Author response to the Referee Comments by Reviewer 2 Céline Heuzé on the manuscript:

Evaluation of MITgcm-based ocean reanalyses for the Southern Ocean

Yoshihiro Nakayama, Alena Malyarenko, Hong Zhang, Ou Wang, Matthis Auger, Yafei Nie, Ian Fenty, Matthew Mazloff, Köhl Armin, and Dimitris Menemenlis

submitted to Geoscientific Model Developments (https://doi.org/10.5194/egusphere-2024-727)

We thank the Reviewer for the constructive comments. The reviewer's comments are displayed in *italics*, replies are shown in **bold text**.

Good paper that will be heavily cited. I have one somewhat-major suggestion to make it even more useful to the overall Southern Ocean community, and a series of minor ones that are mostly esthetic. In the following, the ocean reanalyses and state estimates that are the topic of this paper are referred to as "ORAs".

The one major comment: Show us the mixed layer depth. ORAs are too often used instead of observations for both physics and BGC studies when one wants to look at the time evolution of their favourite process (e.g. carbon cycle, primary production). Sure, I am biased in what I read, but this process often relies on a realistic representation of the mixed layers. Besides, it is surprising that you comment on the Weddell Sea hydrography and gyre strength and lament at the representation of both AABW and shelf processes and yet do not check that the mixed layers there are not-too-spurious.

I would recommend either a table listing for each ORA, for summer and winter separately, ideally for the Weddell and Ross seas separately, the minimum – median – maximum mixed layer depths, or at least a map of the temporal maximum mixed layer depth for the entire Southern Ocean. The time evolution (for each sea) of the spatial maximum in a style similar to Fig 2 would also be very useful.

As suggested, we add two figures (Figures 6 and 7) showing the MLD spatial distributions and timeseries. We also include new text describing our new results (Lines 196-211).

Minor comments, in order of appearance: The text contains many typos, starting with the title: ocean reanalysEs, not reanalysIs; and affiliation 5: InsTitute (T missing). I trust a careful reading of the text by the authors and then the copy editors will fix the others.

Thank you for pointing this out. We have carefully checked the manuscript before resubmission.

Table 1: Add which product used which atmosphere (see later comment), and the year of the bathymetry, since you use this year disparity to explain some of the differences later in the text.

As suggested, we include the names of atmospheric reanalysis products in Table 1. We are not able to include the year of bathymetry as this differs depending on regions and observational campaigns. Instead, we include the reference of bathymetric products.

In section 2, the methods used to calculate the sea ice extent and volume are not defined. In particular, was a threshold in sea ice concentration used, and if so, which? There is also no information about the sea ice observation source aside from a reference to "satellites" in the figure caption.

Thanks for the reminder. We have included the information about the threshold of sea ice concentration and data sources as in section 2.5.

Figure 2 onwards: I appreciate that your colour scheme was consistent across figures. I would however recommend that you modify it to increase readability. Red and magenta are hard to distinguish – swapping red for a yellowish orange should work better, and make it compatible with red/green colour-blindness. Dark blue and black are also hard to distinguish – swapping dark blue for a lighter blue such as rgb(0, 150, 255) should work while still remaining distinct enough from green.

Done.

Figure 3: The observational line is missing. You should have data since the early 2000s at least (Envisat)

The observational reference is now added in Figure 3. Two products are used here; one is derived from the Envisat (ES) and CryoSat-2 (CS2) radar altimetry measurements (Hendricks et al., 2018a, 2018b), which covers the period from June 2002 to April 2017. The other was obtained by altimeters on board the satellites SARAL (Ka-band) and CS2 (Ku-band) and covers from May 2013 to October 2018. A detailed description of the observation data has now been added in section 2.5 of the revised manuscript.

Line 195: I acknowledge that the main objective of the paper is to describe the performance of the ORAs, not to explain their biases. Yet here and there, you do, relating the biases to bathymetry or shelf processes. This line / entire section presents an opportunity for another one: GECCO3 is systematically weaker than the other ORAs, and it also is the only one that uses NCEP rather than ERA5. Can there be an explanation in the NCEP winds?

Done (Line 243-245).

Figures 9-10 + 12-13 and Table 3: For panels b,d,f,h,j and the trends, can you please compare them on a similar time period, or at least a similar time span? There could be interannual variability masking or enhancing the trends that you show. Or filter all the data to make sure that you really are showing trends, but that is more effort and not feasible anyway for the short ORAs.

The primary purpose of this manuscript is to raise awareness among readers (users) about the importance of model drift when using global ocean simulations. For our analyses, we present two examples from the Weddell and Ross Seas, where model drifts significantly influence the results. These examples illustrate that model drifts can have a substantial impact, making it challenging to distinguish realistic ocean signals from model-induced trends, especially without extensively long model spin-ups. Thus, calculating trends for the same period or filtering all the data may convey a misleading message to readers. We emphasize these points in the revised manuscript (Lines 300-307).

Figures 11 and 14: The observational row is missing. You have the data since you show them on Figs 9-10 and 12-13

Done.

Section 3.5: I have no more comment, but wanted to say that I really liked this section.

Thank you!