within the time scale of biological changes (e.g. growth); however, the presence of sea ice and clouds often obscures remote-sensing measurements in the Ross Sea (Arrigo et al., 1998).

At the mesoscale (days-weeks, 1-10 km), gliders are a relatively new and effective means to characterize phytoplankton variability, and the development of ice-avoidance algorithms has enabled the use of gliders in the Ross Sea for these purposes. For example, a glider equipped with bio-optical sensors was directed along a section near 76° 40′ S in austral summer 2010 - 2011 and provided valuable estimates of biomass variability on short time scales (Kaufman et al., 2014). Estimates of the POC:Chl ratio from the glider optical sensors suggested a transition from a *Phaeocystis antarctica* to a diatom-dominated assemblage over several days (Kaufman et al., 2014; Thomalla et al., 2017). Moreover, Jones and Smith (2017) used glider observations from austral summer 2012-2013 to distinguish three phases of the Ross Sea bloom and identified high frequency (hours) associations between wind-driven mixing and biomass. A perennial challenge when using glider data (as well as ship-based data), however, is separating the effects of time and space (Kaufman et al., 2014; Little, 2016).

Numerical models are another approach for examining phytoplankton variability in the remote Ross Sea, providing an effective means for coordinating knowledge and understanding the underlying system complexities (Leonelli, 2009; Vallverdú, 2014). Furthermore, numerical simulations offer the ability for experimental manipulations that would be impractical or impossible in the real system. Such manipulations were implemented in the scenario experiments described by Kaufman et al. (2017a) to investigate how projected climate changes might alter the dynamics of the phytoplankton assemblage. These experiments showed that earlier availability of low light resulting from sea ice reduction was the primary driver of projected increases in production and export and composition change over the next century.

Data assimilation, which refers to methodologies that systematically combine a mathematical model with observations, is often used in biogeochemical applications (Hofmann and Friedrichs, 2001, 2002) to improve estimates of model parameters that are frequently poorly known (Lawson et al., 1995, 1996; Matear, 1995; Fennel et al., 2001; Friedrichs, 2002; Schartau and Oschlies, 2003; Hemmings et al., 2004; Bagniewski et al., 2011; Doron et al., 2013; Xiao and Friedrichs, 2014a,b; Melbourne-Thomas et al., 2015; Song et al., 2016; Gharamti et al., 2017, Schartau et al., 2017). This entails a smoothing or optimization procedure, in which elements of the model are adjusted to minimize differences between the model output and the observations. Typically, an aggregate measure of the differences between observations and model output is provided by calculation of a cost function, defined as the model-data misfit, and an optimization algorithm searches for model parameters that minimize the value of this cost function.

In this study, data assimilation is used to obtain an optimal representation of Ross Sea lower trophic levels. Specifically, observations from an autonomous glider are assimilated into a biogeochemical model of the Ross Sea (Kaufman et al., 2017a) to better understand the spatial and temporal variability of phytoplankton in this region. Assimilation experiments also examine how the space and time characteristics of observational sampling frequency impacts the ability of observations to produce optimal system representations.

## 2 Methods

### 2.1 One-dimensional biogeochemical model

Numerical experiments were conducted with the Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification for the Ross Sea (MEDUSA-RS; Kaufman et al., 2017a), a regionally adapted version of MEDUSA-1.0 (Yool et al., 2011). Three phytoplankton groups are represented in the MEDUSA-RS model: colonial *P. antarctica*, solitary *P. antarctica,* and diatoms. Phytoplankton growth in the model is temperature dependent as well as limited by light and nutrient availability. Colonial *P. antarctica*, diatoms, and detritus all sink at distinct rates. The model handles the sinking of large detrital particles implicitly as a fast-sinking group to avoid issues related to the scale of the model time step and to avoid the need for an additional tracer. A ballast scheme is used to allow inorganic materials to "protect" a variable fraction of the sinking organic material from degradation. The model is configured to focus on dynamics within the euphotic zone with a vertical resolution of 5 m from the ocean surface to 200 m. A full description of the model and its set-up within the Marine Model Optimization Testbed (MarMOT; Hemmings and Challenor, 2012), as well as the physical forcings derived from glider observations, are documented in Kaufman et al. (2017a,b).

### 2.2 Data for assimilation

In situ observations used for the assimilation experiments came from an iRobot Seaglider equipped with a Wet Labs ECO Puck sensor and are available in the Biological and Chemical Oceanography Data Management Office data repository (http://www.bco-dmo.org/dataset/568868). Glider dives from 22 November 2012 to 01 February 2013 covered a horizontal area spanning 76.83 - 77.44 ˚S and 168.9 - 171.97 ˚E (Fig. 1). Data spanning the upper 200 m of the water column were binned by means into hourly, 5-m vertical bins. Concentrations of chlorophyll (Chl) and particulate organic carbon (POC) were derived, respectively, from fluorescence and optical backscatter counts measured by the sensor and converted using regression equations (Kaufman et al., 2017a). These bio-optical quantities were used for calculating model-data misfits during assimilation.

Author
**Deleted:** BCO-DMO

### 2.3 Cost function

The ‘cost function’ (*J*), defined as a measure of misfit between a particular model simulation and observational data, is computed as a weighted average of the squared differences between simulated and observed values:

$$\mathcal{J} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sigma_{chl}^{2}}\left(x_{i,chl} - y_{i,chl}\right)^{2} + \frac{1}{\sigma_{poc}^{2}}\left(x_{i,poc} - y_{i,poc}\right)^{2}\right)$$

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; σ 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).

### 2.4 Cost function minimization

Model parameters were optimized in MarMOT by finding the minimum of the cost function (Sect. 2.3) through a combination of the micro-genetic algorithm (µGA) and Powell’s non-gradient direction set algorithm. The µGA runs first and identifies sets of parameter values that produce low cost values; this is achieved by "evolving" a population of various parameter sets over successive iterations, called generations. The low-cost parameter sets identified by the µGA are then used as starting points for the direction set method, which performs successive linear searches to identify nearby lower cost solutions.

Genetic algorithms, including the µGA, are a subtype of computational methods known as evolutionary algorithms, so-called because of their inspiration from, and metaphoric relationship to, biological evolution. Described using this metaphor, a genetic algorithm procedure modifies a population of candidate solutions over successive generations by variation and selection processes to converge on a single solution or solution area. GAs have several

Author
**Deleted:** , σ is the standard deviation of each variable

Author
**Deleted:** .

Author
**Moved down [1]:** Model parameters were optimized in MarMOT by finding the minimum of this cost function through a combination of the micro-genetic algorithm and Powell’s non-gradient direction set algorithm, as described below.

Author
**Deleted:** **Implementation of micro-genetic algorithm and direction set algorithm**

Author
**Moved (insertion) [1]**

Author
**Deleted:** this

Author
**Deleted:** , as described below

Author
**Deleted:** micro-genetic algorithm [

Author
**Deleted:** ]

advantages for optimization, including their intrinsic parallelism, suitability for systems with multiple local minima, and their generalizability (Bajpai and Kumar, 2010; Ward et al., 2010). The μGA uses three steps to transition from one generation to the next, described following the biological metaphor as: selection, crossover, and resampling (Krishnakumar, 1990; Črepinšek et al., 2013). An advantage of the μGA is its reduced risk of premature convergence, resulting from reinitializating after each convergence, and generating new random populations while maintaining the best fit individual from the previous set (Schmitt, 2001).

