Response to Referee #2

First, I apologize for being so late with the review, due to exceptional circumstances.

The paper aims at assessing the impact of 4 remotely-sensed datasets on different versions of a global model, and in particular examine the consistency in between the datasets. The impact of assimilating one or multiple datasets into the models is examined in terms of different model variables. The datasets covers physics and biology, and the impact of physics on biology is also examined. The inverse is not examined, as there is no feedback from biology to physics. Some interesting -and sometimes maybe counter-intuitive- conclusions are obtained, e.g. that the assimilation of certain datasets degrades certain model variables.

The methods and the results are clearly described. Given that the models (physics and biogeochemical) and data assimilation methods are already extensively described and validated in previous papers, they form a very sound basis for the present study. There are some limitations or deficiencies in the modeling system, but they are known and acknowledged. Thus, it does not require to be validated again in the present paper. Some aspects of the data assimilation procedure feel a little like a cooking recipe, but I guess all modelling systems have these kind of safety nets in their implementation of the data assimilation (e.g. skip assimilation some days when it renders the model unstable, or at least play with the error covariances...).

In general, the paper is clearly structured, very well written, and I did not find typos. The paper is very interesting and timely, because it provides the kind of information needed for near-future versions of biogeochemical model simulations, both reanalysis and operational. As the author mentions, the impact of data assimilation of physics on biogeochemistry is still insufficiently studied, in particular methods for mitigating the impact of spurious vertical currents, but this point is not the topic of the present paper. This study about the remaining problems linked to co-assimilation of different variables is very welcome.

Thank you for your review of the manuscript, and your positive comments.

Minor remarks:
*** although the author refers extensively to the relevant papers for the physical model, biogeochemical model, and data assimilation method, it could help to very briefly provide a few key facts. For example, for the physical model, he could mention the Nemo version (around line 62 page 3). For the BGC model, he could mention how many state variables there are, etc.**

The NEMO version is 3.4, and HadOCC has six state variables. I propose expanding the description in Section 2 to include this information, as well as additional details regarding the one-way coupling between physics and biogeochemistry, the use of first guess at appropriate time (FGAT) by the data assimilation, and the use of conservation of mass and estimates of phytoplankton growth and loss errors by the nitrogen balancing scheme.

*** the results seem valid for both the 1 degree and 1/4 degree models (independently from the fact that the higher resolution seems to present better results, even though the double penalty). But little detail is given regarding the resolution, and potentially, generalisations to even higher resolutions such as used in other regional models.**

I propose adding an extra sub-section of the Results section examining variability in the Tropical Pacific in response to ENSO in runs at each resolution. Please see my response to

Referee #1 for details. I also propose to expand the discussion of resolution in the Summary and Conclusions section. In the original manuscript I simply stated:

"These conclusions apply to both the 1° and 1/4° configurations of the model, though the higher resolution model was better able to simulate surface fCO_2, with and without data assimilation."

I propose expanding this to:

"Conclusions about model and assimilation performance, and consistency and variability, apply similarly to both the 1° and 1/4° configurations of the model. The higher resolution model was better able to simulate surface fCO_2, with and without data assimilation. This may be due to improved representation of processes in the 1/4° configuration, or may reflect differences in initialisation of DIC and alkalinity fields, which model fCO_2 has been shown to be sensitive to (Lebehot et al., 2019). The two resolutions show comparable temporal variability, with data assimilation having a similar impact. It is likely that conclusions about multivariate consistency are broadly generalisable to other resolutions and potentially regional models, though as all models and configurations have their own particular properties and biases, exact results may vary."

Lebehot, A. D., Halloran, P. R., Watson, A. J., McNeall, D. J., Ford, D. A., Landschützer, P., et al. (2019). Reconciling observation and model trends in North Atlantic surface CO_2. Global Biogeochemical Cycles, 33, 1204–1222. https://doi.org/10.1029/2019GB006186.

*** can the author give precisions how his conclusions may ultimately lead to refinements or improvements in the models (as he says in the abstract that this is a potential benefit of the study) ? Is he thinking of methods like parameter estimations, or is he simply pointing to the known limitations of the current modelling system (such as biases mentionned at line 179, 189, 206).**

I was thinking of both these types of things, and agree this could be expanded on in the manuscript. The Summary section contains the following text:

"But it also revealed complex interactions between the model and assimilation, with the assimilation of individual variables improving some non-assimilated variables while degrading others, and correcting some compensating errors while introducing others. This could potentially be counteracted by assimilating additional variables such as nutrients (Yu et al., 2018), but these cannot be observed from space and so observational coverage remains insufficient. The experiments performed in this study highlight ways in which the model and assimilation could each be developed accordingly, but there does not appear to be any one simple fix."

