Comment on gmd-2021-209

Anonymous Referee #1

Referee comment on "Impact of increased resolution on long-standing biases in HighResMIP-PRIMAVERA climate models" by Eduardo Moreno-Chamarro et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-209-RC1, 2021

Recommendation: moderate revisions

General Comments:

This manuscript presents an assessment of the HighResMIP models regarding four well-known, long-standing biases. Authors analyse simulations of 5 climate models in at least two different resolutions. Four of the five models have increased resolution in the atmospheric and oceanic components simultaneously, whereas 1 model has only increased resolution in the atmospheric component. The set of experiments consists of historical runs from 1950 to 2014 in two versions: AMIP and coupled simulations. The author's results suggest some improvements of the model performance regarding the analysed biases with increased atmospheric resolution. Yet, those improvements are not consistently found in all models. No systematic improvement is shown by increasing the resolution from eddy-parametrized into eddy-permitting ocean models. However, the only eddy-resolving ocean model presents improvements in reproducing the North Atlantic temperatures and the path of the Gulf Stream. Overall this study shows limited benefits for reducing the long-standing biases from the high resolution based on the ensemble mean of HighResMIP. This result yields a recommendation for the modelling community: in addition to model resolution, future efforts should be oriented to improve model physics.

The manuscript is well written and structured, the methodology is correctly explained, and the results are relevant. All in all, I believe that this manuscript is relevant for the climate modelling community, and it is worth to be published in GMD. In some places, the results described in the text are difficult to see in the figures, and I think that the original manuscript could be improved with a moderate revision. I report in what follows some comments and suggestions.

Specific Comments:

1) All the variables assessed in this manuscript come from the atmospheric model component. I think that analysing the model bias of at least sea surface temperature (*tos*) could help complete the study. For example, the known Southern Ocean warm bias or North Atlantic cold bias in climate models are reported for the upper-ocean temperatures. The authors assess the 2m-air temperature (*tas*), which is fair, but it is not the same value as *tos* in the open ocean. Analysing the *tos* bias in climate models has an only sense for the coupled runs. But still, a different dataset (from ERA-interim) as observations should be used, and you could get new results. I would recommend looking at sea surface temperatures too.

2) When discussing model biases regionally (in particular in sections 4.1, 5.1, 5.2), I think that a table similar to Table 2 (which I found very clear and helpful) could help to summarise the results. I suggest including additional tables in which the metrics RMSD and Bias are regionally computed.

3) I may be wrong since I am not a native English speaker, but I find the concept 'improving/worsening the bias' (used several times along the manuscript) a little bit confusing. I would say that the model is improved/worsened by a reduction/increase in bias.

Minor Comments:

L182. Please re-order (Kato et al., 2018; Loeb et al., 2018).

L189-190. 'The statistical significance of the anomalies between models and observations is calculated based on a two-tailed Student's t-test at the 5 % level.' I do not understand this very well. For example, in figure 1, the stippling mask indicates areas where four out of five models agree on the sign of the bias. In the same figure, however, non-significant anomalies at the 5% level are masked white. What do you mean by 'anomalies' here? Are you referring to the bias itself? How did you consider the degrees of freedom? How many independent observations do you consider here? Please, specify.

L222. 'the bias is especially persistent': do you mean persistent across the models? Isn't the term 'persistent' related to time?

L240. In order of appearance, I would change the figure numbering and name Fig. 2 to current Fig. 4. Similarly, Fig. 5 would be Fig. 3.

L308. I guess that it is Fig. S6 instead of S8.

L317-318. Is the dry bias at SH mid-latitudes associated with the warm bias in the SO? I see the dry bias for each LR and HR model (Fig. S4) and even in the AMIP runs (Fig. S3).

L323. What do you mean by: 'an improvement of the cold bias in the Weddell Sea'?

L330-333. A table as table 2 with RMSD and mean bias computed regionally in the SO would help in this kind of statement.

L333-337. Couldn't it also be the ocean model resolution or physics?

L365. 'All models but MPI-ESM1-2', not sure if also EC-Earth3P and CNRM-CM6-1. Again, here would be helpful to have a table with the regional metrics.

L373. 'In all the HR models, the cold bias over the subpolar North Atlantic is replaced by a warm bias'? I don't see it. Are you referring to the warm bias west of Greenland? If it is the case, it is not valid for MPI-ESM1-2.

L379. 'the cloud cover bias remains relatively unchanged' isn't easy to quantify from the figures.

L398. 'tropics' instead of 'tropis'.

L416. How do you see that 'the Gulf Stream separation improves'? In HH, it appears a cold bias close to the coast.

L432. In general, I would prefer the terms' eddy-parametrised/permitting/resolving' for the ocean resolution. But you are using 'eddy-parametrised/present/rich', so eddy-permitting here is not consistent. Besides, I think that you are referring to 'eddy-rich' instead of eddy-present' here.

L455. I guess that is LL instead of LR.

Figures and tables:

Table 2. Adding the global *tos* bias in the coupled runs could help determine the possible added value of increased ocean resolution.

Figure 6. The stippling mask is missing in this figure.