In this μGA implementation, optimizations begin with a population of five individual parameter sets randomly generated for the first μGA generation. The constituent parameter values are selected randomly from within a pre-determined range of allowable values (Sect. 2.5.1). An evaluation of the cost function for each model solution indicates the 'fitness' of each individual. A binary tournament procedure is then followed to select parents from this population for the next generation. The most-fit individuals (i.e., those with the lowest cost function values) are paired with one another and undergo recombination of the bits representing parameter values. After each generation, the proportion of bits differing from those of the fittest individual is calculated to determine whether the population can be deemed converged (though this does not necessarily indicate closeness in parameter space). After the threshold for convergence has been achieved, the population is reinitialized to random individuals, although the fittest individual is maintained. The μGA is terminated upon the first convergence occurring after a minimum number of generations has been reached.

Once convergence has been achieved after a minimum number of μGA generations, Powell's non-gradient direction set algorithm performs a local search using the μGA solutions as starting points. The direction set method performs sequential minimizations in iterative directions, updating the search direction after each iteration (Powell, 1964; Press et al., 1992). Although the μGA is well suited for global search problems partly because of its stochasticity, Powell's direction set algorithm is well suited to searching for a local optimum. Brent's method, which combines root-bracketing with secant and inverse quadratic interpolation (Brent, 1973), is used to numerically locate cost minima between neighboring function evaluations along each direction identified by the Powell algorithm. The direction set algorithm stops when a cost function minimum is located or when a maximum number of iterations is reached. The optimized parameter values are those that generated the cost function minimum.

### 2.5 Selection of parameters to be optimized

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). Although the optimization of more parameters generally lowers the assimilated cost, the increasing potential for equifinality with more parameters means the optimization may find equivalent low-cost solutions with substantially different parameter values. Therefore, before assimilating observations and optimizing parameters, a subset of "free" or "optimizable" model parameters must be chosen. In this study, the parameters to be optimized are selected through a three-step process: defining a range of permitted values for every parameter (Sect. 2.5.1), identifying the parameters to which model outputs are most sensitive (Sect. 2.5.2), and evaluating how many of these sensitive parameters can be reasonably optimized when assimilating the available data (Sect. 2.5.3). Initial values for each parameter, prior to the assimilation, were set to values identified in Kaufman et al. (2017a).

Author
**Deleted:** Even though it is tempting to try

Author
**Deleted:** to optimize

Author
**Deleted:** optimized

#### 2.5.1 Parameter ranges

Upper and lower bounds of the allowable range for each free parameter were defined loosely following Hemmings et al. (2015). Bounds were set to be geometrically symmetric (factor of four for rates; factor of five for half-saturation concentrations) around the initial values. For fractional parameter values, limits were set to +/- 0.25 their initial values, although not allowed to exceed 0.05 or 0.95. Ranges for parameters not expressed as fractions were log-transformed for sampling purposes.

#### 2.5.2 Sensitivity Analysis

Parameters to which model outputs are highly sensitive are important and useful to optimize. In contrast, it is futile to optimize parameters to which model outputs of interest are not sensitive; no amount of varying these parameters will result in improved model performance. Therefore, the first criterion used to designate a parameter as optimizable was the sensitivity of model outputs to the values of that parameter. Model sensitivities were evaluated for assimilated variables (Chl and POC) and carbon fluxes of interest (primary production (PP) and carbon export at 200 m). To quantify the sensitivities of these outputs to each of the 80 parameters in the model, a series of runs were conducted following the approach of Hemmings et al. (2015). Each run used a unique sample of parameter values drawn from within the specified parameter ranges

(Section 2.5.1) using a Latin hypercube. This approach provides more even coverage of the parameter space than Monte Carlo sampling methods that can result in clustered values and unsampled regions (Appendix A). One thousand values were drawn from sequential intervals throughout the range for each parameter. Using this technique, unique parameter sets were constructed such that over the course of all runs, the full range of values for each parameter was represented.

The model was run 1000 times, each time using one of the unique parameter sets resulting from Latin hypercube sampling of the full parameter space. Sensitivity was quantified by evaluating the amount of variance in the output diagnostics explained by each parameter (i.e., by computing the coefficient of determination ($r^2$) between each parameter and each of the four output variables of interest; Fig. 2). All four model outputs (Chl, POC, PP, and export) were most sensitive ($r^2 \geq 0.01$) to attenuation of blue-green light by phytoplankton pigments, diatom maximum growth rate, and C:Chl ratio for solitary *P. antarctica*. Three additional parameters (maximum growth rate of *P. antarctica* colonies, maximum growth rate of solitary *P. antarctica*, and microzooplankton maximum grazing rate) exhibited $r^2 \geq 0.01$ for both chlorophyll and POC. The 21 parameters with $r^2 \geq 0.01$ (Fig. 2) were selected for further evaluation (Sect. 2.5.3).

### 2.5.3 Using twin experiments to select optimizable subset

After selecting the 21 potentially optimizable parameters, Numerical Twin Experiments (NTEs) were conducted to identify an optimizable subset by evaluating the extent to which known values of sensitive parameters could be recovered given the data available for assimilation. The implementation of NTEs involves four primary steps (Hofmann and Friedrichs, 2001). First, the chosen model is run forward in time to create a simulation using a known, "true" parameter set. Second, output from this simulation is sub-sampled to create a so-called "synthetic" data set. Third, the synthetic dataset is then assimilated to optimize model parameters. Fourth, the optimized parameter set is compared to the true parameter set. The assimilation is successful if the optimized values recover the true parameters used to generate the assimilated synthetic data.

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

Author
**Deleted:** e

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 (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 µGA 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).

The procedure followed here for determining the subset of optimizable parameters is similar to that used by Friedrichs et al. (2007). First, a reference simulation was generated using the initial parameter set, and chlorophyll and POC estimates from this reference simulation were subsampled to generate a synthetic data set. Starting with a parameter space defined by the set of 21 parameters deemed sensitive in the Latin hypercube tests (Fig. 2), a series of sequential NTEs was then performed with a progressively smaller number of optimized parameters: after each NTE, the optimized parameter that was most different from its 'true' value was removed from the optimizable parameter set. Thus, after each NTE the number of optimized parameters was reduced by one. The series of NTEs was evaluated to identify the largest parameter set for which the original parameter values were recoverable and the cost function remained essentially zero. From this analysis (Fig. 3), it was determined that optimizing eight parameters would be ideal (Table 1), because values of these eight parameters were recovered much better than larger parameter sets and model-data misfit (cost) remained low.

### 2.6 Assimilation experiments

The µGA optimization procedure was used to assimilate glider data in two sets of experiments that explored aspects of spatiotemporal variability and data availability. Estimates of depth- and time-integrated PP and time-integrated carbon export at 200 m were computed from the full model simulation in each experiment.

#### 2.6.1 Experiment #1

The first set of experiments examined the differences in model simulations resulting from assimilating Chl and POC data from different spatial regions. In Experiment #1a, glider observations were assimilated from the upper 50 m of the full temporal and spatial domain, referred to hereafter as the “Full Assimilation” case (Table 2). (Comparisons showed only minor differences between assimilating data from the upper 50 m vs. the full upper 200 m). Observations from different spatial areas of the glider track were also assimilated. Observations from the glider track were divided into three latitudinal bands (Northern, Central, and Southern bands) as well as into three longitudinal areas constituting Eastern, Central, and Western bands. Glider data from each of these three latitudinal and longitudinal bands were assimilated in Experiment #1b and #1c, respectively (Table 2, Fig. 4), resulting in three cost functions for each of these esperiments.

