**Reviewer 2:**

Review of "Global ocean and sea ice variability simulated in eddy-permitting climate models"

In this manuscript the authors study the capability of an eddy-permitting model SINTEXF3 to simulate inter-annual climate variability over the global ocean and sea ice. The manuscript is submitted as a 'Model Evaluation Paper'. Based on the guidelines of GMD, the title needs to be changed. The authors need to identify the model's name and version number in the manuscript title. Furthermore, as per GMD policy, for a model evaluation paper, the model needs to be described in another paper (or it needs to be under review). I could not find any such paper describing the SINTEX-F3 model. Additionally, it seems there are possibilities of major improvements in the manuscript (which I describe below). Hence, at this point, unfortunately, I cannot recommend publication of this manuscript in this journal.

However, if the authors fix these issues and the editor decides to send a revised manuscript (as a different type other than Model Evaluation Paper) for another round of reviews, I will be happy to review the revised/resubmitted manuscript again. Here are some comments which the authors may consider when submitting a revised manuscript.

We thank the reviewer very much for providing constructive comments on the original manuscript. Those are very helpful for further improving the manuscript. Following the GDM guideline and the reviewer's suggestion, we have changed the title to include SINTEX-F3 and resubmitted the revised manuscript to "Model Description Papers". In the revised manuscript, we have added statistical metrics such as the amplitude of regression values and the pattern correlation in the regions of our interest (Table S1-11) to clarify the differences between the low- and high-resolution models. We have also examined physical processes underlying the improvement of ocean and sea ice variability in the high-resolution models. With this revision, we believe we have addressed major concerns raised by the reviewer. Our point-by-point responses to the reviewer's comments are provided below with a blue front.

Major Comments

1. In Section 3.1, why different climate models demonstrate different biases in SST and Sea Ice? I recommend the authors to cite relevant literature and add a paragraph describing the possible differences between different models that leads to different SST and sea ice. I would also recommend adding a table and quote the mean bias value for (a) entire globe (b) northern hemisphere (c) southern hemisphere for different model configurations used in this study. Also consider adding similar tables for all the lat-lon plots shown in the manuscript.

We have added one paragraph to discuss why different climate models have common or different biases in global SST and sea ice. For example, cold SST bias in the North Atlantic is commonly seen in SINTEX-F3, ECMWF-IFS-LR/MR/HR, HadGEM3-GC31-LL/MM, and MPI-ESM-1-2-XR (Fig. S1). This may be related to insufficient northward ocean heat transport by the AMOC (Wang et al. 2014) and too-zonal North Atlantic Current causing more southward intrusion of fresh and cold

Labrador Sea water (Müller et al. 2018). This is also associated with a high SIC bias in the North Atlantic (Fig. S3), weaker low-level zonal winds, and storm track activities (Athanasiadis et al. 2022). On the other hand, CESM1-CAM5-SE-HR shows stronger warm SST bias in the western boundary current regions than CESM1-CAM5-SE-LR, partly due to the difference between explicit and parameterized ocean eddy heat fluxes (Chang et al. 2020). In the eastern boundary current regions, warm SST bias is commonly seen in most climate models except HadGEM3-GC31-HM/HH. This suggests that the increased atmospheric resolution tends to produce cooler SST in the ocean upwelling region (Small et al. 2015). In the Southern Ocean, warm SST bias and low SIC bias are commonly seen in SINTEX-F2, CESM-CAM5-SE-LR/HR, CMCC-CM2-HR4/VHR4, ECMWF-IFS-LR/MR/HR, and HadGEM3-GC31-MM/HM/HH (Figs. S1, S2). This may be due to more incoming shortwave radiation associated with the underestimation of cloud albedo in the atmospheric models (Bodas-Salcedo et al. 2012) and a poor representation of ocean mesoscale eddies in the ocean models (Hallberg et al. 2013).

Following the reviewer's comments, we have also added supplementary tables summarizing the model biases by regions across models (Tables S1, 2) and referred to them in the main text. For example, SINTEX-F3 shows that cold bias in the Northern Hemisphere (-2.14 °C) is stronger than that in the Southern Hemisphere (-1.28 °C). This may be linked to a more positive SIC bias in the Arctic region (19.8 %) than in the Antarctic region (-3.44 %). We have mentioned this in the relevant paragraph.

