Review of "Arctic Ocean Simulations in the CMIP6 Ocean Model Intercomparison Project (OMIP)"

This study compares the simulated Arctic Ocean across models in the latest OMIP experiments (OMIP-1 and OMIP-2 as designated in the manuscript) to those of previous CORE-II experiments by looking at several diagnostics including mean hydrography, liquid freshwater content, and transports through the major Arctic gateways. As a standard model intercomparison paper, the results are straightforward, and I recommend the manuscript for publication after the following concerns are addressed:

1. Methodology: The inter-model spread is used as a measure of the differences amongst the models and is defined as 1 standard deviation of a given value across the models. However, several of the models used in the OMIP have the same sea ice-ocean components (Table 1). How does this impact the spread and the multi-model mean when effectively some models are being double counted (i.e., those that use NEMO3.6 as the ocean model, or those that are MOM-based models under the hood)? It may not matter a ton here as I think the results are internally consistent by comparing across OMIPs, as long as the standard deviations aren't being used to make statistical inferences, but it might be good to clarify this a bit in the text if it's something that the authors thought about.

2. In a couple of locations in the manuscript it is noted that the AWI model performs best. Presumably, the 'best' model is defined is the one that is closest to observations? However, what exactly are the requirements for defining that especially when spatial variation is involved? The text could use a little more clarification regarding this. Another option is to add some difference maps between the models and observations in the supplementary material if that helps make the comparisons more obvious (I will leave that to the discretion of the authors though).

3. Citations: there are three other Arctic Ocean CMIP6 studies that could be cited in this manuscript in much the same way that the Wang et al. 2022b and Khosravi et al 2022 papers are invoked to connect the results of this study to its CMIP6 counterparts. Specifically, these studies warrant mention in the introduction and summary/conclusions (e.g., line 390) as well as in specific areas such as Sections 2.3 and 2.5.2 (Heuzé et al., *in review*), Section 2.4 (Muilwijk et al 2022), Sections 2.2, 2.5.1, and 2.5.3 (Zanowski et al., 2021, e.g., line 267—Zanowski noted the same volume transport trends in the Bering Strait as discussed here but in CMIP6 coupled models). References below:

   Heuzé, C**.,** H. Zanowski**,** S. Karam and M. Muilwijk: The deep Arctic Ocean and Fram Strait in CMIP6 models (*J. Climate*, in review). Preprint: https://eartharxiv.org/repository/view/3233/

Muilwijk, M., A. Nummelin, C. Heuzé, I. V. Polyakov, H. Zanowski, and L. H. Smedsrud: Divergence in climate model projections of future Arctic Ocean stratification and hydrography (J. Climate; https://doi.org/10.1175/JCLI-D-22-0349.1)

Zanowski, H.**,** A. Jahn, and M.M. Holland, 2021: Arctic Ocean freshwater in CMIP6 ensembles: Declining sea ice, increasing ocean storage and export, *JGR: Oceans,* **126**, doi.org/10.1029/2020JC016930

4. Summary/Conclusions: The manuscript could benefit from further commentary about what it is we learned about the Arctic Ocean simulations from the OMIP comparisons that has not already been concluded in previous CMIP6 studies such as Khosravi et al 2022, Wang et al 2022b, Muilwijk et al 2022, Zanowski et al. 2021, and Heuzé et al. All of these studies note that simulation of the Arctic Ocean has not improved since CMIP5 (based on the diagnostics in those papers, of course). I am not doubting the validity or usefulness of the OMIPs or CORE simulations, but rather it would be helpful to place this study in the context of the other literature that has come out in the last year or two, and there is a nice opportunity to do that here by expanding this part of the manuscript.

5. Lines 265-266 “The reasons for the discrepancy between observations and simulations are unknown and should be further investigated.” Regarding the issues with the models being unable to reproduce the correct sign of the Bering Strait volume transport trend, didn’t Wang et al 2022b (figure 8) suggest that this might be due to changes in the sea surface height gradient between the Arctic and Pacific? That may have been analysis for the future forcing scenarios and not the present-day trends as is the case in this manuscript, but it may also be worth commenting on. I have been wondering about the negative model volume transport trends vs. the positive trends observed by Woodgate et al. as well.

6. A note on rainbow colormaps: It is best to avoid using a rainbow colormap where possible (there’s a lot of information out there about why it’s problematic). Please consider changing the colormaps for the figures to something non-rainbow. If any of the authors use python, cmocean has a nice set of pre-defined colormaps: https://matplotlib.org/cmocean/