#### 2.6.2 Experiment #2

The second set of experiments investigated the assimilation of data at different resolutions mimicking different data sources. In Experiment #2a, glider data were subsampled at ~12-hour intervals (Table 2). The subsampling was repeated 12 times, with each iteration offset from the previous by +1, 2, 3... 11 hours, to generate a series of 12 glider observation sets. The assimilation of these 12 time series yields the “Glider Assimilation” case. In Experiment #2b, glider data were subsampled at a reduced temporal resolution similar to cruise sampling (Table 2). Sampling during cruise missions often takes place for a few days in one location before moving elsewhere, and the ship sometimes returns to the first location after a number of weeks. To roughly mimic this sampling pattern, daily vertical profiles (again down to 50 m) were assimilated for three days in a row, starting from the first day of available glider data (22 Nov), and then three days of data were assimilated two weeks later. Shifting this pattern forward one week at a time generated a series of eight cruise-based observation sets for assimilation in this “Cruise Based Assimilation” case. In Experiment #2c, glider data were assimilated only from the upper 5 m surface layer to produce a data set resembling satellite-derived data. These data were then subsampled at two-week intervals, to represent typical data return from remote sensing observations of ocean color in the Ross Sea, where the availability of satellite image retrieval is frequently limited by excessive, though variable, cloud cover (Arrigo and van Dijken, 2004). The two-week subsampling pattern covered the entire period of glider data (22 Nov - 1 Feb), and was sequentially shifted forward one day at a time to generate a series of 14 satellite-based observation sets for assimilation in this “Satellite-Based Assimilation” case (Table 2).

### 2.7 Predictive Cost Assessment

In addition to the assimilative cost ($\mathcal{J}_A$) calculated during the optimization procedure using assimilated data, a predictive cost ($\mathcal{J}_P$) was calculated to assess model-data misfit computed using the unassimilated data in each experiment. Because predictive costs represent model-data misfit from unassimilated data only (Friedrichs et al., 2006; Ward et al., 2010), it is an objective measure of the skill of an optimized model in reproducing observations at different points in time or space (Gregg et al., 2009). In this case, the aim of these experiments is to assess the skill of each optimized simulation regardless of which subset of the available data is assimilated. By computing the mean and median predictive cost for each experiment (other than the Full Assimilation case), the skill of the resulting simulations can be compared directly with one another.

## 3 Results

### 3.1 Experiment #1

Assimilation of the glider data over the full temporal and spatial domain (Full Assimilation case) improves the model-data fit of both Chl and POC (Fig. 5a,b) and reduces the cost by nearly half (47%) compared to the a priori simulation without assimilation (Table 3). Average Chl and POC concentrations in the upper 50 m are both slightly lower (8% and 12%, respectively) in the optimized simulation. The contribution of each phytoplankton group to total chlorophyll remains similar to the No Assimilation case (Fig. 6a,c), but colonial *P. antarctica* carbon is lower and diatom carbon is higher in December and early January (Fig. 6b,d). Compared to the No Assimilation case, PP is only slightly lower (7%), whereas export flux is nearly 50% higher (Table 3; Fig. 7). Compared to their initial values, colonial *P. antarctica* parameters change the most as a result of the optimization, with reductions between 40-70% for the colonial *P. antarctica* maximum growth rate, maximum sinking rate and C:Chl ratio (Table 4). In contrast, the diatom maximum growth rate and C:Chl ratio increased (~10% and 20% respectively).

Chlorophyll and POC time-series exhibit only minor differences between latitudinal band experiments when data from the northern, central, and southern sections are assimilated independently (Fig. 5c,d) or when data from the eastern, central western sections are assimilated (Fig. 5e,f). Specifically, the optimal simulations for Chl and POC exhibit similar seasonal cycles across the three latitudinal and longitudinal bands, with only slightly higher Chl and POC concentrations when assimilating data from the southern band (Fig. 5c,d) and higher Chl from the western band (Fig. 5e,f). Mean costs are much lower for the latitudinal and longitudinal experiments than for the No Assimilation case, and only slightly higher than the Full Assimilation case (Table 3). This indicates that data sampled from within only one spatial band improved the match between modeled and observed variables in the unassimilated areas as well. Average estimates of PP and export in both the latitudinal and longitudinal experiments are only slightly less (< 5%) than the Full Assimilation estimate (Fig. 7, Table 3).

### 3.2 Experiment #2

Assimilation of data subsampled at a frequency one-twelth that of the original glider data (Expt. 2a) results in twelve model simulations, all of which are similar to the Full Assimilation case, with Chl and POC time series closely following the observed seasonal pattern (Fig. 8a,b). Mean assimilative and predictive costs in the Glider Assimilation case are close to the cost of the Full Assimilation case (Table 3). Mean PP and export estimates are also close to estimates from the Full Assimilation case. The mean optimal parameter values obtained from the Glider Assimilation case are generally within one standard deviation of the optimal values from the Full Assimilation case (Table 4).

Assimilation of data subsampled with a frequency typical of cruise observations (Expt. 2b) results in a wide range of solutions, with several Chl and POC time series exhibiting markedly different peak bloom timings (Fig. 8c,d). Two of the solutions yield substantially higher concentrations of POC in November, and Chl peaks range from mid-November to early January. The mean predictive cost from this experiment (1.24) is roughly three times the assimilative cost for the Full Assimilation case (0.41) and three times the predictive cost for the Glider Assimilation case (0.43; Table 3). The PP estimates from the Cruise-based Assimilation case span a broad range (92 to 156 g C $m^{-2}$ $y^{-1}$) around the Full Assimilation estimate but are generally higher (Fig. 7a). This experiment similarly yields a very large range of export estimates (11 to 33 g C $m^{-2}$ $y^{-1}$) encompassing the results from Experiment #1 (Fig. 7b). Optimal parameter values obtained from the Cruise-based Assimilation case are generally less well constrained (higher standard deviations) than the Glider Assimilation case (Table 4).

Assimilation of data subsampled as satellite-based observations from the surface layer (Expt. 2c) results in Chl and POC concentrations generally higher than the Full Assimilation case (Fig. 8e,f). The predictive costs are similar on average to those of the Cruise-based Assimilation experiment; however, there is less variation (Table 3). The median integrated production is higher (9%) than the Full Assimilation estimate and the Cruised-based Assimilation estimate (Fig. 7a; Table 3); however, the range of PP estimates for this Satellite-based Assimilation case is smaller than those for the Cruise-based Assimilation case (Fig. 7a). Most notably, despite generally higher PP and higher POC concentrations, carbon export from the Satellite-based Assimilation case is substantially lower (41%) than the Full Assimilation estimate (Fig. 7b; Table 3). In fact, export estimates from individual runs in this experiment are all lower (-19% to -56%) than the Full Assimilation estimate (Fig. 7b). Again, the range of export estimates is smaller for the Satellite-based Assimilation than for the Cruise-based Assimilation. When assimilating data at a resolution similar to that of satellite-based observations, mean optimal parameter values were similar to those obtained in the Glider Assimilation and Cruise-based Assimilation cases, with the exception of the fast detritus sinking fraction for diatoms, which was significantly lower in the Satellite-based Assimilation case (0.62 ± 0.14) than in the other experiments (Glider Based Assimilation Case: 0.86 ± 0.05). In contrast to this sinking parameter for mortality from diatoms, the mean maxmimum sinking rate of colonial *P. antarctica* in the Satellite-based case was not significantly different than its value in either the Full Assimilation or Cruise-based cases (Table 4). Standard deviations of optimal parameters for the Satellite-based Assimilation case were generally similar to or lower than those for the Cruise-based Assimilation case, except for the C:Chl ratio for diatoms, which produced a very high optimal value and was particularly poorly constrained (375±187 gC $\mathrm{gChl^{-1}}$; Table 4).

