Response to Peter Strutton (Referee #2)

We thank the reviewer for the constructive comments and suggestions which will be very helpful as we revise the manuscript. Below the complete reviewer comments are shown along with detailed responses to each comment (reviewer comments in black, responses in blue font)

Review:

This is a very useful contribution that explains the benefit that models can derive from the incorporation of satellite and BGC-Argo observations. The paper is timely and clearly written. I recommend publication after minor revisions.

Response: We would like to thank the reviewer for the positive assessment and constructive comments.

Specific comments:

The introduction is comprehensive. It could be shortened a bit (the 3rd and 5th paragraphs could mostly be removed) but this is not essential.

Response: Thanks for these suggestions. The 5th paragraph is to show weakness of satellite observations which served as the motivation of this study and is highly related to some main conclusions. We will consider shortening this paragraph, but we would like to keep it in our revised manuscript. As suggested, the 3rd paragraph will be removed.

Methods switch between present and past tense. Also not a big deal, just disconcerting for the reader.

Response: We will look at this carefully to make sure the tense is used appropriately.

P6 L48: Here and in subsequent equations/text I'm a bit confused. The float and satellite measure bbp700 and bbp670 respectively. So why are we now talking about bbp470? And what is meant by 'validated bbp470'?

Response: We talked about bbp470 because the empirical relationship that we used to estimate phytoplankton and POC was based on bbp470 (please see equ. 2-3). We had to convert measurements of bbp700 and bbp670 to bbp470 based on the equ. 1 before we estimated phytoplankton and POC.

We will clarify this and make it clearer in our revised manuscript.

P6 Eq 2 and 3: What are the units of the terms on the LHS? Please be more specific about what 'Phytoplankton' is. I think it's phytoplankton N.

Response: Yes, the phytoplankton and POC were in unit of mmol N m^{-3}. We actually have mentioned it in P6-7 L58-60 in our original manuscript. We will also revise and make it clearer here as suggested.

P11 L69: Here and in section 3.1, the temporal resolution of the satellite data is not specified. I also think a bit more information here would be useful. How are monthly climatologies of the float profiles created? What distance from the 1D site is considered? Maybe this is described elsewhere and I missed it, but I see the other reviewer asked something similar.

Response: In this study, we used monthly satellite estimates of surface chlorophyll in the parameter optimization. The monthly climatology of float profiles was created by averaging all profiles collected in the Gulf of Mexico into monthly bins. We did this because 1) the BGC-Argo float profiles were sparsely distributed in the Gulf of Mexico and there were insufficient profiles around the 1D site (please see figure 1 in our original manuscript), and 2) the deep ocean part of the Gulf of Mexico is quite homogenous horizontally. As suggested, we will include some description and explanation in our revised manuscript.

P14-15: In the sub-section headings, it wouldn't hurt to remind us what experiments A, B and C are. That is 'satellite only' etc.

Response: Agree. We will revise it as suggested

Figures 3 and 8: Why not just put the parameter labels on the y axes?

Response: Yes, will do as suggested.

For the 3D case, I think it's correct to say that Figure 8 is an average of all model grid cells where the water depth is >1000m. That's a pretty big area. Yes, it's reasonably uniformly low chlorophyll, so one could make the case that this encompasses a contiguous bio-region. But why not choose a smaller box in the middle of the deep part of the basin, and perhaps one from the shelf, to illustrate the model performance? I suspect the latter will not perform as well, but that would still be interesting to know.

Response: To the first point, since the BGC-Argo float profiles are sparsely distributed in the deep part of the Gulf of Mexico (please see Figure 1 in our original manuscript), choosing a smaller box would mean that much fewer float profiles are available for the model-data comparison. Also, since the region is quite homogenous horizontally, we feel it is appropriate to average over all. In Figure 4, we show the interquartile range of the profiles in space and time (black bars). These are very similar if calculated only for July. Thus, the error bars in Figure 4 give a good indication of the spatial variability. We will add this in the revised manuscript.

To the second point, the BGC-Argo floats were deployed only in the deep part of the Gulf, hence no float profiles are available on the shelf. However, we agree that it would be interesting to show the validation in smaller boxes from the deep ocean and shelf, and will include this comparison with satellite surface chlorophyll in the supplementary of our revised manuscript.

The results proceed in a logical fashion through the different experiments, 1D and 3D. The presentation of the results is clear and concise.

Response: Thank you, we really appreciate this comment.

In a paper like this, readers will likely be looking for a clear recommendation: Which of the three options should they choose? I think the conclusions do a good job of summarizing the recommendations and the figures represent what's lost if it's not possible to implement the best case scenario.

Response: Thank you, we are happy to hear this.