

General comments

It is great to see that the authors repeated the parameter estimation experiment with different initial values. However, the results could be presented better and should probably be given more space. First of, the repeat experiments could already be mentioned and motivated in Section 2.4. Currently they are only presented in a relatively short passage in Section 4.1. Here, some statements create more questions than answers: "this was achieved by similar subsets of the optimized parameters" (line 397) Which subset of parameters remained the same, which only in a few experiments?

Thanks, we included the description of the sensitivity experiments in section 2.4 (line 282-285) and listed the optimized parameters in each experiment (line 404-405).

Apparently, the different experiments had similar results according to this passage "[...] and there was up to 76% of the reduction in the model-observational misfit (vs. 58% of the reduction in the reference case; Table B1) 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. (line 395)" Yet, a similar reduction in the cost function value does not imply similar parameter values.

The text has been modified to state more explicitly the sensitivity of the optimized parameter values to the perturbations in initial conditions. The averaged optimized parameter values from the fifteen sensitivity experiments are comparable to those in the original experiments (line 407-408).

In the current version, Table 1 paints a nice picture with uncertainty intervals for some parameters that are locally derived from the Hessian matrix. Having multiple cost function minima with different uncertainty intervals (and different parameters that are optimized/constrained) may distort that picture a bit and the presented uncertainty intervals may just be representative for 1 out of 16 experiments. Because this information is not available to the reader, it is unclear how generalizable the results in Table 1 are with respect to the other 15 experiments. A few more details/results to clarify, and maybe some discussion would be very helpful here.

Thanks for the suggestion. We included a summary table for the fifteen perturbation experiments (Table B2).

In the updated manuscript, it is now much clearer how parameters are "removed" from the optimization. I am a bit surprised that removed parameters are abruptly reset to their initial values. Given the strong correlations that may be present between the parameters, a sudden reset of parameter values after several iterations could lead to "shocks" in other parameters. Or is the estimation restarted with all parameters starting from their initial values and fewer included in the optimization?

We reset those parameters to their initial values because of their unrealistically optimized values from a previous optimization cycle (e.g., higher or lower than their initial parameter values by several orders of magnitude). These parameters tend to co-occur with a few other, cross-correlated

parameters that are also optimized unrealistically. Those pairs end up being removed and reset to their initial values simultaneously, so this does not cause shocks in other reasonably optimized parameters.

Section 4.3 is currently divided into two paragraphs. One is listing the changes brought to ecosystem indices by the optimization, without indicating if these changes are an improvement. The second paragraph then compares the model output to data and results from other studies and only briefly touches on the changes brought by assimilation. Here, it would be really useful to mix the two paragraphs and report the changes with reference to the data that is available.

Rearranged as suggested (line 472-474, 492-507, 510). There is no data available for correlation coefficients between the ecosystem indices so it is difficult to discuss if the adjusted correlations from optimization represent an improvement or not.

Specific comments

l 29 "we discuss fully potential underlying reasons": Sounds a bit convoluted, maybe delete the "fully", also because it is difficult to claim an exhaustive discussion.

Fixed

l 53: "its strength": I would suggest that there are multiple, changing it to "its strengths".

Fixed

l 75: I think the first "dominated" in this sentence refers to large phytoplankton only, and the second one to smaller ones but I still think that two "dominated" are a bit confusing here.

Fixed

l 142: "In principle, optimization should be able to capture the elevated diatom Chl by adjusting free parameters unless: 1) the right parameters are not adjusted and/or the baseline (non-optimized) parameters need significant adjusting, and/or 2) the model equations are not adequate even with the optimized parameters." What if the nutrient initial values are too low, would errors in the state estimates be a third option?

Initial nutrient values are not low ($\text{NO}_3 = 22.8 \pm 1.5 \text{ mmol m}^{-3}$, $\text{PO}_4 = 1.64 \pm 0.11 \text{ mmol m}^{-3}$, mean \pm standard deviation from depth-averaging).

l 145: I am not sure if there has been much evidence for it in the WAP region but could their thick shells be a reason for less preferential grazing on diatoms?