## 4 Discussion

### 4.1 Ross Sea simulation resulting from assimilation of glider data

Data assimilation is a valuable tool for efficiently utilizing limited observational data in remote regions like the Ross Sea. In this study, glider data consisting of both fluorescence-derived chlorophyll and backscatter-derived POC were assimilated into a one-dimensional marine biogeochemical model developed for the Ross Sea. Eight ecosystem parameters, including phytoplankton rates and C:Chl ratios, were optimized resulting in a simulation with a 50% reduced model-data misfit. This optimal simulation yielded lower *P. antarctica* carbon concentrations and higher diatom carbon concentrations, resulting in higher carbon export compared to those generated by the initial hand-tuned simulation (Kaufman et al., 2017a), despite slightly lower estimates of overall annual primary production. Changes in chlorophyll concentrations of diatoms and *P. antarctica* were minor. This optimal simulation was obtained largely via changes in the C:Chl ratios: the colonial *P. antarctica* ratio of C:Chl was lower and the diatom C:Chl was higher than in the original simulation. Although modified from their initial values, the relative differences between these optimized C:Chl ratios for *P. antarctica* and diatoms are consistent with shipboard measurements of C:Chl ratios, which found higher C:Chl in diatom-dominated waters compared to *P. antarctica*-dominated waters: ~200 vs. 90 g C g Chl$^{-1}$ (DiTullio and Smith, 1996), and ~50-100 vs. 20-50 g C g Chl$^{-1}$ (Mathot et al., 2000). Although the authors are not aware of any specific estimates in the literature for the fraction of diatom mortality that becomes fast-sinking detritus, other optimal rate paramaters are consistent with those previously reported in the literature. For example, the optimized growth rates (0.29 - 0.4 m d$^{-1}$) are similar to measured values in the Ross Sea (Smith and Gordon, 1997; Smith et al., 1999; Mosby and Smith, 2015), and the optimized sinking rate of *P. antarctica* colonies (14 m d$^{-1}$) is similar to previous estimates (Asper and Smith, 1999; Asper and Smith, 2003; Smith et al., 2011).

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).

### 4.2 Spatial variation within the glider track

Phytoplankton in the Ross Sea exhibit both spatial and temporal variability. Cruise transects across the continental shelf show a marked spatial variability in both the east-west and north-south direction over short periods of time (Smith et al., 2013). Within the Ross Sea Polynya, ship-based observations show biochemical gradients that suggest patchiness of phytoplankton dynamics on the mesoscale (Hales and Takahashi, 2004; Smith et al., 2017).

Nutrient pools have been found to exhibit gradients from both north to south and east to west (DiTullio and Smith, 1996; Sedwick et al., 2011; Smith et al., 2013; Marsay et al., 2014), and phytoplankton assemblage composition is not necessarily uniform across longitudes (DiTullio and Smith, 1996; Garrison et al., 2003; Smith et al., 2013). In addition, cold and fresh eddies have been observed along the ice shelf edge potentially reshaping the phytoplankton assemblage on short time (<10 days) and space (<20 km) scales (Li et al., 2017).

When analyzing glider data in regions characterized by high mesoscale variability, it is often not apparent whether observed patterns represent spatial or temporal variability. As Rudnick (2016) discusses, "Because gliders can occupy lines, their data can be viewed as traditional sections, such as those measured from a ship. However, because high-frequency variability is projected onto a spatial structure, it is sometimes more convenient to think of the data as a time series from a mooring." This ambiguity led Kaufman et al. (2014) to concede "both spatial and temporal gradients may have played a role in the observed variability" when analyzing physical-biological relationships from glider data in the southern Ross Sea.

Although both temporal and spatial gradients may be present, observations can be presented as either primarily spatial or temporal patterns with simple tests guiding the decision. For example, a comparison of means and standard deviations across spatial sections and time periods was previously used to identify time as the dominant dimension of variability in the 2012-2013 glider observations (Jones and Smith, 2017). In this study, a similar conclusion was reached, using a very different methodology. The assimilation of glider data from six different sub-areas of the study region (separated latitudinally or longitudinally by ~20 km) indicated that the seasonal cycle is similar in phase throughout the region of the glider track. The assimilation of glider data from each of the nine regions yielded similar estimates of POC and Chl, generally within the variance of the glider observations (gray areas of Fig. 5c-f), and similar estimates of temporally averaged primary productivity and export. This further supports the approach of using the glider data as a time series and suggests that temporal patterns at this scale play a greater role than spatial patterns in structuring variability of the phytoplankton assemblage. Moreover, the similarity between predictive and assimilative costs when assimilating the latitudinal and longitudinal bands of data suggests that the parameters are not being over-fit for these experiments. Thus, temporally resolved observations in any of these regions might be expected to provide similar constraints on modeled temporal patterns of the phytoplankton.

Author
**Deleted:** nine

### 4.3 Differences between assimilating glider, satellite-derived, and cruise-based data

Results from experiments that assimilated data at different spatial and temporal resolutions suggest that assimilating only surface observations, as are typically available from remote-sensing platforms, underestimates carbon export and more weakly constrains estimates of productivity relative to assimilation of depth-resolved glider data. The lower estimates of carbon export occurred because the optimal diatom fraction for fast-sinking detritus obtained via the assimilation of surface-only data (0.62 ± 0.14) was significantly lower than that obtained via the assimilation of data throughout the upper 50 m (Expt. 2a: 0.86 ± 0.05; Expt. 2b: 0.86 ± 0.11). These results highlight the importance of assimilating subsurface measurements and of modeling diatom aggregation when estimating carbon export; similar findings were reported in 1D biogeochemical optimization experiments using data from Lagrangian floats in the North Atlantic (Bagniewski et al., 2011). Experimental results also indicate that the assimilation of satellite-derived data provides a weaker constraint on productivity estimates, as seen by the larger range of estimates (114 ± 11 gC $m^{-2}$ $y^{-1}$), as compared to the assimilation of glider data (104 ± 2 gC $m^{-2}$ $y^{-1}$). Although not statistically significant, the higher productivity estimates generated by the assimilation of satellite-derived data is consistent with those of Gregg (2008), who found that assimilation of satellite-based chlorophyll estimates into a three-dimensional global biogeochemical model overestimated primary production. In contrast, results from assimilating satellite-derived chlorophyll concentrations into a one-dimensional model in the equatorial Pacific produced underestimates of primary productivity compared to in situ observations (Friedrichs, 2002).

Although both chlorophyll and POC were assimilated in the present study, chlorophyll alone has been the dominant satellite data product used in biogeochemical assimilation, although other data types are available and can impact the assimilation results. For instance, a study investigating the assimilation of different types of satellite-derived data, including POC and size-fractionated chlorophyll, found that assimilation of satellite-derived POC estimates worsened the model estimates of chlorophyll, whereas the assimilation of chlorophyll did not substantially impact the POC estimates (Xiao and Friedrichs, 2014b). Additionally, satellite-based sampling bias could be reduced by concurrently assimilating export flux data derived from sediment trap measurements (Friedrichs et al., 2007), or by assimilating satellite measurements such as remote-sensing reflectance directly (Jones et al., 2016). It is also worth noting that when assimilating actual satellite data, the biases suggested by this study resulting from assimilation of only surface data would be compounded with biases inherent in the satellite retrieval algorithms (Saba et al., 2011; Stukel et al., 2015).

Assimilating cruise-based data in the highly variable Ross Sea may also yield potentially large errors in primary production, as well as in carbon export estimates, depending on which specific days are sampled. Estimates of bloom timing from the assimilation of cruise-based observations may also vary substantially (Fig. 8c,d). This echoes the results of a series of reduced resolution data interpolations, from which Hales and Takahashi (2004) reported that cruise-based observations in the Ross Sea were likely able to capture average conditions well, but miss some mesoscale phenomena. Likewise, a subsampling analysis of physical-biological correlations from 2010 Ross Sea glider data demonstrated the possibility of lower resolution data obscuring or biasing biogeochemical interpretations (Kaufman et al., 2014). The results provided by the data assimilative study described here can be used to help guide decisions of when and how long to sample certain locations in the Ross Sea; this is especially important given the limitations of ship-based sampling in such a remote region (Smith et al., 2014). In fact, the use of data collection from other sampling platforms may decrease the pressure to conduct repeated transects by ship, and allow limited vessel-time to be used for more thorough process-based investigations uniquely-suited for research vessels.

