Response to Referee #1

The authors present a parameter estimation and sensitivity study, applying a variational optimization technique to a 1-dimensional ecosystem model for a station off the coast of Antarctica. The manuscript is concise (sometimes too much so) and presents a sophisticated optimization framework with an interesting application. It could be improved by showing if and to what extent the results generalize to different initial parameter values.

Thank you for your positive feedback. Please note that, based on reviewer comments, the revised manuscript utilizes a new model optimization and analysis with more ecologically realistic optimized values of bacterial, microzooplankton, and krill maximum growth rates than in the previous version (1.06 d^{-1}, 1.18 d^{-1}, and 1.02 d^{-1} in the revised version vs. 5.17 d^{-1}, 6.30 d^{-1}, and 2.11 d^{-1} in the previous version which are too high for the maximum growth rates of bacteria, microzooplankton, and krill, respectively). Our response to each of your specific comments as well as figures and tables in the manuscript are now based on the results from this new model optimization. We also changed the words "optimized parameters" to "constrained parameters" and "updated parameters" to "optimized parameters" for terminological consistency with other studies which our model was originally derived and modified from. Some of the figures and tables were rearranged in the revised version as well. Finally, in an effort to address one of your comments, we paired the optimized model results with the initial (unoptimized) model results in all of the main figures and tables, and updated section 4 accordingly.

The manuscript is generally well written but in places it overestimates the project specific knowledge of the reader. Some implementation details are hinted at, but not described fully or until later in the manuscript. I have pointed out several of those instances in the specific comments, one example is the "removal" of parameters from the optimization which is a key aspect of this study but not described in much detail. According to the description, an optimized parameter with sigma_f larger than 50% is updated but then removed from the optimization procedure, but according to Table 2 most of the parameters are never updated and somehow removed before the update. It is unclear how this happens.

Our apologies for the confusion. We have attempted to clarify the text to better describe the model procedure used to address the fact that our model has many more parameters than can be appropriately optimized against the field data. Most of the parameters remained the same as their initial parameter values in Table 1 because they were never included in the subset of parameters chosen for optimization to begin with. We formed an initial parameter subset chosen for optimization consisting of 10 different model parameters (out of a total possible 72 model parameters) using the criteria now explicitly described in the text but this was never explained in the previous version (see our response to your comments below "l 359: "Among the total of 72 optimizable model parameters …"). We revised section 2.4 to clarify this point (line 249-264, line 270-279).

It is nice to see a large Monte Carlo experiment with 1000 simulations to examine the model solution in the parameter space close to the optimized parameter values. However, I would argue that a more important aspect of an uncertainty analysis was left out of the manuscript. How sensitive are the results presented here to the initial choice of the parameters? If starting in another place in parameter space, does the algorithm end up in a different local minimum? Is the same subset of 14 parameters selected consistently? Do the results of the Hessian matrix analysis differ significantly in a different local minimum? Is an 80% reduction in the model-observation misfit typical? These are questions that may be more important to a modeler who would like to perform a similar parameter estimation experiment. I would suggest that the authors perform a second (but perhaps smaller) Monte Carlo experiment and assess the sensitivity of their results to the initial parameter values. This would not need to include every aspect of their original analysis and the results for their "reference" experiment could stay and be augmented with the new results.