Interesting point, but we are not aware of such evidence.

l 250: "or estimated using a subset of the observations, without examining the effects of the initial parameter values on the model results prior to optimization": It's not clear how this should work.

A subset of observations is used but the effect of the parameter values is never examined? What is the subset of observations used for then?

Apologies for the confusion, we followed the literature values listed instead of using a subset of the observations.

l 258: "with one parameter per each state variable, the change of which yields the largest decrease in the total cost function": How was this determined? Was the one parameter per state variable put in place first or did it turn out that the parameters yielding the largest effect, were one for each state variable?

It was the latter case. The parameters yielding the largest change in cost function were the ones we selected for the initial parameter subset, which also happened to be usually one per each state variable.

l 271: "If parameters are optimized to ecologically unrealistic values, they are kept back to the initial parameter value": Even if they have undergone some changes in the previous steps, they are reset to the initial values? If so, could this have an effect on the other parameters which may be correlated?

We addressed our response to this above under general comments.

Eq 6: It would be good to clearly state the difference between the mean that is used in the computation of CV and the climatological mean that multiplied with it.

Fixed.

l 300: "J equivalent to J/M hereafter": Why not introduce it immediately?

We did not change J to J/M in order to preserve the original cost function equation, the way typically written in other literatures. We instead chose to calculate and discuss J/M to just make it easier to discuss against Chi-square value ranges.

l 317: What about increased wind-driven turbulence as the ice disappears, is this a concern?

It could be, but at best our model may simulate wind impacts indirectly, rather than directly as a prognostic model would do, given that the diagnostic MLD field in the model is derived from CTD observations that may reflect the impact of wind forcing on the vertical structure of the water column.

l 319: "Also, because our model simulates only the spring-summer growth season, winter sea-ice growth is less of a concern.": Use a different term for sea-ice growth or change first instance of growth to something like "phytoplankton growth", so that phytoplankton growth won't be confused with sea-ice growth in this sentence.

We changed "sea-ice growth" to "the impact of winter sea ice on ecosystem dynamics".

l 331: I know that I had a question about this in my last review and I still think it should be explained better or just made more explicit. "Initial conditions are prepared by first optimizing the full growth seasonal cycle forced by climatological physics and assimilated with climatological observations and with the same bottom boundary conditions used in the optimization of the 2002-2003 growth season" I think it should be pointed out here what kind of optimization is performed. Talking about the initial conditions, one could assume that optimization implies state estimation here, i.e. adjusting the initial conditions directly. However, based on the comments to my question, it appears that parameters were estimated for a climatological simulation which was then used to create the initial conditions. But where do the initial conditions for that climatological simulation come from? I think my problem is that I don't still understand what exactly "first optimizing the full growth seasonal cycle" really means.

Thanks for this good point that we had missed to clarify. Initial conditions for the very first climatological model simulation were prepared by adjusting initial conditions manually following literature values (e.g., Luo et al. 2010), while those for the following rest of the simulations were prepared by parameter optimization (line 339-341).

l 390: "presented in the manuscript": Change to "presented above".

We deleted that phrase in case "above" may be misleading upon rearranging of the paragraphs after publication.

l 471: Is the decreased or increased (for NCP and POC) correlation realistic?

We do not have literature values to discuss if the adjustment is realistic, rather than knowing the higher correlation indicates stronger coupling and vice versa.

Fig. 4: Use the same coordinate system in both plots. Preferably, combine both plots into one, with different symbols for prior and posterior solution and different colors for the different observation types.

Please note that we had tried the way you suggested but as a result a significant number of data points overlapped with each other and compromised legibility.

Fig. 6A/B: Join into the same figure, just like Fig 5B.

Thanks for your suggestion, but we intend to show A for "initial/unoptimized results" and B for "optimized results" for consistency.