## 5 Summary and Conclusions

A series of experiments investigating spatiotemporal variability of the phytoplankton assemblage and potential effects of assimilating data from different observation platforms highlighted the benefits and challenges of combining data and biogeochemical models in the Ross Sea. The assimilation of glider data reduced model-data misfit by 50%, and resulted in reduced depth-integrated primary production and higher carbon export at 200 m. Additional experiments for different spatial regions reduced predictive costs with respect to unassimilated data by ~35%, suggested that the model parameters were well constrained, and implied that using glider data as time series in these local studies is a reasonable approach. This may further suggest the value of using moorings or buoys, or even deploying gliders in a “virtual mooring” mode. However, the effects of mesoscale variability were apparent when assimilating data at a frequency characteristic of cruise-based sampling, which resulted in a wide range of primary production and export estimates depending on the sampling times. Results of assimilating data characteristic of satellite-based sampling suggest that assimilating satellite-derived data will result in underestimated carbon export. These findings can be used to help avoid potential sources of error when using ship-based or satellite-based observations alongside the development, calibration, or running of biogeochemical models. The combination of high-resolution glider data and modeling in this study underscores the importance of considering how the timing at which observations are collected affect the subsequent interpretations.

Author

**Deleted:** Ross Sea

Author

**Deleted:** , and it is believed that further comparison of data from additional autonomous gliders throughout the Ross Sea and between years will provide greater clarity of the spatiotemporal variability on multiple scales.

*Data Availability.* Data from the autonomous glider are available from the BCO-DMO data repository (http://www.bco-dmo.org/dataset/568868), and other data to support this article are available at W&M Publish (https://doi.org/10.21220/V5RT5C) and upon request from the authors (dkauf42@gmail.com, marjy@vims.edu).

Author
**Deleted:** will be made

Author
**Deleted:** http://doi.org/xxxx

Author
**Deleted:** dekaufman@vims.edu

## Appendix A: Latin hypercube sampling (Sect. 2.5.2)

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.

*Competing Interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This material is based upon work supported by the U.S. National Science Foundation's Office of Polar Programs (NSF-ANT-0838980). The authors thank Drs. Elizabeth A. Canuel, Eileen E. Hofmann, and Elizabeth H. Shadwick for constructive comments. Additional thanks go to Michael S. Dinniman for helping with model forcings. This work was performed (in part) using computational facilities at the College of William and Mary which were provided by contributions from the National Science Foundation, the Commonwealth of Virginia Equipment Trust Fund and the Office of Naval Research. This

paper is contribution 3675 of the Virginia Institute of Marine Science, College of William and Mary.

Author
**Deleted:** XXXX

## Literature Cited

Arrigo, K. R., and McClain, C. R.: Spring phytoplankton production in the western Ross Sea, *Science*, *266*(5183), 261–263, doi:10.1126/science.266.5183.261, 1994.

Arrigo, K. R., Robinson, D. H., Worthen, D. L., Schieber, B., and Lizotte, M. P.: Bio-optical properties of the southwestern Ross Sea, *J. Geophys. Res. Ocean.*, *103*(C10), 21683–21695, doi:10.1029/98JC02157, 1998.

Arrigo, K. R., and van Dijken, G. L.: Annual changes in sea-ice, chlorophyll a, and primary production in the Ross Sea, Antarctica, *Deep Sea Res. II*, *51*(1–3), 117–138, doi:10.1016/j.dsr2.2003.04.003, 2004.

Arrigo, K. R., van Dijken, G. L., and Bushinsky, S.: Primary production in the Southern Ocean, 1997–2006, *J. Geophys. Res.*, *113*(C08004), 1–27, doi:10.1029/2007JC004551, 2008.

Asper, V. L., and Smith, W. O.: Particle fluxes during austral spring and summer in the southern Ross Sea, Antarctica, *J. Geophys. Res.*, *104*(C3), 5345–5359, doi:10.1029/1998JC900067, 1999.

Asper, V. L., and Smith, W. O.: Abundance, distribution and sinking rates of aggregates in the Ross Sea, Antarctica, *Deep. Res. Part I Oceanogr. Res. Pap.*, *50*(1), 131–150, doi:10.1016/S0967-0637(02)00146-2, 2003.

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.

Bajpai, P., and Kumar, M.: Genetic algorithm–an approach to solve global optimization problems, *Indian J. Comput. Sci. Eng.*, *1*(3), 199–206, 2010.

Brent, R. P.: "An algorithm with guaranteed convergence for finding a minimum of a function of one variable." Algorithms for Minimization Without Derivatives, pp. 61–80, Prentice-Hall, Inc, 1973.

Črepinšek, M., Liu, S.-H. and Mernik, M.: Exploration and Exploitation in Evolutionary Algorithms: A Survey, *ACM Comput. Surv.*, *45*(3), 1–33, doi:10.1145/2480741.2480752, 2013.

DiTullio, G. R., and Smith, W. O.: Spatial patterns in phytoplankton biomass and pigment distributions in the Ross Sea, *J. Geophys. Res.*, *101*(C8), 18467–18477. doi: 10.1029/96JC00034, 1996.

Doron, M., Brasseur, P., Brankart, J.-M., Losa, S. N., and Melet, A.: Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical–biogeochemical model, *J. Mar. Syst.*, *117–118*, 81–95, doi:10.1016/j.jmarsys.2013.02.007, 2013.

El-Sayed, S. Z., Biggs, D. C., Stockwell, D., Warner, R., and Meyer, M.: Biogeography and metabolism of phytoplankton and zooplankton in the Ross Sea, Antarctica, *Antarct. J. U.S.*, 131–133, 1978.

Fennel, K., Losch, M., Schröter, J., and Wenzel, M.: Testing a marine ecosystem model: Sensitivity analysis and parameter optimization, *J. Mar. Syst.*, *28*(1–2), 45–63, doi:10.1016/S0924-7963(00)00083-X, 2001.

Fretwell, P., Pritchard, H. D., Vaughan, D. G., Bamber, J. L., Barrand, N. E., Bell, R., Bianchi, C., Bingham, R. G., Blankenship, D. D., Casassa, G., Catania, G., Callens, D., Conway, H., Cook, A. J., Corr, H. F. J., Damaske, D., Damm, V., Ferraccioli, F., Forsberg, R., Fujita, S., Gim, Y., Gogineni, P., Griggs, J. A., Hindmarsh, R. C. A., Holmlund, P., Holt, J. W., Jacobel, R. W., Jenkins, A., Jokat, W., Jordan, T., King, E. C., Kohler, J., Krabill, W., Riger-Kusk, M., Langley, K. A., Leitchenkov, G., Leuschen, C., Luyendyk, B. P., Matsuoka, K., Mouginot, J., Nitsche, F. O., Nogi, Y., Nost, O. A., Popov, S. V., Rignot, E., Rippin, D. M., Rivera, A., Roberts, J., Ross, N., Siegert, M. J., Smith, A. M., Steinhage, D., Studinger, M., Sun, B., Tinto, B. K., Welch, B. C., Wilson, D., Young, D. A., Xiangbin, C., and Zirizzotti, A.: Bedmap2: improved ice bed, surface and thickness datasets for Antarctica, *Cryosph.*, *7*(1), 375–393, doi:10.5194/tc-7-375-2013, 2013.

Friedrichs, M. A. M.: A data assimilative marine ecosystem model of the central equatorial Pacific: Numerical twin experiments, *J. Mar. Res.*, *59*, 859–894. doi:10.1357/00222400160497544, 2001.