Thank you for this excellent suggestion. We have done a series of new optimizations using a modified set of initial parameter values to assess where our optimization is finding a local minimum in the cost function. Ideally, this sensitivity experiment to different initial parameter values would be done by optimizing the model from hundreds or thousands of randomly-sampled (Monte Carlo) initial parameter values, but this is not computationally feasible given the time-consuming nature of optimization (e.g., 1-2 days take per optimization). Instead, we formed 15 different sets of the model initial parameter values as follows. To directly compare with the reference optimization case presented in main figures and tables, in the first group of sensitivity experiments we only altered initial values for parameters within a combination of the 10-parameter subset chosen for optimization and the 12-optimized parameters for the reference case (Table 1). For these sensitivity experiments (n = 12), optimized parameters' initial values were changed by +/-50% for two (μ_{DA}, μ_{CR}), four (μ_{DA}, μ_{CR}, α_{DA}, α_{CR}), six (μ_{DA}, μ_{CR}, α_{DA}, α_{CR}, Θ, μ_{BAC},), eight (μ_{DA}, μ_{CR}, α_{DA}, α_{CR}, Θ, μ_{BAC}, g_{BAC}, μ_{MZ}), and ten parameters (μ_{DA}, μ_{CR}, α_{DA}, α_{CR}, Θ, μ_{BAC}, g_{BAC}, μ_{MZ}, μ_{KR}, g_{MZ}). A second, smaller group of sensitivity experiments (n = 3) was conducted on a subset of unoptimized parameters (A_E, *wnsv*, *remin* and g_{DA}, g'_{DA}, g_{CR}). For each sensitivity experiment, any 3 of the unoptimized parameters' initial values were changed by +/-50%, with -50% only for A_E, *wnsv*, *remin* because +50% for these three are ecologically too high for the study site.

These experiments showed that the optimized model results presented in the manuscript (i.e., the reference case) were not sensitive to the initial choice of the parameters. The 15 different initial parameter sets resulted in a range of initial model-observation misfits, some substantially larger than the reference case (14.25-28.24 vs. 14.85 for the reference case, Table B1). However, the total normalized optimized cost values of the 15 sensitivity experiments (5.79-7.19) were similar to that of the reference case of 6.42. In the sensitivity experiment #12, the initial model-observation misfit was ~2 times larger than that of the original optimization experiment, and there was up to 76% of the reduction in the model-observational misfit (vs. 58% of the reduction in the reference case). These results suggest that no matter where in parameter space the optimization started from, the adjoint/optimization scheme took the model cost function to similar local minima. Importantly, this was achieved by similar subsets of the optimized parameters. The uncertainties of the optimized parameters were also similar among different optimizations, with most of the relative

errors < 0.5. We revised section 4.1 to discuss these points (line 385-399) and included the results of the sensitivity experiments in Table B1.

When comparing the results of the optimized model with the initial one (Fig. 5 vs Fig B3), the most striking difference to me was the presence of much higher frequency variation in the optimized solution. What is causing this variation and is it realistic? If so, were the initial parameters based on particularly bad guesses, since they do not show the variation? Where possible, it would be useful to overlay these plots with the data used in the optimization procedure. Also, I think lots of readers may be interested in the before and after optimization comparison and I would suggest to add Fig. B3 into the manuscript body for comparison, and perhaps even show corresponding panels side-by-side.

The new optimization results in the revised version no longer exhibit such high-frequency variations (Figure 5B). As suggested, we included the observations of each data type used in data assimilation (Figure B2, line 415-418), paired all of the main figures and table showing the results from optimization with those before optimization (Figures 4-7 for Taylor diagrams, model state variables, ecosystem indices, and annual mean carbon flows and Table 1 for the average model bias), and updated the results accordingly (sections 4.2-4.4).

Along the same lines, why is a comparison to the initial solution avoided in so many places, for example why not include its results in the Taylor diagram in Figure 4? But maybe I am misinterpreting the initial solution a little: A reader may assume that the initial parameter values were the best values one could come up with, without going into a rigorous parameter optimization exercise such as the one presented in this manuscript. But here it appears more like the initial parameter values are based on a first (informed) guess without yet checking the effect of the parameter values on the model state. Is there any indication how "good" the initial values are, are the initial state estimates somewhat reasonable?

As is typical for many types of models, a collection of what appear to be reasonable initial parameter estimates can result in relatively poor overall system behavior because of system-level interactions of different model components. In most marine ecosystem models, these initial parameter values are subjectively adjusted by hand to improve the simulation, and the simulations with the initial, unadjusted parameter values are rarely shown. Here we have chosen a more objective optimization approach, and following your comment we now more explicitly compare the initial and optimized solution.

