The manuscript describes a study assimilating temperature and salinity observations into a global physics-biogeochemistry ocean model, with the aim of improving the modelled air-sea CO_2 flux. The assimilation brought the model temperature and salinity closer to the assimilated observations, and had a mixed impact on the carbon variables and wider biogeochemistry. The global mean change was small, but could be regionally significant, with the mechanisms explored.

The experiments are well conceived, and the manuscript generally well written and well presented. I just have some comments where aspects could use clarifying or expanding on.

L51: “Data assimilation (DA) has been employed …” This paragraph doesn’t need to be comprehensive, but could be modified and expanded a little to more fully represent the available literature. Valsala and Maksyutov (2010, https://doi.org/10.1111/j.1600-0889.2010.00495.x) ran a global assimilation for 1996-2004; not multidecadal but almost as long as the present study. The paragraph states “In each of these studies, an Adjoint or Green’s Function DA approach is used”, but the Gerber et al. (2009) study referenced used an EnKF – another non-adjoint/Green’s function example is While et al. (2012, https://doi.org/10.1029/2010JC006815) who used a sequential analysis correction scheme to assimilate pCO2. The paragraph opens by talking about “DA studies of the air-sea CO2 flux” in general terms, only semi-clarifying later that it’s focussing on studies which directly assimilated pCO2 data. There have also been other studies which, like the present one, looked at the impact of assimilating other variables on the air-sea CO2 flux, e.g. Ciavatta et al. (2016; https://doi.org/10.1002/2015JC011496) and other papers from that group, and Ford and Barciela (2017, https://doi.org/10.1016/j.rse.2017.03.040).

Thank you for the references to the literature, great! We acknowledge these.

L65: “It was shown that assimilating ocean physics at the initial state of a model simulation has a stronger and more positive impact on the modeled carbon cycle than assimilating the BGC initial state (Fransner et al., 2020).” In no way diminishing the motivation for this current study – which is undoubtedly important for the reasons stated in Fransner et al. (2020) and others – it could be clarified that this was a single model study and may or may not hold in general. The relative importance of physics vs biogeochemistry initialisation on different variables and time scales remains an open question – see e.g. the discussion in Section 4.4 of Lebehot et al. (2019, https://doi.org/10.1029/2019GB006186) and indeed the ultimate conclusions of this current manuscript.

Yes, this definitely needs to be put into context. Thank you again for the literature references.

L67: “Therefore the question arises which processes are most important when altered physics change CO2 fluxes in DA approaches.” I think I understand the meaning of this sentence, but it could be reworded for clarity.

“This raises the question which mechanisms are responsible for the response of the \ce{CO2} flux in physics DA approaches.”

L68: “to improve” – a better wording could be “to aim to improve”?

Yes, we changed this.

L75-79: The issues discussed by Park et al. (2018) and others, mentioned later in the manuscript, could be introduced at this point.

Okay, we moved this here.

L103: “Alkalinity is restored by a fictional surface flux of 10m/yr.” Is there a reference for this, or was it introduced in this study?

We follow the set-up of Gurses et al. (2023). This alkalinity restoring has been used by Hauck et al. (2013) and Schourup-Kristensen (2014) as well.

Gurses: doi.org/10.5194/gmd-16-4883-2023
Hauck: doi.org/10.1002/2013GB004600
Schourup-Kristensen: doi.org/10.5194/gmd-7-2769-2014

L121: “After each assimilation step, corrections are applied to the analysis state to ensure the consistency of model physics.” Can you give an indication of whether these corrections need to be applied regularly or just occasionally?

While the correction is necessary at each step for about 10\% of SSH updates and \SI{1e-3}{}\% of temperature values, the correction of salinity is never needed.

L148: How is the weekly-resolution SSS used in the daily assimilation?

To clarify, we have rephrased: “ESA-CCI contains daily data at a spatial resolution of 50~km, albeit not capturing temporal variability below weekly.”

ESA-CCI: doi.org/10.1029/2021JC017676

We use the daily ESA-CCI data for the daily assimilation steps. It is not necessary for the observations to contain the day-to-day variability, as the data assimilation has a comparatively slow effect: For example, it takes several months of assimilation to achieve the maximum feasible correction of a large-scale model bias.

L153: “model values are computed as the average of the grid points of the triangle enclosing” – what’s done in the vertical?

“To assimilate the profiles, the observations are assigned to the respective model layers (depth range) in the vertical.” - added to the manuscript.

L171: “For the comparison …” – this paragraph would benefit from a clearer explanation of what adjustments have been made to what products and why, including the model estimates from this study (which presumably have no river carbon inputs?).

