

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.

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 pattern and amplitude of interannual variability as well as seasonal cycle are also important to evaluate model's performance.

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.

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?

4. I don't think Fig. 9a is necessary. Instead, it would be great if authors show spatial pattern of AWI-CM3 SST variability. Does this model show seasonal locking of ENSO?

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

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

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.

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

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

10. Line 225: The list of observational datasets used to calculate all mean absolute errors is also given in Appendix B -> Appendix C?

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 high-resolution DART simulation?

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

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

14. Line 424: “m<sup>2</sup>” should be included in the unit of  $\alpha$ .

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

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

17. Figure B1 should be Appendix D.

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

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

19. Please check reference list.

E.g. Line 580 and 588.