

Thanks to the authors' diligent and continuing efforts to address the reviewer's comments, I find that the manuscript is much improved since it was first submitted.

However, I still have recommendations, some related to potential clarification of new additions, and others pointing out the few but important sentences where the authors could make a more compelling case. In particular, I do not share the same view as the authors in their concluding remarks, as they are formulated now. I am detailing these recommendations below.

Thanks to the reviewer for their comments. We have revised the manuscript in response to the review except in a few cases where we disagree with the suggested changes.

General comments

1. I find that some of the vocabulary related to climate models used in this paper is not always adequate or precise.

- L10: '13 climate models that were included... are used to investigate...' → '13 climate model simulations...' or 'outputs'. Otherwise the formulation implies that the authors ran the simulations themselves.

Response: "simulations" was added after "climate models".

- L12: 'the model simulations are compared with existing proxy records...' → 'the simulated variables are compared...' or 'the variables of interest in simulations are compared'

Response: We do not think it is necessary to change the wording here. It is obvious that selected model variables are compared with proxy records and other model studies.

- L19: 'LGM precipitation anomalies are simulated differently between models' → 'models show different...'. Otherwise it could imply that the precipitation processes are represented differently in the models.

Response: The wording was changed to "Models simulate a range of LGM precipitation anomalies over the region".

- L15: 'models do not have a robust response' → 'model simulations do not show a robust response'

Response: The wording was changed as suggested.

- L181: 'Many modelling studies have focused on the LGM as this period is one of the main 'entry card' experiments for PMIP'. The relevancy of this statement is questionable, as it reverses cause and consequence. There are scientific reasons why modellers are interested in simulating the LGM climate (which is why it has been defined as an entry card in PMIP).

Response: This part of the sentence was deleted.

-L211: 'Evidence' is a slightly strong word for a simulated change. I would use a synonym (e.g. 'signs') to avoid confusion.

Response: (L147) "Evidence for change" was replaced with "Simulated changes".

2. Although the authors have clarified the knowledge gap, the way it is formulated is not very compelling.

- L8: 'remain uncertain': According to what: proxy studies?

Response: The uncertainty reflects disagreement in proxy studies as well as previous modelling studies. This sentence is an introductory one, with the details provided in the Introduction section of the paper. We think the existing wording should be sufficiently clear to the reader.

- L8-9: 'including the [list] of changes': Is there a way to clarify the scientific problem you are tackling? This formulation using a list introduces a disconnection between your variable of interest. On the contrary, the emphasis could be on how these variables are connected (why you want to examine them together in the same paper), which tell us something about the processes.

Response: The aim of this paper is to provide a preliminary overview of the climate of Australia as simulated in PMIP LGM simulations. The study does not aim to address one specific variable or aspect of the climate circulation, but rather summarise the extent of model to model agreement and the extent of agreement between models and proxy records. This will then hopefully stimulate further studies into particular aspects of the Australian region at the LGM (e.g. monsoon, ENSO, frontal systems, etc.), and will be of relevance for interpretation of proxy records. Focusing on several different variables and modes of variability is not uncommon for such overview studies, e.g. see Grose et al. (2020) study of Australian climate in CMIP5 and CMIP6 models.

Reference: Grose, M. R., Narsey, S., Delage, F. P., Dowdy, A. J., Bador, M., Boschat, G., ... & Power, S. (2020). Insights from CMIP6 for Australia's future climate. *Earth's Future*, 8(5), e2019EF001469.

- L60: 'Questions about the climate of Australia include [list]'. Same as before, this is not a very compelling way to phrase a scientific question. Perhaps a more efficient way of telling the reader this would be to do it later in the paragraph, by explicitly relating this statement with L65-66 which mentions processes.

Response: Please note that this sentence was added in response to a previous review. We think the current version is adequate to introduce the themes of the paper. Any further edits here appear to be personal stylistic preferences of the reviewer, rather than relating to the scientific content or clarity of the paper. We feel that further editing of this paragraph is scientifically unnecessary and may not improve the manuscript, but we will follow the Editor's guidance.

3. I remain unconvinced by the perspectives.

General response: While we appreciate the reviewer's suggestions, we feel that it is appropriate for us as the authors to determine our own conclusions and perspectives. We are happy to follow the Editor's guidance and make revisions if there are scientific errors or the text is unclear.

- L20: 'suggesting that caution is required when interpreting model output'. While true, this is also an obvious statement, so probably a wasted opportunity to teach something new to the reader. I find your new element of conclusion about different land-sea mask over Sahul to be much more insightful for modellers, and possibly worth a mention in the abstract.

Response: We wished to provide a warning to those using output from climate models to compare with proxy records, particularly when relying on a single model. We feel that this general point is still relevant so we would like to retain it here. Regarding the land-sea mask, this is a relatively technical conclusion so we feel it would be better addressed in the Conclusions (following explanation of the LGM land masks in the models) rather than presented out of context in the Abstract.

