We thank Dr. Bell for his careful reviews. Please see our point-by-point reply below in red text.

Summary

The authors have replied satisfactorily to most of the technical points in my review. However the replies to my questions on the sustainability of the ARCO methodology were weak and did not respond very directly to my questions (other than the last one) in the normal way. More importantly the manuscript is largely unchanged in this section so the questions that I suggested many readers would have unanswered in their minds when reading the original draft are not answered by the revised version. Andy Hogg also asked that the discussion on the sustainability issue include a more objective discussion of the pros and cons and the reply on that point was similarly weak. As this issue lies at the heart of the paper, the authors really need to go through another round of revisions focused on this part of the paper. Providing the authors give a reasoned and objective response on this issue I would still expect to recommend that the paper is accepted.
We apologize for not having addressed the referee's previous concerns thoroughly. We hope our replies below and edits to the manuscript have satisfactorily addressed them.

More detailed points

These more detailed notes refer to the sections and paragraph numbers of the original review.

Presentation of results

1. The responses on this point are very thorough and the paper has been revised appropriately.
   Thank you.

2. Yes, a detailed examination of this issue could be quite extensive. I'm willing to accept that it is outside the scope of this paper.
   Thank you.

3. The authors have changed the colour scale in figure 3 but the dynamic range is no better. So I still think that the colour scales in figures 2, 3 and 5 could be improved but will not insist that they are.
   We have changed the colormap for Figure 2. Regarding Figures 3 and 5, we have tried to choose color schemes which increase monotonically in color saturation and brightness in order to be friendly to readers with color-vision deficiency (cf. Crameri et al., 2020).

4. The authors give a reason for keeping the figures as they are. This is a minor presentational point so I am content with the response.
   Thank you.

5. The paper has been revised appropriately.
   Thank you.

6. I'm concerned that the results presented do not properly support the statement "Considering the difference between simulations tidally forced and not, it is likely that in order to emulate the upcoming SWOT observations, applying tidal forcing is a key aspect in addition to model resolution (Savage et al., 2017a, b; Arbic et al., 2018)" and that this could be quoted out of context. It is clear that to compare with SWOT or altimeter data, tidal forcing and atmospheric pressure loading need to be taken into account. But there are pros and cons to doing this interactively. Do the references given supply that evidence? The text that has been added in line 149 does not fit into the rest of the paragraph properly so needs to be revised.
   We have added three more citations which address the importance of tidally-forced simulations in examining the eddy-wave interaction, which SWOT is expected to observe. We have also changed the sentence to: "The difference we find between simulations tidally forced and not is consistent with previous studies which argue that in order to emulate the upcoming SWOT observations, applying tidal forcing is a key aspect in addition to model resolution…"

7. Both the figure and supporting text have been improved.
   Thank you.

8. Fine
   Thank you.

9. The authors make a very valid point in response. But it would be helpful to use a common range (8 10-9) for 3 models (GIGATL, HYCOM50 and FESOM-GS) and 1.2 10-8 for eNATL60 and LLC4320. The Ce range for GIGATL could be made 0.07 to be in line with eNATL60 and FESOM-GS. I hope these minor changes would be easy to do. They would facilitate comparison.
   Done.

10. This point was worded slightly differently from the one I questioned in point 6 above. This wording is OK.
    Thank you.

11. The additional explanation of the calculation of C(t) is an important addition and the additional plots were requested by other reviewers. I still think that one would usually plot C(t)*MLI on the x-axis. The slope of the scatter fit would then be shallower than the 1:1 line – which is what one expects to see when the fit is not particularly good. If this is relatively easy to do the authors should do that.
    As the actual diagnosed submesoscale flux ($\overline{w^s b^s}$) takes both signs but C(t)*MLI only takes positive values, we argue that having $\overline{w^s b^s}$ on the x axes makes the figure easier to read with the one-to-one line plotted against axes with logarithmic scaling.