As suggested and explained above, we included the Taylor diagram of the initial solution in the revised version (Figure 4A) and discussed the results (line 409-415). As you pointed out, the initial parameter values were from our initial “best” guessed values of individual parameters, either derived from existing literatures or a subset of the observations; effects on the model states (system-level dynamics) were not examined prior to optimization (line 249-256). Some of these parameter values have been measured experimentally at the study site or in the similar environment (e.g., bacterial growth rate of 0.82 d^{-1}, total phytoplankton (including large cells like diatoms) growth rate of 0.33-0.55 d^{-1}, nanophytoplankton (corresponding to cryptophytes) growth rate of

0.52-0.99 d^{-1}, and microzooplankton growth rate of up to 1.0 d^{-1}; Garzio, L.M., Steinberg, D.K., Erickson, M. and Ducklow, H.W., 2013. Microzooplankton grazing along the Western Antarctic Peninsula. *Aquatic microbial ecology*, *70*(3), pp.215-232; Caron D.A., Dennett M.R., Lonsdale D.J., Moran D.M., Shalapyonok L. Microzooplankton herbivory in the Ross sea, Antarctica. Deep Sea Research Part II: Topical Studies in Oceanography. 2000; 47(15-16):3249-72), but due to the error ranges of themselves we assigned the reasonably close initial values to the corresponding parameters (line 427-434).

l 18: "Here we developed a one-dimensional, data assimilation planktonic ecosystem model [...] the pre-existing food-web and biogeochemical components of the WAP-1D-VAR model": It is not clear here if the model was developed from scratch, as the first sentence states, or that parts of the model existed before and were used here.

Thanks, the parts of the model were originally derived and modified from other data assimilation testbed models, rather than being developed from scratch. We rephrased line 20-21 to make this clear.

l 24: What are "intercompartmental flows"? Later in the abstract, "model state variables" is used and it would be good to use one expression consistently.

As suggested, we changed “intercompartmental flows” to “flows between those variables”.

l 27: "... and comparable values of the assimilated and non-assimilated model state variables and flows to other studies": This is a run-on sentence and difficult to understand, I would suggest to break it up and rephrase the second part.

Thanks, rephrased (line 27-28).

l 188: At this point it is not clear what the water depth at the model site is, how deep the model extends and how it reacts to a growing (deepening) ice cover or if this is even a concern at the site. If the authors want to keep this implementation aspect general, it would be good to make this explicit and mention under which circumstances the modeling framework can be used.

We added the depth of the modeling site (~200 m, line 307). The model depth can be extended to as deep as needed. Conceptually, the application of the 1-D model framework makes the most sense for the upper water column dominated by local seasonal processes, and extension of the model into deeper water well below the maximum seasonal mixed layer becomes more problematic because of the growing importance of lateral advective process that are not well captured in the 1-D model framework. As noted, the vertical structure of the water column can be affected by growing sea ice due to reduced wind-driven turbulence and brine rejection during winter, but this is what a prognostic, coupled ocean-ice 1-D model can offer to simulate, not our diagnostic forcing based model used in this study. Also, because our model simulates only the spring-summer growth season, winter sea-ice growth is less of a concern. We revised section 3.1. to discuss these points (line 312-320).

l 232: "using the new set of modified parameters": After reading a bit ahead, it is unclear if these modified parameters refer to modified parameter values or new parameters, as some have been removed from the optimization. In general, I would suggest to use the term "parameter value" instead of "parameter" whenever the text refers to the values rather than the parameter itself.

Those are "modified" parameters optimized from the previous optimization that serve as the basis for the next optimization, as a new "set" (line 245). As suggested, we also fixed the wording issue throughout.

l 240: The removal should be mentioned and explained in the description above.

