Response to reviewer #1

This study describes a new version of the AWI-CM3 and shows that it has good skills in representing the observed climatology and better computational efficiency compared to previous versions. Also, authors discuss model biases and give some suggestions to make further improvement. The development of new version is very impressive and represents the cutting edge of Earth system modeling. The results are noteworthy and can potentially help understand model biases. I believe this manuscript provides a useful roadmap for ongoing challenge to develop next generation of Earth system model including development of high resolution climate model. I recommend publishing this paper after some minor revisions.

> *We would like to thank the reviewer for the encouraging and very positive feedback.*

Suggested improvements are as follows:

1. Evaluation of model performance.

Authors calculated climate model performance indices based on the ideas of Reicher and Kim (2008). What about temporal variability? Does this method consider the variability? If not, I would suggest adding more analysis on climate variability. Regarding the interannual variability authors show ENSO timeseries only. Spatial patten and amplitude of interannual variability as well as seasonal cycle are also important to evaluate model's performance.

> *The method of Reichler and Kim (2008) can be used to consider basic climate variability. To do so we can infer the climate variability from both the observational data, as well as from CMIP6 and AWI-CM3. We have added the standard deviation of sea surface temperature and sea surface height as general measures of model variability. For both the existing and newly added variables in the performance index we show biases with seasonal time resolution.*
>
> *Furthermore we added mixed layer depth as one of the key areas of future work on AWI-CM3.*
>
> *We also updated the list of CMIP6 model simulations against which we normalize our model performance (see: https://github.com/JanStreffing/cmpi-tool/issues/1). We ensured that for each model family we use only the highest resolution, lowest complexity (in terms of ESM-components) version. In doing so the CMIP6 ensemble considered has become somewhat more skillful, leading to a slight drop in the relative performance of our AWI-CM3 prototype.*

Reichler, Thomas, and Junsu Kim. "How well do coupled models simulate today's climate?." Bulletin of the American Meteorological Society 89.3 (2008): 303-312.

2. Figure 1 and Table 1: It would be great if authors can combine the information in Fig.1 and Table 1. But, if it looks too busy, ignore my suggestion.

We agree with the suggestion and combined Figure 1 and Table 1.

3. Figure 5 and its description: Please add plots for observations to compare spatial distribution and temporal evolution of sea ice. And, what is the dotted lines in Fig. 5I?

We added GIOMAS reanalysis for the spatial distribution of sea ice thickness. We believe that the sea ice extent plot is already quite busy, and that the thickness reanalysis gives a good overview of the extent match/mismatch. The dotted lines are merely mean as visual aid for a plot that is rather wide.

4. I don't think Fig. 9a is necessary. Instead, it would be great if authors show spatial patten of AWI-CM3 SST variability. Does this model show seasonal locking of ENSO?
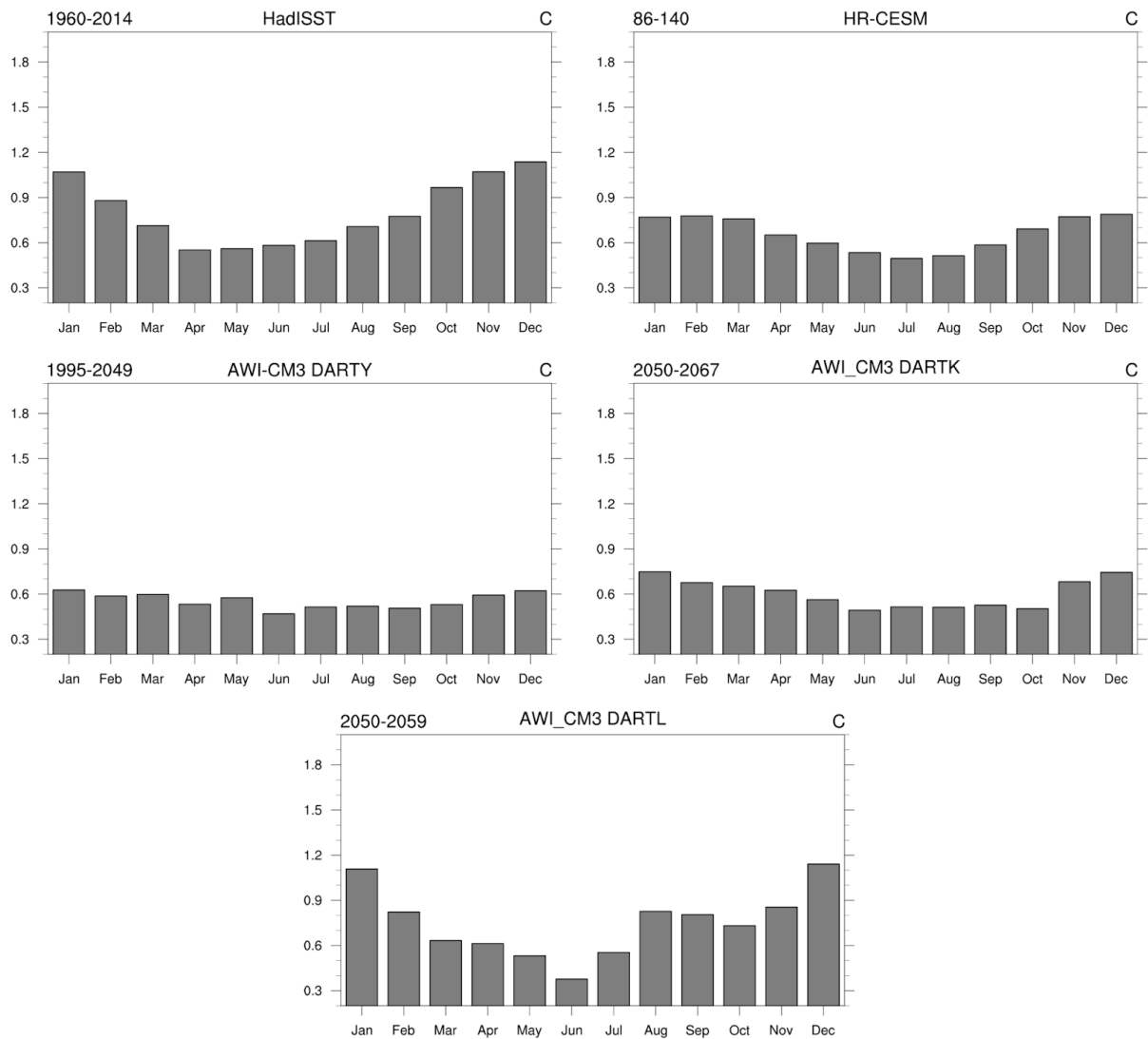
We have modified Fig. 9a such that it shows the EOF1 as correlation based on: https://ajdawson.github.io/eofs/latest/userguide/overview.html#eof-analysis-with-eofs