2. In Figure 4, it seems that the SINTEX-F2 is closer to the OISST2_hi data compared to SINTEX-F3 most of the time-why? The trend in SINTEX-F3 is different compared to the observations. The authors do not clearly mention the possible mechanisms behind these differences. Also, the climate models seem to be very different compared to both SINTEX-F2 and SINTEX-F3.

As mentioned in the Discussion section, SINTEX-F3 shows weaker ENSO variability than SINTEX-F2 (Fig. 4a), probably because the model simulates weaker Walker Cell as mean state so that the SST anomalies excited in the tropical Pacific cannot develop largely through local air-sea interaction. We need further efforts to improve the mean state bias in the tropical Pacific, but this is beyond the scope of this study. Second, we have tested the statistical significance to the difference in the standard deviation of tropical climate indices between CMIP6-LR and CMIP6-HR. We have found that only ATL3 index (Fig. 4c) shows a significant difference between the models, indicating the benefits of increased model resolutions for better presentation of Atlantic Niño/Niña as well as mean state (Richter and Tokinaga, 2020). The SINTEX-F3 also captures the peak of the ATL3 index in June, while SINTEX-F2 fails to simulate the peak. This leads to the amplitude of regression values closer to the observation and higher pattern correlation of global SST anomalies associated with Atlantic Niño/Niña (Table S5). We have mentioned these results in the relevant paragraphs.

3. Figure 8-10: How sensitive are the model results to the choice of the black box? Is it possible to choose multiple such boxes and conduct more detailed analysis?
We would like to first apologize that the rectangular box in the Agulhas Retroflection region was wrongly displayed in the figures, inconsistent with the description of the box region (15-30° E,

40-42° S) in the figure caption. So, we have corrected the rectangular box in the figures. Second, we have tested different boxes in the Agulhas Retroflection region, but we have found that the results are almost similar to the case when using the original box. For example, the corresponding figures using a larger box (20-35° E, 40-44° S) are shown in Figs. S9-13. The spatial patterns of the warmer SST, larger upward surface heat flux, shallower mixed-layer depth, higher SSH, and higher SLP regression values do not differ from those using the original box. Therefore, we have remained using the original box to discuss the air-sea interaction process in the section.

4. At present, the evidence does not convincingly show that SINTEX-F3 performs significantly better than SINTEX-F2. I encourage the authors to substantiate any claimed improvements with robust statistical metrics and appropriate significance testing.
Following the reviewer's comments, we have added supplementary Tables S3-9 that summarize the amplitude of ocean and atmospheric variables in the core regions of climate indices and the pattern correlation of global SST regression values associated with tropical climate variability. As mentioned in our reply to the above comment #2, SINTEX-F3 shows significantly weaker ENSO and the IOD than SINTEX-F2, probably due to weaker Walker Cell (Tables S3, 4) in the Indo-Pacific region, but simulates significantly stronger Atlantic Niño/Niña and the associated teleconnection pattern, consistent with the CMIP6-HR model results (Table S5). In the Agulhas Retroflection region, SINTEX-F3 shows warmer SST associated with significantly larger upward surface heat flux than SINTEX-F2, closer to the observation and the CMIP6-HR model results (Table S6). Similarly, better performance of SINTEX-F3 is obtained in the Dakar Niño/Niña region with significantly warmer SST and larger upward surface heat flux regression values (Table S7). Similarly, in polar regions, SINTEX-F3 shows significantly higher SIC, larger upward surface heat flux and negative SLP regression values, closer to the observation and CMIP6-HR model results (Table S8), while the model does not show better performance in the Arctic region (Table S9). These statistical metrics and statistical testing further support better performance of SINTEX-F3 in the analysis regions. We have mentioned these metrics in the relevant paragraphs of the Results section.

