

Review of GMD-2022-260-version2

Thank you for addressing my comments for the first version of the manuscript. I am pleased to see that you have picked out specific models that stand out and provided some interpretation of the model results. I find the 2nd version of the manuscript satisfactory and recommend it be published after a few minor adjustments. These are specified below.

Comments

OMIP versus “standard” CMIP6 models

For the reader who is unfamiliar with the OMIP initiative, it is perhaps not so clear what the difference is compared to the CMIP6. Perhaps you could clarify; are the OMIP models global or regional domain? How do they differ from the fully coupled models used in CMIP6? Just one-two sentences before the Methods.

L75: I presume you mean “have not been evaluated in a systematic/collective manner?” I would assume that some of the models have been evaluated individually before?

L115: Perhaps you can add a sentence mentioning that observations are also not without problems/biases and should not be taken as the complete truth.

L130: Using the potential temperature at 400m depth to evaluate the AW makes sense when comparing to previous studies. I would still consider looking at the maximum temperature within a depth interval (or 0-degree isotherm), which could account for differences in the depth of the AW layer between models. By choosing a fixed depth you may bias the results. At least, it is something you should check.

L173-175: You state that there are no improvements in the mean state in the MMM. This is somewhat depressing – so I wonder if there are any improvements in any individual model?

L183: “meter” → “meters”

L212: It would be nice if you could include a small paragraph about the model’s ability to represent stratification in the Beaufort Sea in particular; i.e., how well do the simulate Pacific Summer Water (PSW). This would be interesting to report on since the models seem to capture the volume through BS quite well.

L311-313: Are the biases in the BG due to incorrect atmospheric forcing or model physics? Please clarify.

L317: Can you please clarify if the freshwater transport include both liquid and solid freshwater contributions? What is the relative importance of solid/liquid transport?

L132: What about vertical resolution?

L354: “... to 2018 and the trend ...” → “...although the trend”

L358: Why is the spread in volume transport for the models so low compared to the observations?

L374: “The OMIP models obtained upward trends ...” → “The OMIP models also obtained positive trends”

L380: It's interesting that the models using NEMO seem to have some common biases. Can you comment on why this might be the case? Perhaps mixing schemes? Too strong AMOC preconditioning a stronger inflow across the Greenland-Scotland Ridge? Or representation of bathymetry in the inflow region (e.g Heuzé and Årthun 2019 10.1525/elementa.354)?

L465: "... implying that changes in the upper Arctic Ocean are captured in the models": In terms of what? Please specify. It's possible for the model to simulate the SSH correctly but still not capture the water mass structure accurately.

L495: Great! Thanks for suggesting this – I agree.

L502: I wonder if you really resolve eddies at 4.5 km outside the deep Arctic basin, e.g., on the shelf regions in the Barents and Kara Sea where you have Atlantic Water? Consider changing/adding the reference to Wang et al 2020 ([https://doi.org/ 10.1029/2020GL088550](https://doi.org/10.1029/2020GL088550)) which uses a horizontal resolution of 1 km.

Figures

Fig 3+4: If possible I would change the colormap for a,b, c and f, g h to one that is not divergent. For example "Thermal" for a,b, c and "Haline" for f, g,h. The divergent colormap implies differences/anomalies.

Figure 9, 10, 11, 13: Same as above. There are better colormaps available for displaying these metrics (see cmocean) and will help the communicate the figures to the reader more effectively. Note that people associate certain colormaps with specific things.

The comments on the figures are minor details, but I would still encourage the authors to consider updating the figures with more appropriate colormaps. I will leave it to the authors and the editor to make that decision.