In response to both reviewers pointing this out, we have rephrased:

“We present \ce{CO2} flux estimates for the period 2010-2020, that are compared to the 'Regional Carbon Cycle Assessment and Processes 2' (RECCAP2) global air-sea \ce{CO2} flux estimates \citep{devries2023}. To make the RECCAP2 estimates comparable with our estimate stemming from a model without river carbon input, we apply a river flux adjustment \citep{friedlingstein2023,regnier2022} to the RECCAP \ce{pCO2} products. Thus, we quantify the anthropogenic perturbation of the ocean carbon sink \citep[as $\mathrm{S_{OCEAN}}$ in the Global Carbon Budget][]{friedlingstein2023,hauck2020}, and not the contemporary net air-sea \ce{CO2} flux with outgassing of river carbon into the atmosphere (as in RECCAP2).”

L206: “we define the improvement as” – I’m in two minds whether calling the statistic “improvement” is good as it’s clear and intuitive, or if it should be more objective and phrased as “reduction in mean absolute difference” or something equally dry. On balance I’m happy how it is, given it’s clearly defined, but will keep this comment here for completeness. It can be a little odd when positive and negative improvement gets discussed (e.g. L254, L258).

The term ‘improvement’ was used before (see e.g. Losa et al., 2012: https://doi.org/10.1016/j.jmarsys.2012.07.008, with positive and negative improvements in Figure 1 and 2).

L220: “EN4-OA” – this is a reasonable product to use for comparison, but my understanding is that it includes no observations beyond the assimilated data, just interpolation between data points. So calling it “partly-independent” or “non-assimilated” (L244) may be misleading. Furthermore, it could have been introduced in the previous section.

Yes, we refrain from this wording because EN4-OA and EN4 are not independent. We have also moved the introduction of EN4-OA to the previous section.

L228: “in particularly” – in particular

Thank you, done.

L240: “particularly much” – “particularly”

Thank you, done.

L241: “Albeit negative side effects of temperature assimilation” – how is it judged that the temperature assimilation is responsible?

“Tests with the assimilation of temperature alone show negative side-effects of temperature assimilation on SSS in some locations. In the final set-up with combined assimilation, negative effects on SSS are found in 9\% of the observed area.” - added to the manuscript.

Fig. 1 and others: My instinct would be to plot ASML – OBS rather than ASML – FREE. However, I’ve argued about this with coauthors on papers before, and appreciate others strongly feel ASML – OBS is the better choice. So I'm merely flagging it as something to consider, I can see the argument both ways.

We have chosen ASML - FREE throughout the manuscript because it allows us to visualize comparatively small changes in some of the biogeochemical variables. On the one hand, for temperature and salinity, ASML-OBS provides a clear picture of the model error after data assimilation. On the other hand, for the biogeochemical variables, FREE-OBS and ASML-OBS are visually too similar to recognize the differences. Showing ASML-FREE for all variables allows us to recognize similarities between the effects of DA on different variables.

L275: "see Appendix Text A1 for further discussion". Appendix Text A1 is a single short paragraph, I don't understand why it's in an appendix. It would be better in the main manuscript, either here or in the Discussion section.

We have moved it here.

L276: "Thus, it can be assumed that the velocities in the upper part of the ocean are also well represented." I don't think you can make this assumption, certainly not for vertical velocities. See e.g. Raghukumar et al. (2015, https://doi.org/10.1016/j.pocean.2015.01.004) and Gasparin et al. (2021, https://doi.org/10.1016/j.ocemod.2021.101768). The data assimilation will continually update the observed variables to better match the observations, without necessarily leading to improvements in non-observed variables such as velocities – although of course that's the aim. The current study certainly doesn't seem to have the issues with vertical velocities the above studies do, but without providing assessment of the wider circulation there's no guarantee it's improved.

We agree that there is no guarantee that it's improved and have therefore rephrased: "This can be interpreted as an indication that the velocities in the upper part of the ocean are also well represented."

The advantage of referring to T and S observations is that these are directly comparable, which is not the case for velocities (that are partly parametrized in FESOM). Furthermore, the modeled boundary layer cannot be directly compared to a classically defined mixed layer either.

L280: "4 Results" – Section 3, "Effect of DA on ocean physics" is also results. Perhaps Section 4 should be "Effect of DA on ocean biogeochemistry".

Yes, we have adjusted the section titles.

L282: "The ocean absorbs 2.78 Pg C dec^{-1}" – is this the correct unit? From Fig. 4a, it looks to be absorbing 2.78 Pg C yr^{-1} on average over the decade.

Thank you. Indeed, this was a typo and is now fixed.

L290: "air-sea CO2 flux (negative: into the ocean)" – if negative's into the ocean shouldn't it be "sea-air CO2 flux"?

