

## Response to reviewers

For easier reading, we repeat the reviewer's comments in black and answer then in *blue*.

### Reviewer 2

The manuscript provides a first evaluation of the Last Glacial Maximum (LGM) simulations performed for PMIP4-CMIP6 and a comparison with PMIP3-CMIP5. The authors focus on evaluating changes in temperature, the hydrological cycle, as well as atmospheric and oceanic circulation between PI and LGM simulations. The manuscript also provides an overview over how simulation results changed between these two rounds of PMIP. For their model-data comparisons they use new proxy data compilations as compared to the previous round of PMIP. This manuscript will provide a useful reference for a wide audience in the field. I did not find any major shortcomings with the manuscript, and I recommend this manuscript for publication after a few minor and technical comments and suggestions as detailed below have been addressed by the authors.

*We thank the reviewer for these supportive comments.*

### Main points

1. The model ensembles in PMIP3 and PMIP4 are very different, only two modelling groups participating in PMIP3 had submitted LGM simulations from their new model versions to PMIP4. This is not the fault of the authors, but certain conclusions cannot be drawn in this situation. E.g. I don't think that the statement (abstract, lines 6-8) that "PMIP4-CMIP6 are globally less cold and less dry than the PMIP3-CMIP5 simulations, most probably because of the use of a more realistic specification of the northern hemisphere ice sheets in the latest simulations" is justified. This could be tested by simulating the new and old ice sheet configuration using the same models, but certainly not by just averaging two very different model ensembles. Similarly, in lines 281-282 "This, together with colder temperatures, could help explain why the PMIP4-CMIP6 models simulate a stronger AMOC at the LGM." Here at least the wording is more appropriate ("could help explain" instead of "most probably"), but I still think that this could very well be due to completely different reasons (AMOC is sensitive to a lot of things in different models). In summary, I believe the comparison PMIP3-PMIP4 is of limited use when it comes to attributing causes and/or processes to different outcomes, and I think it would strengthen the manuscript if the authors would make this more explicit, e.g. by mentioning this in the introduction and/or the methods section, and revise the statements mentioned above (and further lines 267-269; lines 342-344; lines 382-384; lines 388-389).

*The results of two new models, IPSLCM5A2 and CESM1.2, for which previous versions existed in PMIP3. However, we agree that the models in the PMIP3 and PMIP4 sets are not simply different versions of the same models. We will modify the text to highlight this. This is also shown by Figure 1, which simply displays the difference in mean annual precipitation for PI, between the PMIP3 and PMIP4 ensembles.*

2. Given that the PMIP ensembles are relatively small anyway, it is a bit disturbing that not all PMIP4 models seem to have delivered all diagnostics, particularly since AMOC and northward heat transport are pretty standard, and all modelling groups are represented as co-authors on this manuscript as far as I can see. I would urge the authors to get the missing data in place for a revised version of a manuscript. Otherwise, in lines 227 and 240-241, I would not list the models that provide the data, but those who are missing (because they are fewer) and why.

*We will attempt to add the missing diagnostics in for the next version of the manuscript and will modify the text as suggested. We actually have collected a lot of the missing data.*

3. Outliers: The authors are classifying some model results as outliers, apparently without any objective criterion. Why is CNRM-CM5 an outlier regarding tropical temperature change (line 198,

Fig. 2)? The model is consistently warm over all latitude bands, and there are PMIP4 models that show almost the same weak cooling in the tropics. Likewise (lines 270-271, Fig. 7), why is CNRM-CM5 an outlier? I see that iLOVECLIM has a very strong reduction of precipitation in the tropics (is this due to the reduced complexity of the model compared to the others?). On the other hand, the zonal mean behaviour of INM-CM4 looks much stranger, so why isn't this model an outlier?

