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 SINTEX-F3 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.

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.
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.
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?
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.

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.
2. Figure 3: Are the model results sensitive to the 15% threshold?
3. The link to download HighResMIP model doesn't seem to work.
4. I also recommend not to put all the figures in the end.
5. It would be beneficial to have the manuscript's English professionally reviewed to enhance its overall quality.

Additional Comments

L28: I would recommend shortening the sentence.

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

L94-144: Please consider shortening.

L149: Please change 0.1° to 1°

L151: Please change ½ to 0.5

L311: Change 'difference' to 'different'

L342, L420: Citation needed