Please see if the revised section 2.4 made this part clear.

l 262: "for growth seasons' relatively complete data coverage for modelling purposes": I am not sure what this means, please rephrase. Listing the years without any further information is not very helpful to the reader, does it imply that a new CV is estimated for every one of those time periods?

For each data type, a single, average CV was calculated using all the observations over the 2002-2012 period (line 291-294). We also revised line 294-295 to explain why those years were chosen for the CV calculation.

l 289: "optimized model simulation": Does this imply parameter estimation?

Yes, we rephrased to "by first optimizing the growth seasonal cycle…" (line 331-335).

l 291: It would be beneficial to the reader to rephrase "2002-2003, 2003-2004, 2004-2005, 2005-2006, 2006-2007, 2008-2009, 2009-2010, 2010-2011, and 2011-2012 period" to something like "9 growth periods between 2002 and 2012 (the 2007-2008 growth period is missing because of X)"

Fixed there, as suggested, as well as wherever applicable (line 293-297, line 331-335).

l 293: "averaging across all these 9 years did not reflect distinct seasonal phytoplankton peaks": What is meant by "reflect" here? Am I understanding this correctly, that phytoplankton blooms occurred at different times in the years that were examined, and so the "simple" climatology did not show any bloom? Then a more complex climatology was constructed using the time shifts to align all blooms. Was the model then optimized using parameter estimation? Were the optimized parameters used to inform boundary conditions or was the climatology used for that? More information would be useful here.

Exactly, your understanding is correct that the model was optimized for the "time-shifted" climatological year, with the same bottom boundary conditions as in the presented year (line 331-335).

l 308: "but given cryptophytes being the second dominant species in the water samples they are considered to represent non-diatom species": I assume I know what is meant here but it would be

good to clarify this point a bit more: The dataset contains data for diatoms and "non-diatoms", and here in this modeling study cryptophytes are assumed to represent all non-diatoms?

Fixed as suggested (line 350-352).

l 310: "POC(N)": I presume this is meant to mean "POC and PON" and not "POC and N" in this context. Here the "()" are used a little differently than they would normally be used and I fear it could be confusing to some readers. I would suggest to use "POC (PON)" or make the sentences a bit longer use "POC and PON" and avoid the "()".

Fixed.

l 349: "Some of these model biases cases were evidently shown on a point-to-point basis": It is not clear what this means, please rephrase.

That was to refer the point-to-point comparison figures. Revised to make it clear (line 407-408).

l 353: "The data types with relatively high correlation coefficients tended to have relatively low centred RMSD and vice versa.": In Figure 4 it looks like "BP" with a "relatively high" correlation coefficient (according to the definition from the preceding sentence) has maybe the highest centered RMSD value.

Thanks, that sentence was removed.

l 354: "the model fitted average observations slightly better": It would be nice to add different colors for those in Figure 4. Otherwise, the reader may have to go back to a few sections to review which observations were averaged.

That sentence was no longer valid after the new optimization experiments and was removed in the revised version.

l 359: "Among the total of 72 optimizable model parameters, subsets of 14 [...] parameters changed": It is still not entirely clear to me how this is done in the algorithm, are the remaining parameters dropped from the optimization before the first iteration? Also, "subsets" appears to imply multiple experiments with different results, yet Table 2 only shows one subset, i.e. 14 parameters with a value for p_f.

We revised section 2.4 to demonstrate how we chose an initial parameter subset chosen for optimization. We also fixed the word “subsets” to “a subset”.

l 369: "The optimized model results at each model time step and grid were associated with generally small errors derived from the Monte Carlo experiments (Figure B2).": This sentence is difficult to understand, is it mean to say that the ensemble of state estimates obtained from the Monte Carlo experiment, that was conducted following the optimization, has a low standard deviation? As an aside, I would suggest not to mix the terms error and uncertainty here: shown is the uncertainty in the state estimate not an error that needs to be corrected.