Friedrichs, M. A. M.: Assimilation of JGOFS EqPac and SeaWiFS data into a marine ecosystem model of the central equatorial Pacific Ocean, *Deep-Sea Res. II*, *49*, 289–319. doi:10.1016/S0967-0645(01)00104-7, 2002.

Friedrichs, M. A. M., Hood, R. R., and Wiggert, J. D.: Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data, *Deep-Sea Res. II*, *53*(5–7), 576–600, doi:10.1016/j.dsr2.2006.01.026, 2006.

Friedrichs, M. A. M., Dusenberry, J. A., Anderson, L. A., Armstrong, R. A., Chai, F., Christian, J. R., Doney, S. C., Dunne, J., Fujii, M., Hood, R., McGillicuddy Jr., D. J., Moore, J. K., Schartau, M., Spitz, Y. H., Wiggert, J. D.: Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups, *J. Geophys. Res. Ocean.*, *112*(C08001), 1–22, doi:10.1029/2006JC003852, 2007.

Garrison, D. L., Gibson, A., Kunze, H., Gowing, M. M., Vickers, C. L., Mathot, S., and Bayre, R. C.: The Ross Sea Polynya Project: Diatom- and Phaeocystis-dominated phytoplankton assemblages in the Ross Sea, Antarctica, 1994-1996, in *Biogeochemistry of the Ross Sea, Antarctic Research Series*, vol. 78, pp. 53–76, 2003.

Gharamti, M. E., Samuelsen, A., Bertino, L., Simon, E., Korosov, A., and Daewel, U.: Online tuning of ocean biogeochemical model parameters using ensemble estimation techniques: Application to a one-dimensional model in the North Atlantic, *J. Mar. Syst.*, *168*, 1–16, doi:10.1016/j.jmarsys.2016.12.003, 2017.

Gregg, W. W.: Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model, *J. Mar. Syst.*, *69*, 205–225, doi:10.1016/j.jmarsys.2006.02.015, 2008.

Gregg, W. W., Friedrichs, M. A. M., Robinson, A. R., Rose, K. A., Schlitzer, R., Thompson, K. R., and Doney, S. C.: Skill assessment in ocean biological data assimilation, *J. Mar. Syst.*, *76*(1–2), 16–33, doi:10.1016/j.jmarsys.2008.05.006, 2009.

Hales, B., and Takahashi, T.: High-resolution biogeochemical investigation of the Ross Sea, Antarctica, during the AESOPS (U. S. JGOFS) Program, *Global Biogeochem. Cycles*, *18*, 1–24, doi:10.1029/2003GB002165, 2004.

Hemmings, J. C. P., Srokosz, M. A., Challenor, P., and Fasham, M. J. R.: Split-domain calibration of an ecosystem model using satellite ocean colour data, *J. Mar. Syst.*, *50*(3–4), 141–179, doi:10.1016/j.jmarsys.2004.02.003, 2004.

Hemmings, J. C. P., and Challenor, P. G.: Addressing the impact of environmental uncertainty in plankton model calibration with a dedicated software system: the Marine Model Optimization Testbed (MarMOT 1.1 alpha), *Geosci. Model Dev.*, *5*(2), 471–498, doi:10.5194/gmd-5-471-2012, 2012.

Hemmings, J. C. P., Challenor, P. G., and Yool, A.: Mechanistic site-based emulation of a global ocean biogeochemical model (MEDUSA 1.0) for parametric analysis and calibration: an application of the Marine Model Optimization Testbed (MarMOT 1.1), *Geosci. Model Dev.*, *8*(3), 697–731, doi:10.5194/gmd-8-697-2015, 2015.

Hofmann, E. E., and Friedrichs, M. A. M.: Biogeochemical Data Assimilation, in *Encyclopedia of Ocean Sciences*, edited by J. H. Steele et al. pp. 364–370, Elsevier, 2001.

Hofmann, E. E., and Friedrichs, M. A. M.: Predictive Modeling for Marine Ecosystems, in *The Sea, Volume 12: Biological-Physical Interactions in the Sea*, edited by A. R. Robinson, J. J. McCarthy, and B. J. Rothschild, pp. 537–565, John Wiley & Sons, Inc., New York, 2002.

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.

Jones, E. M., Baird, M. E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., Margvelashvili, N., Matear, R. J., Wild-Allen, K., Robson, B., Rizwi, F., Oke, P., King, E., Schroeder, T., Steven, A., Taylor, J.: Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef, *Biogeosciences*, *13*, 6441–6469, doi:10.5194/bg-13-6441-2016, 2016.

Jones, R. M., and Smith, W. O.: The influence of short-term events on the hydrographic and biological structure of the southwestern Ross Sea, *J. Mar. Syst.*, *166*, 184–195, doi:10.1016/j.jmarsys.2016.09.006, 2017.

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.

Kaufman, D. E., Friedrichs, M. A. M., Smith, W. O., Queste, B. Y., and Heywood, K. J.: Biogeochemical variability in the southern Ross Sea as observed by a glider deployment, *Deep Sea Res. I*, *92*, 93–106, doi:10.1016/j.dsr.2014.06.011, 2014.

Kaufman, D. E., Friedrichs, M. A. M., Smith, W. O., Hofmann, E. E., Dinniman, M. S., and Hemmings, J. C. P.: Climate change impacts on southern Ross Sea phytoplankton composition, productivity, and export, *J. Geophys. Res. Ocean.*, *122*(3), 2339–2359, doi:10.1002/2016JC012514, 2017a.

Kaufman, D.E., Friedrichs, M.A.M., Smith, W.O., Jr., Hofmann, E.E., Dinniman, M.S., Hemmings, J.C.P.: Associated Dataset: Climate change impacts on southern Ross Sea phytoplankton composition, productivity and export. Virginia Institute of Marine Science. College of William and Mary. http://doi.org/10.21220/V5PC71, 2017b.

Krishnakumar, K.: Micro-genetic algorithms for stationary and non-stationary function optimization, *Proc. SPIE. 1196, Intell. Control Adapt. Syst.*, *289*, doi:10.1117/12.969927, 1990.

Lawson, L. M., Spitz, Y. H., Hofmann, E. E., and Long, R. B.: A data assimilation technique applied to a predator-prey model, *Bull. Math. Biol.*, *57*(4), 593–617, doi:10.1007/BF02460785, 1995.

Lawson, L. M., Hofmann, E. E., and Spitz, Y. H.: Time series sampling and data assimilation in a simple marine ecosystem model, *Deep-Sea Res. II*, *43*(2–3), 625–651, doi:10.1016/0967-0645(95)00096-8, 1996.

Leonelli, S.: “The Impure Nature of Biological Knowledge and the Practice of Understanding.” Scientific Understanding: Philosophical Perspectives, Pittsburgh University Press, 1–27, 2009.

Li, Y., McGillicuddy Jr., D. J., Dinniman, M. S., and Klinck, J. M.: Processes influencing formation of low-salinity high-biomass lenses near the edge of the Ross Ice Shelf, *J. Mar. Syst.*, *166*, 108–119, doi:10.1016/j.jmarsys.2016.07.002, 2017.

Little, H.: Quantifying spatial and temporal scales of phytoplankton variability in the Sub-Antarctic Ocean using a high-resolution glider dataset, University of Cape Town, Thesis, 2016.

Löptien, U., and Dietze, H.: Constraining parameters in marine pelagic ecosystem models – is it actually feasible with typical observations of standing stocks?, *Ocean Sci.*, *11*(4), 573–590, doi:10.5194/os-11-573-2015, 2015.