While the direction of air-sea CO2 flux is not uniformly defined in the literature, the term ‘air-sea’ is commonly used for both for some reason, see e.g. Global Carbon Budget (Friedlingstein et al., 2023): ‘air-sea flux’ is positive into the ocean; and Roobaert et al. (2023): ‘air-sea exchange’ is negative into the ocean.

L301: While STSS+ is broadly the northern bit and STSS- southern, it’s a bit more nuanced than that and that should be reflected in the text.

We have rephrased this (with credits to the other reviewer’s suggestions):

“The part of the STSS characterized by a positive \ce{CO2} flux difference between ASML and FREE, which we call the STSS+ and in which the \ce{CO2} uptake is reduced through the assimilation, roughly forms an outer (northern) ring around the STSS region.”

“In contrast, the part of the STSS characterized by a negative \ce{CO2} flux difference between ASML and FREE, which we call the STSS- and in which the \ce{CO2} uptake is increased through the assimilation, is fragmented and roughly consists of segments of an inner (southern) ring.”

Fig. 5: Add to the caption that the lines in a and b denote the regions, and the hashing (striping?) denotes STSS+.

Added this.

L462: “a pCO2-independent proxy for primary production” – I’m not sure “pCO2-independent” is needed here, I don’t quite understand what’s meant.

We agree that it is not needed here and have deleted this word.

Originally, we meant to point out that there is no direct relationship of chlorophyll and pCO_2 through carbonate chemistry - unlike for all other variables (T, S, DIC and Alk) that are included in the observation comparisons.

L480: “as the modelled phytoplankton growth is temperature-dependent” – how sure are you the change is due to the direct temperature dependence rather than the indirect influence of stratification and mixing changes?

We cannot separate these effects and have therefore rephrased the text:

“Surface chlorophyll changes follow SST changes (\cref{fig:chl} and \cref{fig:SST_glob}). As the modeled phytoplankton growth is temperature-dependent \citep{gurses2023}, the similarity of spatial patterns indicates a direct temperature effect. In addition, indirect temperature effects on plankton dynamics due to stratification and mixing changes may contribute, but the link between sea surface temperature and mixing is not straight-forward (not shown).”

As the link between sea surface temperature and mixing is not straight-forward, the temperature-dependence of growth is a more likely candidate to explain the similar spatial patterns of SST and chlorophyll changes (Figure R1).

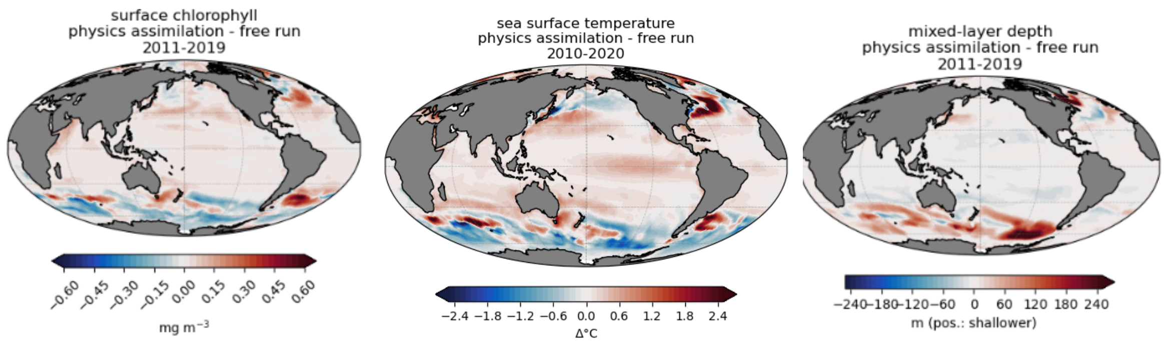

Figure R1: Spatial patterns of the difference ASML-FREE for surface chlorophyll, SST and boundary layer depth.

L515: “There are two other data assimilating BGC model approaches” – there are many other data assimilating BGC model approaches! Perhaps a more accurate phrasing might be: “We compare here to two other data assimilating BGC model approaches …”

Thank you for the rephrasing suggestion, we used it.

L524: “suggesting that a flawed representation of ocean physics as an argument for the models underestimating the CO2 flux trend is unlikely” – I broadly agree, though it may depend on how well the wider circulation is represented.
L559: “suggests that the physical processes are already well represented in FREE” – again I broadly agree, but there may still be pertinent limitations, especially depending on the time and space scale.
L565: “the adjustment of the ocean’s carbon cycle to changes in the circulation” – true, though it’s also possible that this might itself introduce biases in the carbon chemistry. See e.g. Lebehot et al. (2019, https://doi.org/10.1029/2019GB006186).

We agree with the reviewer's last three points and will mention these limitations in our discussion.