Review of gmd-2023-123
Title: Impact of high resolution on Arctic Ocean simulations in Ocean Model Intercomparison Project phase 2 (OMIP-2)

General comments
The topic of the study is very interesting and important. The ocean component of current coarse resolution climate models (CMIP6) has large biases in the Arctic Ocean, and future climate projections of the hydrography in the Arctic Ocean show a large model spread. Thus, it is necessary to investigate approaches that can reduce the model biases, such as using higher resolution versions of the ocean components. This is the focus of this study, to investigate how higher resolution versions improves the hydrography and ocean circulation in the Arctic Ocean compared to lower resolution versions. The authors use 5 different ocean components with realistic atmospheric forcing, where each component has a high and a low resolution version.

The manuscript is overall nicely written and has very interesting results with good figures. However, there are two major comments: **(1)** Some discussions points are missing in the present study, as listed below. Although discussion points, I think it is important to discuss some of these points early in the manuscript, perhaps as part of the approach used in the study (i.e., in Section 2). **(2)** The structure of Section 4 (Conclusion and Discussion) is to some extent difficult to follow. I therefore recommend major revision. I believe the authors will be able to improve on these two points. In addition, there are parts of the manuscript text that could be clarified and/or improved. I've pointed to some of these places further below (see 'Comments per section').

(1) The manuscript needs discussion:
- On how representative the 5 ocean models are compared to other CMIP6 ocean models (Shu et al., 2023). It seems that some of the 5 ocean models used in this study have poor performance compared to other CMIP6 ocean models.
- On the very different sets of resolution for each model. For instance, one ocean model has 9km as low resolution and 3.6km as high resolution, whereas another model has 72km as low resolution and 6.8km as high resolution.
- On the impact of increased horizontal resolution when the vertical resolution is changing at the same time (in some cases the parameterization or sea ice model also changes when moving from low to high resolution version). This is to some extent discussed in Section 4.
- On the deep-water balance, as the spin-up does not appear to be long enough.
- On the benefits of high resolution version when the low resolution has already good performance.*

*The study shows that the high resolution versions outperforms the low resolution versions when the latter is showing poor performance (3 out of 5 models, where the low resolutions are 9km, 32km, and 51km). But when the low resolution has good performance, the benefits of the high resolution are not that clear (AWI-FESOM). In addition, one model shows that higher resolution does not necessarily lead to improvements (IAP-LICOM).
*The good performance of the low resolution of AWI-FESOM is mentioned in Section 4 (lines 610-612), and the discussion on lines 613-614 is very interesting. Could you please expand on this discussion; why is low resolution AWI-FESOM working quite well despite its coarse resolution?

(2) The structure and message of Section 4 needs to be clearer.
- Section 4 is, as it stands now, sometimes repetitive and needs more structure.
- The results could be more concrete; the results are sometimes a bit vague. For instance, how much is the performance improved by increasing the horizontal resolution? Could it be quantified?
- The results seem to be sometimes simplified, e.g., line 570. Could the discussion be more nuanced? For instance, one model shows good performance for both versions, one model shows overall poor results, whereas three models show poor results for low resolution but improved for the high resolution.
- As the five different models shows quite different results (and the configurations are different from model to model), it might be useful to discuss each of the models separately.
- I think it would benefit the manuscript to split Section 4 into two sections; first Discussion and thereafter summarize and conclude the study in Conclusion.

Comments per section

Abstract
Line 6: The main improvements appear to be seen in the Fram Strait and Davis Strait, but not much improvement seems to be the case in the Bering Strait. Suggest making the 'Arctic gateways' more nuanced.

Introduction
Line 15: Note that the amplified Arctic Ocean warming compared to global is for the future, or is the 'two times faster' referring to present?

Please clarify which models are ocean components forced with realistic atmosphere, coupled, regional or global models.

Results
Line 168-169: How does this relate to Heuze et al. (2023; this study is already in your reference list) that finds that the ocean heat transport through the Fram Strait is reasonable, but the volume transport and temperature is not correct?

Line 175: The recent study by Richards et al. (2022) might be useful here.

Line 172-180: Here the boundary currents are discussed. Have you looked into the ocean circulation/boundary currents in the models?

Line 184-185: This sentence needs a reference. Could there be cold sources in the deep also entering through the Fram Strait, such as Norwegian Sea Deep Water?

Line 192: The results described here are difficult to see from the figures.

Line 198: It is not fully clear how the absence of the sea ice dynamics impacts the results.

Line 219: The discussion on how the sea surface salinity restoring impacts the results is not fully clear to me.

Line 225: A reference is needed.

Line 254: Poleward propagation of warm anomalies in the Atlantic Water layer is discussed in Årthun et al. (2017).

Line 259: How is sea ice decline leading to intensified inflow?

Line 260: Please be consistent, use either Arctic basin or Arctic Basin.

Line 270-272: This is not so easy to see from the figures.

Line 280: Please explain why these two periods are compared.

Line 349: Note that a different approach is taken in Muilwijk et al. (2023) studying the halocline in CMIP6 models.

Line 363: Is it the deepening of the Atlantic Water layer that causes the deepening of the halocline, or opposite, or other factors causing the deepening of the layers?

Line 368: Should I expect to see that 1990s is warmer than 2010s for FSU-HYCOM? Please clarify.

Line 417: A positive salinity bias both in the Amerasian Basin and the Eurasian Basin?

Conclusion and Discussion
Please see the comments above regarding Section 4.

Several places in this section it is mentioned that higher resolution helps, but at the same time other findings reveal that the answer is not that straightforward (for instance AWI-FESOM and IAP-LICOM show different results than the three other models). This comment is similar to my comment above, about giving a bit more nuanced discussion.

Line 629: The results from Pan et al. (2023) are relevant here.

Figures
Fig. 14: The dark red color in figure 14j is very difficult to see. Perhaps write a note in the caption that the Amerasian Basin is deep red in 14j.

References:
Årthun, M., Eldevik, T., Viste, E. *et al.* Skillful prediction of northern climate provided by the ocean. *Nat Commun* 8, 15875 (2017). https://doi.org/10.1038/ncomms15875

Richards, A. E., Johnson, H. L., & Lique, C. (2022). Spatial and temporal variability of Atlantic Water in the Arctic from 40 years of observations. Journal of Geophysical Research: Oceans, 127, e2021JC018358. https://doi.org/10.1029/2021JC018358