We rephrased the mentioned part (line 435-437) and changed the word “error” to “uncertainty” wherever applicable.

Figure 1: In the figure description: "oragnic"

Fixed.

Response to Referee #2

This work develops a one-dimensional data assimilation model for the West Antarctic Peninsula and compares the model output with data from the Palmer Long-Term Ecological Research site. Overall, this paper provides a useful model to aid in data assimilation techniques and modeling efforts for the WAP.

Thank you for your positive feedback. Please note that, based on reviewer comments, the revised manuscript utilizes a new model optimization and analysis with more ecologically realistic optimized values of bacterial, microzooplankton, and krill maximum growth rates than in the previous version (1.06 d^{-1}, 1.18 d^{-1}, and 1.02 d^{-1} in the revised version vs. 5.17 d^{-1}, 6.30 d^{-1}, and 2.11 d^{-1} in the previous version which are too high for the maximum growth rates of bacteria, microzooplankton, and krill, respectively). Our response to each of your specific comments as well as figures and tables in the manuscript are now based on the results from this new model optimization. We also changed the words "optimized parameters" to "constrained parameters" and "updated parameters" to "optimized parameters" for terminological consistency with other studies which our model was originally derived and modified from. Some of the figures and tables were rearranged in the revised version as well. Finally, in an effort to address one of the reviewer #1 comments, we paired the optimized model results with the initial (unoptimized) model results in all of the main figures and tables, and updated section 4 accordingly.

One general comment is that it would be beneficial to describe how well the optimized and updated parameters compare to measured parameter values from the field, lab, or both. That is, are the optimized and update parameters realistic or is there no baseline?

Thanks for this great suggestion. Though not many, there are several field and lab-based studies at the study site or in a similar polar environment that reported the ranges of the values of the model parameters used in our model. These include the bacterial growth rate of 0.82 d^{-1}, total phytoplankton (including large cells like diatoms) growth rate of 0.33-0.55 d^{-1}, nanophytoplankton (corresponding to cryptophytes) growth rate of 0.52-0.99 d^{-1} calculated from dilution experiments (Garzio, L.M., Steinberg, D.K., Erickson, M. and Ducklow, H.W., 2013. Microzooplankton grazing along the Western Antarctic Peninsula. *Aquatic microbial ecology*, *70*(3), pp.215-232) and the microzooplankton growth rate of up to 1.0 d^{-1} (Caron D.A., Dennett M.R., Lonsdale D.J., Moran D.M., Shalapyonok L. Microzooplankton herbivory in the Ross sea, Antarctica. Deep Sea Research Part II: Topical Studies in Oceanography. 2000; 47(15-16):3249-72) serving as the useful basis for evaluating the optimized parameter values in our study. The optimized values of the maximum bacterial, diatom, cryptophyte, and microzooplankton growth rates are 1.06 d^{-1} (0.93-1.20 d^{-1}), 0.77 d^{-1} (0.68-0.88 d^{-1}), 0.72 d^{-1} (0.61-0.85 d^{-1}), and 1.18 d^{-1} (1.10-1.26 d^{-1}), which are comparable to the measured values above. We revised section 4.2 to include these points (line 427-434).

Lines 72-73--Can you please provide support for the appropriateness of including just two phytoplankton groups (diatoms and cryptophytes) in this region?

Typically, the coastal WAP is associated with large phytoplankton blooms dominated by large (> 20 μm) diatoms, but nanoplankton, or nanoflagellates (< 20 μm) are also an important component of the food web, dominated by cryptophtyes (Schofield, O., Saba, G., Coleman, K., Carvalho, F., Couto, N., Ducklow, H., Finkel, Z., Irwin, A., Kahl, A., Miles, T. and Montes-Hugo, M., 2017. Decadal variability in coastal phytoplankton community composition in a changing West Antarctic Peninsula. *Deep Sea Research Part I: Oceanographic Research Papers*, *124*, pp.42-54). Mixed flagellate, prasinophytes, and type-4 haptophytes are also found in the region, but we chose to model two dominant phytoplankton compartments in order to avoid too many free (optimizable) parameters associated with each phytoplankton group. We revised section 2.1 to include these points (line 74-80).