*We agree that we used the word "outlier" too loosely here. CNRM appears to stand out from the PMIP3 model results for the LGM tropical temperatures (Fig 2) but this is not seen on the zonal mean plot because many PMIP4 models mostly stand between CNRM and the other PMIP3 models. We will attempt to use the word more precisely in the next version of the manuscript, in particular stating which ensemble we consider when pointing to an "outlier".*

4. The description of changes in AMOC (lines 228-233) is a bit unclear: "Two of the PMIP4-CMIP6 models show a deep NADW cell reaching the ocean floor in the North Atlantic, whereas four of the PMIP4-CMIP6 models simulate some Antarctic Bottom Water (AABW) in the North Atlantic." The AWIESM1 and the MIROC-ES2L have a very deep cell already in the PI. The only model showing a strong change is iLOVECLIM. What do the authors mean by "some AABW". Likewise: "The intrusion of AABW cell into the North Atlantic was shown by some of the PMIP3-CMIP5 simulations, but not as much as the PMIP4-CMIP6 simulations". I have difficulties seeing this in Fig.4. In PMIP3, only CCSM4 shows an "intrusion of AABW" from PI to LGM, and in PMIP4 I do not see this at all(?). If there is a clear change of AABW intrusion between PMIP3/4 maybe an additional figure could help? Please consider revising the description of overturning changes.

*We agree that this description is not very precise and will update it in the next version of the manuscript. In particular we will add the models' names after the assertions and add the definitions of NADW and AABW on the plots. We will also include the notions of change in ocean circulation state, like for iLOVECLIM, to descriptions of PI and LGM states. We also plan to add an analysis of AMOC strength, AABW strength, and AMOC depth.*

Other points:

line 47-48: "...several of the PMIP4-CMIP6 models having substantially higher climate sensitivity than the PMIP3-CMIP5 versions of the same models, and thus the range of climate sensitivity sampled by the PMIP4-CMIP6 models is much wider." It should be mentioned already here, that this is not true for the models actually included in this study (as stated further below, lines 122-125).

*This is right and will be added in the text. We were actually hoping to include those models at a second stage, but this probably will not be the case.*

lines 159-160: I do not understand "...we only use the data synthesis for comparisons here." Is other data used for other purposes in this work? Please clarify.

*We meant that we only use the information at data sites, not the global result from the data assimilation. In the new version of the paper, we will actually show both, which raises interesting points about representativity of the sites.*

lines 187-188: "As expected, the simulations show larger changes over the land than over ocean." It would be interesting to read how much (i.e. maybe state the average land-sea gradient for PI and LGM?).

*Fine, we will add the quantification in the next version of the manuscript, by computing the regression between the two quantities.*

lines 358-359: "The PMIP4-CMIP6 models are more consistent with the temperature reconstructions over tropical Asia, but show poorer agreement with the precipitation reconstructions than the PMIP3-CMIP5 models." I would say "slightly more consistent". As to the precipitation reconstructions

over tropical Asia, it could be mentioned that PMIP4 models at least agree on the sign of change in contrast to PMIP3.

*It turns out that this conclusion was drawn from a very few points over this region, so we have removed it from the new version of Figure 12.*

Please complete Table 1 as far as possible (some references are missing, and it should be possible to collect information on PMIP3 spinup duration at least for MPI and MIROC).

*Table 1 is now much more complete, in particular for the PMIP4 models which are documented here altogether for the first time.*

Technical:

lines 12-13: "...remain large so,..." (?) -> maybe "...remain large, and although..."

*This has been corrected, thank you.*

line 28: delete "rise"

*OK, done.*

lines 34-35: "Atmospheric greenhouse gases (GHGs) were lower than pre-industrial (PI) values,..." -> "Atmospheric greenhouse gas (GHG) concentrations were lower than during the pre-industrial (PI) period,..."

*OK, done.*

line 83: alternate -> alternative (?)

*yes, this is what we meant.*

line 206: the authors probably mean "significant" not "real".

*We will reformulate the sentence.*

line 332: "Although" does not make sense, consider rewording the beginning of this sentence.

*This is right. The sentence will be modified, taking into account an updated method to quantify model-related variability.*