We have worked on ENSO phase locking in the 6 months since we ran the simulations upon which this paper is based. Initially phase locking was found to be weak. After improving the near equators SST and Precipitations biases phase locking looks better in newer versions of the model. We added one sentence about the weak phase locking in the model versions we present here.

Since this work is outside the scope of the project and partially by non co-authors, we do not add it into the paper. Here is a preliminary plot for the reviewer showing ongoing work on ENSO phase locking:

5. Figure 6: I don't see any description for Fig. 6c.

We have added the missing descriptions for Fig. 6c.

6. Figure 7: I guess this is annual mean climatology. Please specify clearly.

We have clarified that the plots indeed show annual mean biases

7. Figure 7 & 11 label bar: To describe the results precisely, authors should indicate each values

of color bar since the scale does not increase/decrease uniformly.

The color bar is symmetric around 0 and omits every other value. The values that have no values on the positive side have one on the negative side and vice versa.

8. Section 2.4: I think XIOS works for OpenIFS only. Then, is there any IO scheme for FESOM?

We have added a sentence on the bespoke parallel IO scheme included in FESOM2

9. Line 104: I'm wondering if the ocean basins with narrow outflow can be included in the

DART resolution.

Indeed the DART mesh with its resolution down to 4km does include many of these narrow outflow channels. We have added a sentence that states this explicitly.

10. Line 225: The list of observational datasets used to calculate all mean absolute errors is

also given in Appendix B -> Appendix C?

There seems to be something amiss with the appendix naming. We think this is a problem with the GMD latex template. We swapped the figures and table around until the Apendix letter matched the positions. Hopefully this will be double checked during typesetting.

11. Line 320: If this cold bias in the IO is likely due to inability to resolve narrow strait on the

FESOM2 CORE2 mesh, then can we see some improvements of the performance in the highresolution DART simulation?

First results from shorter (100y long) simulations do indicate this. We consider this a bit outside the scope of this paper, as the high resolution model is still being tuned.

12. Line 325: Biases presented in 8 explain -> Biases presented in Figure 8 explain

Fixed.

13. Line 388: 20 Sv along with the cooling of the upper ocean (Figure 3c) -> Figure 3b

From the context of the sentence, we were talking about the change in the time series once the total mass fixer was turned on. This timeseries is Figure 3c

14. Line 424: "m2" should be included in the unit of α.

Fixed.

15. Line 433: What is the TCR of AWI-CM1?

The TCR of AWI-CM1 is in fact the same as that of AWI-CM3, barring rounding errors (2.1°C). We rewrote this sentence to express this.

16. Line 461: What is main model improvement TCo319L137-DART compared to TCo159-CORE2.

We have added a few sentences outlining the major improvements stemming from increased model resolution.

17. Figure B1 should be Appendix D.

See #10

18. Please check use of abbreviation in the text.

E.g. Line 14: The evolution of coupled climate models between phases of the Coupled Model Intercomparison Project (CMIP) is advancing

Line 470: We ran a set of experiments closely resembling the Coupled Model Intercomparison Project phase 6 (CMIP6) DECK -> We ran a set of experiments closely resembling the CMIP6 DECK

We double checked our abbreviations. We decided to introduce abbreviations in the conclusion independently from the main text, as readers will frequently skip the text on the first skim through.

Line 439-440: the Simulated Years Per Day (SYPD) and the computational cost measured in Core Hours per Simulated Year (CHSY)

Line 448: Tables 2 and 3 list the Simulated Years Per Day (SYPD) and core hours per simulated year (CHSY) values -> Tables 2 and 3 list the SYPD and CHSY values

Fixed