Marsay, C. M., Sedwick, P. N., Dinniman, M. S., Barrett, P. M., Mack, S. L., and McGillicuddy Jr., D. J.: Estimating the benthic efflux of dissolved iron on the Ross Sea continental shelf, *Geophys. Res. Lett.*, *41*, 7576–7583, doi:10.1002/2014GL061684, 2014.

Matear, R. J.: Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P, *J. Mar. Res.*, *53*, 571–607, doi:10.1357/0022240953213098, 1995.

Mathot, S., Smith, W. O., Carlson, C. A., Garrison, D. L., Gowing, M. M., and Vickers, C. L.: Carbon partitioning within *Phaeocystis antarctica* (Prymnesiophyceae) colonies in the Ross Sea, Antarctica, *J. Phycol.*, *36*, 1049–1056, doi:10.1046/j.1529-8817.2000.99078.x, 2000.

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.

Melbourne-Thomas, J., Wotherspoon, S., Corney, S., Molina-Balari, E., Marini, O., and Constable, A.: Optimal control and system limitation in a Southern Ocean ecosystem model, *Deep. Res. Part II Top. Stud. Oceanogr.*, *114*, 64–73, doi:10.1016/j.dsr2.2013.02.017, 2015.

Mosby, A., and Smith, W. O.: Phytoplankton growth rates in the Ross Sea, Antarctica, *Aquat. Microb. Ecol.*, *74*, 157–171, doi:10.3354/ame01733, 2015.

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.

Peloquin, J. A., and Smith, W. O.: Phytoplankton blooms in the Ross Sea, Antarctica: Interannual variability in magnitude, temporal patterns, and composition, *J. Geophys. Res.*, *112*(C08013), 1–12, doi:10.1029/2006JC003816, 2007.

Powell, M. J. D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives, *Comput. J.*, *7*(2), 155–162, doi:10.1093/comjnl/7.2.155, 1964.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipes in C: the art of scientific computing, 2nd ed., Cambridge University Press, 1992.

Rudnick, D. L.: Ocean Research Enabled by Underwater Gliders, *Ann. Rev. Mar. Sci.*, *8*(1), 519–541, doi:10.1146/annurev-marine-122414-033913, 2016.

Saba, V. S., Friedrichs, M. A. M., Antoine, D., Armstrong, R. A., Asanuma, I., Behrenfeld, M., J. Ciotti, A. M., Dowell, M., Hoepffner, N., Hyde, K. J. W., Ishizaka, J., Kameda, T., Marra, J., Mélin, F., Morel, A., O'Reilly, J., Scardi, M., Smith, W. O., Smyth, T. J., Tang, S., Uitz, J., Waters, K., and Westberry, T. K.: An evaluation of ocean color model estimates of marine primary productivity in coastal and pelagic regions across the globe, *Biogeosciences*, *8*(2), 489–503, doi:10.5194/bg-8-489-2011, 2011.

Schartau, M., and Oschlies, A.: Simultaneous data-based optimization of a 1D-ecosystem model at three locations in the North Atlantic: Part II - Standing stocks and nitrogen fluxes, *J. Mar. Res.*, *61*(6), 795–821, doi:10.1357/002224003322981156, 2003.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses : parameter identification in marine planktonic ecosystem modelling, *Biogeosciences*, *14*, 1647–1701, doi:10.5194/bg-14-1647-2017, 2017.

Schine, C. M. S., van Dijken, G., and Arrigo, K. R.: Spatial analysis of trends in primary production and relationships with large-scale climate variability in the Ross Sea, Antarctica (1997-2013), *J. Geophys. Res. Ocean.*, *120*, 1–19, doi:10.1002/2015JC011014, 2015.

Schmitt, L. M.: Theory of genetic algorithms, *Theor. Comput. Sci.*, *259*(1–2), 1–61, doi:10.1016/S0304-3975(00)00406-0, 2001.

Sedwick, P. N., Marsay, C. M., Sohst, B. M., Aguilar-Islas, A. M., Lohan, M. C., Long, M. C., Arrigo, K. R., Dunbar, R. B., Saito, M. A., Smith, W. O., and DiTullio, G. R.: Early season depletion of dissolved iron in the Ross Sea polynya: Implications for iron dynamics on the Antarctic continental shelf, *J. Geophys. Res.*, *116*(C12019), 1–19, doi:10.1029/2010JC006553, 2011.

Smith, W. O., and Nelson, D. M.: Phytoplankton bloom produced by a receding ice edge in the Ross Sea: spatial coherence with the density field, *Science*, *227*(8), 163–166, doi:10.1126/science.227.4683.163, 1985.

Smith, W. O., and Gordon, L. I.: Hyperproductivity of the Ross Sea (Antarctica) polynya during austral spring, *Geophys. Res. Lett.*, *24*(3), 233–236, doi:10.1029/96GL03926, 1997.

Smith, W. O., Nelson, D. M., and Mathot, S.: Phytoplankton growth rates in the Ross Sea, Antarctica, determined by independent methods: temporal variations, *J. Plankton Res.*, *21*(8), 1519–1536. doi:10.1093/plankt/21.8.1519, 1999.

Smith, W. O., Dinniman, M. S., Tozzi, S., DiTullio, G. R., Mangoni, O., Modigh, M., and Saggiomo, V.: Phytoplankton photosynthetic pigments in the Ross Sea: Patterns and relationships among functional groups, *J. Mar. Syst.*, *82*(3), 177–185, doi:10.1016/j.jmarsys.2010.04.014, 2010.

Smith, W. O., Shields, A. R., Dreyer, J. C., Peloquin, J. A., and Asper, V.: Interannual variability in vertical export in the Ross Sea: Magnitude, composition, and environmental correlates, *Deep-Sea Res. I*, *58*, 147–159, doi:10.1016/j.dsr.2010.11.007, 2011.

Smith, W. O., Tozzi, S., Long, M. C., Sedwick, P. N., Peloquin, J. A., Dunbar, R. B., Hutchins, D. A., Kolber, Z., and DiTullio, G. R.: Spatial and temporal variations in variable fluoresence in the Ross Sea (Antarctica): Oceanographic correlates and bloom dynamics, *Deep-Sea Res. I*, *79*, 141–155, doi:10.1016/j.dsr.2013.05.002, 2013.

Smith Jr., W. O., Goetz, K. T., Kaufman, D. E., Queste, B. Y., Asper, V., Costa, D. P., Dinniman, M. S., Friedrichs, M. A. M., Hofmann, E. E., Heywood, K. J., Klinck, J. M., Kohut, J. T., and Lee, C. M.: Multiplatform, multidisciplinary investigations of the impacts of Modified Circumpolar Deep Water in the Ross Sea, Antarctica, *Oceanography*, *27*(2), 180–185, doi:10.5670/oceanog.2014.36, 2014.

Smith, W. O., McGillicuddy, D. J., Olson, E. B., Kosnyrev, V., Peacock, E. E., and Sosik, H. M.: Mesoscale variability in intact and ghost colonies of Phaeocystis antarctica in the Ross Sea: Distribution and abundance, *J. Mar. Syst.*, *166*, 97–107, doi:10.1016/j.jmarsys.2016.05.007, 2017.

Song, H., Edwards, C.A., Moore, A.M., and Fiechter, J.: Data assimilation in a coupled physical-biogeochemical model of the California current system using an incremental lognormal 4-dimensional variational approach: Part 3—Assimilation in a realistic context using satellite and in situ observations, *Ocean Model.*, *106*, 159–172, doi:10.1016/j.ocemod.2016.06.005, 2016.

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.

Stukel, M. R., Kahru, M., Benitez-Nelson, C. R., Decima, M., Goericke, R., Landry, M. R., and Ohman, M. D.: Using lagrangian-based process studies to test satellite algorithms of vertical carbon flux in the eastern North Pacific Ocean, *J. Geophys. Res. Ocean.*, *120*, 1–15, doi:10.1002/2015JC011264, 2015.