Lines 130-135—The authors mention microzooplankton had a limit on the amount of diatoms they grazed and instead grazed cryptophytes to be able to simulate elevated diatom Chl. Can you comment on the ecological appropriateness of these grazing dynamics? That is, is there evidence to support that this prey switching occurs? What other mechanisms might lead to elevated diatom chlorophyll beyond microzooplankton changing their food preferences?

In principle, optimization should be able to capture the elevated diatom chlorophyll by adjusting free parameters unless: 1) the right parameters are not adjusted and/or the baseline (non-optimized) parameters need adjusting, and 2) the model equations are not adequate even with the optimized parameters. We did our best assigning ecologically-appropriate initial parameter values for parameters associated with the time rate of change of diatom biomass as well as performed sensitivity experiments decoupling the diatom-zooplankton relationship. However, the model did not capture the elevated diatom chlorophyll, so we went with the second scenario by altering the equation to limit grazing on diatoms by microzooplankton, as one option; for example, we could have added a quadratic term to the mortality or loss term for zooplankton to increase diatom chlorophyll as well. It is challenging to provide the exact evidence of prey switching, or of zooplankton mortality thresholds, due to the lack of observations/experiments at the study site. To a certain extent, this is to show that data assimilation enables to "correct" the existing model equations to better match model output to observations, a valuable lesson that this data assimilation model can offer to modeling and experimental communities. We revised section 2.2 to include these points (line 142-149).

Lines 159-160 state that microzooplankton growth is based on grazing on cryptophytes and bacteria, while krill growth is based on grazing on diatoms and microzooplankton. However, in lines 130-135, the authors mention there is a limited amount of microzooplankton grazing on diatoms. Please reconcile this information. Does it mean the limit grazing mentioned in lines 130-135 is no grazing on diatoms by microzooplankton? That would contradict lines 74-75. Line 317 also mentions microzooplankton only grazing on bacteria and cryptophytes.

Diatoms are still grazed by microzooplankton, but with a grazing limitation term assigned to the equation. Please see our response above to your comment.

Relatedly, in Fig. 1, there are two grazing arrows going from diatoms to microzooplankton and no grazing arrows for cryptophytes. Please reconcile this figure with information in the text.

We fixed Figure 1 – somehow the figure in the previous version had the shifted arrow from cryptophytes to diatoms for this.

Line 224 and Table 2 – Consider indicating the literature sources from which the model parameters were taken.

Thanks, added the reference.

Fig. B1 – can you please clarify what you mean by "Errors represent how much larger model output is compared to observations"?

We rephrased that sentence.

Fig. 5—Can you please clarify what is leading to the oscillating patterns seen in the model state variables such as diatoms and cryptophytes?

The revised Figure 5 from the new optimization experiments which no longer show those oscillating patterns.

Line 153 – should the sentence read "…by remineralizing NH4 and PO4 if C is in short" rather than "if C in short"?

Fixed.

Line 428—should this read "There could be several additional…" instead of "There would be several additional…"?

Fixed.

Lines 467-468, 473—the citation formatting changed. Please fix it to be consistent

Fixed.

Fig. 3, 5, 6 – consider putting the year on the x-axis

Fixed.

Fig. 5 – the caption should be updated with the correct figure (presumably not "Figure SX")

Fixed.

Fig. 7 – The arrows in this figure are presumably for the same processes in Fig. 1. Consider labeling them and/or referencing the reader to the arrows in Fig. 1, assuming that is appropriate.

Thanks for the suggestion. We referenced the arrow number in Figure 1 when discussing the results from Figure 7 (line 532-534).