I propose to expand this discussion to read:

"But it also revealed complex interactions between the model and assimilation, with the assimilation of individual variables improving some non-assimilated variables while degrading others, and correcting some compensating errors while introducing others. The experiments performed in this study highlight ways in which the model and assimilation could each be developed accordingly, but there does not appear to be any one simple fix. Addressing known model biases in variables such as SST and chlorophyll should reduce such issues. Impacts on non-assimilated variables could also be counteracted by assimilating additional observations such as nutrients (Yu et al., 2018), but these cannot be observed from space and so observational coverage remains insufficient. The biogeochemical model could potentially benefit from spatially-varying parameterisations, with parameter values updated in combined state-parameter estimation (e.g. Simon et al., 2015)."

Simon, E., Samuelsen, A., Bertino, L., & Mouysset, S. (2015). Experiences in multiyear combined state–parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the Ensemble Kalman Filter. Journal of Marine Systems, 152, 1-17.

*** maybe the text about the safety-nets of the assimilation procedure can be re-written to make it feel more rigourous (as it certainly is)**

I agree that this could have been better written. The original text reads:

"In most cases, assimilation increments were applied at all model grid points. However, for model stability a few exceptions were required. No increments were applied in the Baltic Sea in the 1° runs, which is treated as an enclosed sea at this resolution. No assimilation was performed on 18 January 2000 in 1° runs including SLA assimilation, as a few large SLA observations were causing the model to fail. On a few occasions the assimilation caused LOW_SLA and LOW_OC_SST_SIC_SLA_T&S to fail near the Antarctic coast; in these cases no increments were applied for a short period in the surrounding region. Similarly, no increments were applied in the Malvinas Current region on a small number of dates in HIGH_SLA and HIGH_OC_SST_SIC_SLA_T&S. On all dates, no biogeochemical increments were applied in grid boxes with SIC greater than 0.01, which is a relaxation of the conditions imposed by Ford et al. (2012) and Ford and Barciela (2017). Furthermore, phytoplankton nitrogen increments were limited in magnitude to 1.0 mmol m^{-3} in a region surrounding the Amazon river outflow, prior to running the Hemmings et al. (2008) nitrogen balancing scheme, in order to avoid spuriously large DIC increments at very low chlorophyll concentrations. These cases were generally indicative of issues with the model and assimilation procedure under specific circumstances, rather than of errors in the observation products."

I propose rewriting this as follows:

"In most cases, assimilation increments were applied at all model grid points. However, for model stability the following exceptions were required:

- *No increments were applied in the Baltic Sea in the 1° runs, as it is treated as an enclosed sea at this resolution.*
- *No biogeochemical increments were applied in grid boxes with SIC greater than 0.01, in a relaxation of the conditions imposed by Ford et al. (2012) and Ford and Barciela (2017).*
- *Phytoplankton nitrogen increments were limited in magnitude to 1.0 mmol m^{-3} in a region surrounding the Amazon river outflow, prior to running the Hemmings et al. (2008) nitrogen balancing scheme. This was to avoid spuriously large DIC increments at very low chlorophyll concentrations in the region of freshwater influence.*
- *No assimilation was performed on 18 January 2000 in 1° runs including SLA assimilation, as a few anomalously large SLA observations resulted in unrealistic increments.*
- *Near the Antarctic coast during February and March, sparse SLA and T&S observations located in melt ponds occasionally led to unrealistically large increments being generated in LOW_SLA and LOW_OC_SST_SIC_SLA_T&S. In these cases, no increments were applied for a short period in the surrounding region until the ice had melted further.*
- *SLA assimilation is designed to be performed in combination with T&S assimilation (Lea et al., 2014), and assimilating SLA data on its own can sometimes result in adverse changes to subsurface density structure in energetic regions. This occasionally led to a model instability in the Malvinas Current region in HIGH_SLA,*

and so to prevent this no increments were applied in this region on 12 dates during the run. This was also required on one date during HIGH_OC_SST_SIC_SLA_T&S."