Thomalla, S. J., Ogunkoya, A. G., Vichi, M., and Swart, S.: Using Optical Sensors on Gliders to Estimate Phytoplankton Carbon Concentrations and Chlorophyll-to-Carbon Ratios in the Southern Ocean, *Front. Mar. Sci.*, *4*(February), 1–19, doi:10.3389/fmars.2017.00034, 2017.

Vaillancourt, R. D., Marra, J., Barber, R. T., and Smith, W. O.: Primary productivity and in situ quantum yields in the Ross Sea and Pacific Sector of the Antarctic Circumpolar Current, *Deep-Sea Res. II*, *50*(3–4), 559–578, doi:10.1016/S0967-0645(02)00584-2, 2003.

Vallverdú, J.: What are Simulations? An Epistemological Approach, *Procedia Technol.*, *13*, 6–15, doi:10.1016/j.protcy.2014.02.003, 2014.

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., and Oschlies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, *J. Mar. Syst.*, *81*(1–2), 34–43, doi:10.1016/j.jmarsys.2009.12.005, 2010.

Xiao, Y., and Friedrichs, M. A. M.: Using biogeochemical data assimilation to assess the relative skill of multiple ecosystem models in the Mid-Atlantic Bight: effects of increasing the complexity of the planktonic food web, *Biogeosciences*, *11*(11), 3015–3030, doi:10.5194/bg-11-3015-2014, 2014a.

Xiao, Y., and Friedrichs, M. A. M.: The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model, *J. Geophys. Res. Ocean.*, *119*, 2691–2712, doi:10.1002/2013JC009433, 2014b.

Yool, A., Popova, E. E., and Anderson, T. R.: Medusa-1.0: a new intermediate complexity plankton ecosystem model for the global domain, *Geosci. Model Dev.*, *4*(2), 381–417, doi:10.5194/gmd-4-381-2011, 2011.

**Table 1:** Eight parameters optimized in this analysis.

| Parameter Name | Initial value (Kaufman et al., 2017a) | Bounds (lower, upper) |
|---|---|---|
| **Diatom max growth rate at 0˚C** | 0.375 ($d^{-1}$) | 0.09375, 1.5 |
| ***P. antarctica* solitary cells C:Chl ratio** | 30 (gC $gChl^{-1}$) | 7.5, 120 |
| ***P. antarctica* colonies max growth rate at 0˚C** | 0.5 ($d^{-1}$) | 0.125, 2 |
| ***P. antarctica* solitary cells max growth rate at 0˚C** | 0.5 ($d^{-1}$) | 0.125, 2 |
| **Diatom C:Chl ratio** | 150 (gC $gChl^{-1}$) | 37.5, 600 |
| **Fast detritus sinking fraction of diatom losses** | 0.75 | 0.5, 0.95 |
| ***P. antarctica* colonies max sinking rate** | 20 (m $d^{-1}$) | 5, 80 |
| ***P. antarctica* colonies C:Chl ratio** | 40 (gC $gChl^{-1}$) | 10, 160 |

870

**Table 2:** Time, depth, and time-space resolution of glider-based observations of Chl and POC assimilated for each experiment.

| **Experiment** | **Depth (m)** | **Temporal Resolution** | **Spatial Area(s)** |
|---|---|---|---|
| **Expt 1a: Full Assimilation** | 0 - 50 | Hourly | Full glider track |
| **Expt 1b: Latitudinal Assim.** | 0 - 50 | Hourly | North, Central, South Latitudinal bands |
| **Expt 1c: Longitudinal Assim.** | 0 - 50 | Hourly | East, Central, West Longitudinal bands |
| **Expt 2a: Glider Assimilation** | 0 - 50 | ~ twice per day, separated at a minimum of 12 hours. | Full glider track |
| **Expt 2b: Cruise-based Assim.** | 0 - 50 | 3 days in a row, and then another 3 consecutive days two weeks later | Full glider track |
| **Expt 2c: Satellite-based Assim.** | 0 - 5 | 1 day every two weeks | Full glider track |

875

**Table 3:** Depth- and time-integrated primary production (PP), carbon export flux at 200 m, and costs for the No Assimilation run (cost = 0.77), Experiment #1 and #2.

| **Simulation name** | **PP (g C m$^{-2}$ y$^{-1}$)** | **Export (g C m$^{-2}$ y$^{-1}$)** | **Predictive Cost ($\mathcal{J}_P$)** | **Assim. Cost ($\mathcal{J}_A$)** |
|---|---|---|---|---|
| **No assimilation** | 111.7 | 18.8 | - | - |
| **Expt 1a: Full Assimilation** | 104.2 | 27.2 | - | 0.41 |
| **Expt 1b: Latitudinal Assim.** | 101.8[*] ±3.3 | 26.1 ±2.1 | 0.49 ±0.13 | 0.43 ±0.14 |
| **Expt 1c: Longitudinal Assim.** | 103.2 ±2.1 | 26.9 ±2.1 | 0.50 ±0.10 | 0.46 ±0.13 |
| **Expt 2a: Glider Assim.** | 103.7 ±1.8 | 27.0 ±1.2 | 0.43 ±0.01 | 0.43 ±0.03 |
| **Expt 2b: Cruise-based Assim.** | 113.1 ±22.3 | 24.8 ±6.6 | 1.24 ±0.95 | 0.52 ±0.19 |
| **Expt 2c: Satellite-based Assim.** | 114.1 ±10.7 | 16.7 ±2.7 | 1.04 ±0.36 | 0.26 ±0.16 |

[*] costs represent mean ± one standard deviation of assimilative runs.

**Table 4:** Initial parameter values (No Assimilation) and optimal parameter values after conducting the Full Assimilation, Glider, Cruise-based, and Satellite-based Assimilation experiments

| **Parameter Name** | **Initial Value** | **Expt 1a Full Assimilation** | **Expt 2a Glider*** | **Expt 2b Cruise-based*** | **Expt 2c Satellite-based*** |
|---|---|---|---|---|---|
| **Diatom max growth rate at 0˚C ($d^{-1}$)** | 0.375 | 0.40 | 0.43 ±0.01 | 0.42 ±0.15 | 0.41 ±0.09 |
| ***P. antarctica* solitary cells C:Chl ratio (gC $gChl^{-1}$)** | 30 | 29.7 | 25.84 ±5.16 | 37.3 ±26.7 | 51.5 ±26.8 |
| ***P. antarctica* colonies max growth rate at 0˚C ($d^{-1}$)** | 0.5 | 0.29 | 0.22 ±0.10 | 0.45 ±0.58 | 0.29 ±0.17 |
| ***P. antarctica* solitary cells max growth rate at 0˚C ($d^{-1}$)** | 0.5 | 0.39 | 0.45 ±0.06 | 0.75 ±0.70 | 0.79 ±0.51 |
| **Diatom C:Chl ratio (gC $gChl^{-1}$)** | 150 | 176.4 | 166.6 ±50.17 | 252.4 ±164.28 | 374.86 ±187.82 |
| **Fast detritus sinking fraction of diatom losses** | 0.75 | 0.87 | 0.86 ±0.05 | 0.86 ±0.11 | 0.62 ±0.14 |
| ***P. antarctica* colonies max sinking rate (m $d^{-1}$)** | 20 | 10.7 | 10.1 ±3.66 | 20.1 ±20.5 | 12.8 ±9.27 |
| ***P antarctica* colonies C:Chl ratio (gC $gChl^{-1}$)** | 40 | 14.0 | 14.2 ±2.29 | 42.7 ±41.6 | 34.3 ±26.5 |

* mean ± one standard deviation of assimilative runs.