Sustainability of ARCO methodology

As I said in the summary, the responses for this section didn't respond directly to most of my questions. More importantly only very minor changes to the wording of the paper have been made in response to my comments. The same questions would come up in my mind reading the revised paper as the original one. Could you produce a revised version of this section?
Regarding the referee's previous point: "*The largest data sets have to be stored on cheaper forms of data storage like cartridges that are slow to load up on a system. If a data set is spread across 100s of cartridges access to the data will be slow/expensive. So to be analysis-ready, data has to be sub-setted in a way that suits the type of analysis that will be performed.*", the Zarrification of the model outputs to optimize them for cloud computing can be considered as subsetting the data. The step of Zarrification, however, can be omitted if the outputs were directly saved in Zarr format. As for a case in which no subsetting was applied, we can point to the example of the LLC4320 analysis. Their entire, hourly 3D outputs are stored on the NASA Ames supercomputer and made publicly accessible via their data portal. While the bandwidth of the portal was indeed a bottleneck in analyzing the LLC4320 data, we were still able to apply our cloud-based parallelized analyses to it in a consistent manner with the other model outputs. This exemplifies that terabytes of data can be systematically analyzed without specific subsetting, albeit the efficiency being dependent on the data format (i.e. LLC4320 data is stored in binary format).
As the referee also previously noted, geographical proximity between the data and cloud-computing resources is indeed important.
The management of the ECCO data portal is outside of Pangeo Forge and is done independently by NASA so we have added the points above in Section 2.1 (lines 94-98) as: "Regarding LLC4320, the data were accessed via the ECCO data portal. While there was no particular sub-setting applied to their dataset prior to analyses, the data portal and cloud-based JupyterHub being within geographical proximity (within the U.S.) facilitated the data access. The combination of `llcreader` of the `xmitgcm` Python package to access their data in binary format (as opposed to NetCDF) also enhanced the efficiency (Abernathey, 2019; Abernathey et al., 2021)."
Please also see the "Changes-to-manuscript.pdf" document enclosed in our revision.

On the point about 1000 Euros per month I still don't really understand what this means. How many users can be supported for 1000 Euros per month and how does the cost scale as the number of users and the number of ARCO data sets are increased?
We have added: "... adds up to roughly 1000€ per month for up to three simultaneous full-time users… As of writing, we have consumed 3.5 tera hours of CPU and 92.1 terabytes of RAM monthly on average."
The scaling of cost is dependent on the deal negotiated between the party of interest and Google Cloud Platform (GCP). Here, our contract with GCP allowed for roughly 1000€ per month as of writing. We have added this in line 291 as: "(We note that the operational cost somewhat depends on the contract negotiated amongst the party of interest, GCP and 2i2c.)"
Regarding the cloud storage of ARCO data sets, all of the major cloud providers have public dataset programs to support free hosting of scientific data. For this reason, the cost of storage for this type of data is not exactly a commodified product with discrete unit pricing (like hard drives might be). For Pangeo Forge, the OSN storage allocation is part of the

project grant, which currently is not associated with monetary expense for the storage. We have re-wrote the lines 278-282 as: “Currently as of writing, the JupyterHub on Google Cloud Platform (GCP) is funded by a Centre National d'Études Spatiales (CNES), grant acquired by the MultiscalE Ocean Modeling (MEOM) group at the Institut de Géosciences de l’Environnement, and the operational cost of fluxing data to the OSN cloud storage by an NSF grant acquired by the Climate Data Science Laboratory at Columbia University. (The OSN storage itself allocated to Pangeo Forge is not associated with monetary expense nor any egress fees; https://www.openstoragenetwork.org/get-involved/get-an-allocation/.)”

Minor points

These are fine. I just suggest that on line 17 “each party of interest (often an independent group)” is changed to “each of the interested parties (or an independent group)”.
Adopted.

Reference

- Abernathey, R.P. Petabytes of Ocean Data, Part I: NASA ECCO Data Portal. https://medium.com/pangeo/petabytes-of-ocean-data-part-1-nasa-ecco-data-portal-81e3c5e077be, 2019.
- Abernathey, R.P. et al. `xmitgcm`: Read MITgcm mds binary files into xarray. doi:10.5281/zenodo.596253, 2021.
- Crameri, F., Shephard, G.E. & Heron, P.J. The misuse of colour in science communication. Nature Comm. doi:10.1038/s41467-020-19160-7, 2020.