Minor Comments
1. Most of the figures are not colorblind friendly. I suggest the authors to kindly use a different colormap (for example see: https://www.fabiocrameri.ch/colourmaps/), use colorblind friendly plotting techniques, and get the figures checked by Coblis Colorblind Simulator before submitting a revised manuscript. For example: in Figure 4a, dashed and dotted lines can be used.
We have modified the colors of all figures to make them colorblind friendly following the above website.

2. Figure 3: Are the model results sensitive to the 15% threshold?
We have calculated Arctic sea ice extent climatology using different thresholds of 10 % and 20 % (see figures below), and found that the results are not sensitive to the thresholds. We have mentioned this in the caption of Fig. 3.

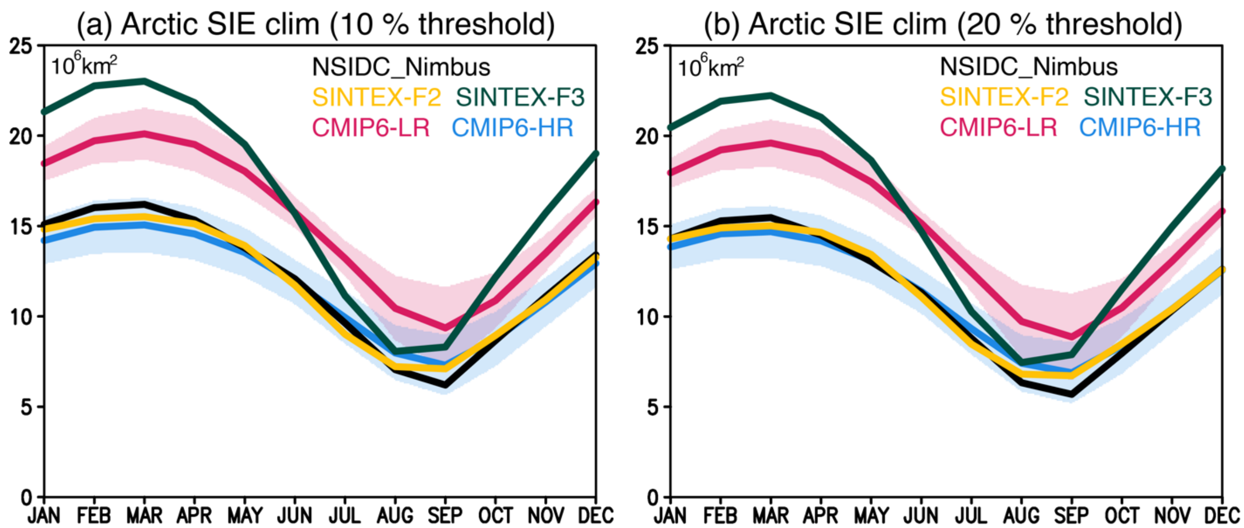


Figure R1: (a) Monthly climatology of the Arctic sea ice extent (SIE; in $10^6$ $km^2$) for the NSIDC_Nimbus (black) dataset, and SINTEX-F2 (yellow), SINTEX-F3 (green), CMIP6-LR (magenta), and CMIP6-HR (blue) models. The color shades indicate plus and minus one standard deviation of the model spreads. Here we defined the Arctic SIE as the total area with the SIC above 10 % in the Northern Hemisphere. (b) Same as in (a), but for the 20 % threshold.

3. The link to download HighResMIP model doesn't seem to work.

We have corrected the link.

4. I also recommend not to put all the figures in the end.

We have put all figures and tables (except supplementary files) in the main text near the location where they are first mentioned.

5. It would be beneficial to have the manuscript's English professionally reviewed to enhance its overall quality.

We utilized a professional English editing service for scientific journals and had the manuscript proofread by native English speakers to improve the readability.

Additional Comments

L28: I would recommend shortening the sentence.

We have shortened the sentence.

L64, L66, L77: A few more recent citations might be helpful.

We have added a few more references.

L94-144: Please consider shortening.

We have shortened those paragraphs.

L149: Please change 0.1° to 1°

We have modified.

L151: Please change 1/2 to 0.5
We have modified.

L311: Change ‘difference’ to ‘different’
We have corrected.

L342, L420: Citation needed
We have provided the references.