Comment on gmd-2021-293
David Bailey (Referee)

This manuscript describes a regionally refined (Arctic) version of E3SM and compares to the RASM model both in JRA forced sea ice-ocean configuration. Overall this is well-written and easy to understand. However, I do have some thoughts here regarding the details of the model and also the originality of the work. I will also note that the categories of "Scientific Significance" and "Scientific Quality" do not really apply here. I believe that the manuscript only provides information about model performance with an without regional refinement and really says nothing about Arctic science. This is probably fine for GMD, but why are these a part of the evaluation?

1. My first major comment is that this is kind of a rehashing of Petersen et al. 2019. While these simulations are forced with JRA as opposed to CORE, the model details are the same in both papers as I see it. The regionally refined aspect is new, but it would be helpful for the authors to contrast the model details from this paper and Petersen et al. 2019. Perhaps a table would be helpful here.

2. One of the challenges of regionally refined grids is the scale aware parameterizations. I would like to see more details here about how the parameterizations adapt from the coarse resolution part of the domain to the finer resolution part of the domain. Again, this would be an interesting thing to contrast with Petersen et al. 2019.

3. I am curious why the authors chose 3 cycles of JRA? I see the salt and temperature trends are still very strong here and particularly above 1000m depth. This seems to accelerate in the 3rd cycle. Are you using some kind of temperature or salinity restoring? This is often required in a data atmosphere forced mode. I see now where the restoring is 1 year. This is fairly weak. A discussion of how the restoring would impact the drift would be a good addition here.

4. With the drift, it is sort of difficult to compare the third cycle of the E3SM simulations to observations or only to one cycle of RASM. Perhaps it would be better to compare the first cycles.

5. It is interesting that the MOC seems to be fairly well equilibrated (Figure 5b) after 1 cycle despite the heat and salt trends. Also, the MOC max transport is significantly higher in the E3SM-Arctic simulation. It would be interesting to contrast with Bryan, Gent, and
Tomas (2014) here in which they found that differences in parameterizations at one degree and 0.1-degree led to differences in the poleward heat transport and hence sea ice differences.

6. Figures 7, 8, and 9 are interesting. The Nares strait is not resolved in the LR simulations as mentioned. I would like to see more discussion on how this impacts the downstream transport and the MOC. Is there a shift in transport from the Canadian Arctic to Fram Strait?

7. Figures 10, 11, and 12 are helpful in understanding the vertical distribution differences. How much do these change between cycles? That is, how are they impacted by the drift.

8. Figure 14 and 15 show the central Arctic profiles of temperature and salinity. This is one of the areas where I am worried about the comparison of the third cycle in the E3SM runs to the first cycle in the RASM runs. The drift in the model causes these profiles to move further from the initialization from PHC. This is always an issue with ocean models. That said, are the RASM runs initialized differently? What is going on here with the RASM temperature profiles? Is vertical resolution playing a role here?

9. In addition to the above, some of the most interesting seasonal differences in 14 and 15 are above 100m. Instead of density, perhaps you could zoom in on the top 100m instead. The surface temperature and salinity biases are indicative of some ice-ocean coupling issues. Do the salt and fresh water fluxes assume 4 psu in the sea ice?

10. Figure 16 is interesting, but might be a candidate for removal. Perhaps a discussion and leading to another paper. I feel like there could be so much more here and the discussion is fairly limited.

11. It is nice to see that the sea ice volume and area are consistent across the three cycles for the most part. They impacted a bit by the upper ocean drift. There is something funny in the volume before 1995 and after. The trend is faster than PIOMAS before 1995 and then slower after 1995. I worry this is an artifact of the JRA forcing.

12. On the JRA forcing. Are you doing corrections over the sea ice? There are definitely some biases in the JRA in the polar regions. While there were corrections added some biases remain different than the CORE forcing.

13. What are the lateral boundary conditions for the sea ice and ocean in the RASM runs?

14. In Figure 19, I find it interesting that all of the model simulations have a thickness bias in the Beaufort Sea. It might not be fair to compare longer periods in the model to such a short period of IceSat. There could be some substantial interannual variability here. The biases near the CAA might just be something in the IceSat data. I’m not sure this figure is helpful. Also, I wonder if the "bias" in the Beaufort Sea thickness could be related to freshwater content in the Beaufort gyre? Maybe adding the RASM thickness in Figure 18 instead?

15. Back to my earlier point about scale aware parameterizations. Does this feed into the discussion and conclusions about the regionally refined configuration versus the low res configuration? Also, the JRA biases are key here. Ultimately, even though one improves resolution in the sea ice, the thickness distribution in particular is generally set by the wind patterns as mentioned earlier in the manuscript.

Minor comments:

1. On line 65, please delete "occasionally".
2. How are the sea ice albedo parameters set/tuned?

3. I see this is the B-grid versions of the models. Any thoughts about C-grid? How does the sea ice compare to Turner et al. (2021) which is in review?

4. Some more details of the RASM simulation would be helpful.