- L22: 'is required to determine the drivers... and to identify the most plausible set of LGM simulations'. This statement puts on the same level a process-understanding objective ('determine the drivers') and a second part implying in its formulation that the end goal is to exclude some models based on their performance to get some kind of realistic ensemble. I would argue that the whole point of a multimodel comparison study is not to elect Mr. Best Model Out There (for all models are wrong, although some may give better results than others depending on the variables we are looking at) or exclude outliers, but really to use the model evaluation and the evidenced biases to learn something about the processes and determine how the model representation could be improved. **Same remark for L792-793**

Response: There are many points of view on the goals of palaeoclimate modelling studies, including to evaluate and improve models as well as to better understand the climate of the past (as discussed in Kageyama et al. 2017 and numerous other papers). Given this, we attempted to address both sets of goals with our concluding statement. We have now reworded the sentence to the following to remove reference to the most plausible set of model simulations:

"Further analysis based on model evaluation and quantitative model-proxy comparison is required to better understand the drivers of LGM climate and atmospheric circulation changes in this region."

At line 792-793: We think the discussion in Section 4.4 is reasonable, as it includes both perspectives based on better understanding past climate via modelling studies as well as possibly selecting models with a better simulation of regional climate. This is consistent with the range of perspectives in the climate and palaeoclimate modelling communities and we do not think further revision of this section of the manuscript is needed.

- L780-781: I do not understand what the authors mean by this statement. Isn't quantitative model-data comparison a way of characterizing model biases (and thus the 'uncertainty')?

Response: We intended to state that given large uncertainty (including over the sign of change) in both models and proxies, it is not currently possible to make robust conclusions about whether it was wetter or drier over northern Australia at the LGM based on the proxy and model evidence.

This conclusion follows from the discussions in the paper - there is uncertainty in interpretation of hydroclimate proxy records due to offsetting changes in both precipitation and evaporation as well as CO₂ effects in vegetation proxies, so it is difficult to conclude that it was clearly wetter or drier in many regions. In addition, models simulate inconsistent precipitation changes, with some models simulating wetter and others simulating drier conditions over northern Australia.

We have attempted to rewrite this sentence to more clearly summarise our point.

"Given the large uncertainty (including over the sign of change) in both models and proxies, it is not possible to make robust conclusions about whether it was wetter or drier at a regional scale over Northern Australia at the LGM based on the available evidence. "

- L789-790: Please justify this statement, as it is the first mention of a potential use for future climate. Alternatively, this could be justified in the introduction when outlining the aim of the present study.

Response: A brief mention of the relevance of understanding past changes in the region for constraining uncertainty in future projections is now added in the Introduction with reference to two relevant papers (Grose et al., 2020 and Narsey et al., 2020). A more detailed discussion of the relevance of past modelling for future projections in this region is beyond the scope of the current study but will be addressed in future work.

4. The addition of the different LGM land-sea masks is indeed interesting to see. I have a few related recommendations.

- Figures 1, 4 and 9: I wonder if the use of an interpolated contour for the LGM land-sea mask makes the most sense. Possibly, a sharp delimitation (without interpolation) between the wet and dry grid cells would let the reader see more clearly the model mesh and resolution. However, it might deteriorate visibility.

Response: Thanks for the suggestion. However we have decided not to plot the land-sea mask in this way as it would be difficult to see the climate variables of interest in the same plots.

- For consistency, Figures 1, 4 and 9 should also plot the pre-industrial land-sea masks in thin lines, and not just the high-resolution modern topography.

Response: We do not think it is necessary to plot the pre-industrial land-sea masks as well as the LGM land masks. Numerous other papers describing PMIP LGM simulations use only modern coastlines or add the LGM coastlines but without including equivalent pre-industrial coastlines (e.g. Yan et al., 2018; Brown et al., 2020; Kageyama et al., 2021; Wang et al., 2023). We will follow the Editor's guidance on this matter.

- L360 'the thin black lines indicate modern coastlines': Same, we cannot assume that the pre-industrial land-sea mask from individual model will follow well the high-resolution modern coastlines.

Response: We do agree that the piControl land-sea mask configurations may differ across models and not exactly follow the high-resolution modern coastlines. However, we do not think this is a major consideration compared with the dramatic changes in LGM coastlines. We do not think it is necessary to include the pre-industrial land masks on the figures, as discussed in the response to the previous point.

- L544-546: There is no mention of the sharp precipitation gradient that some models simulate at their coastlines (but not others). This could be worth mentioning.

Response: One sentence has been added in Section 3.3 to mention this finding.

- L755-756: As evidenced by the different land-sea mask, the model response may also be related to different boundary conditions (and not just a different model response to the change in boundary conditions).

Response: Thanks for that. One sentence has been added in Section 4.4 to address it.

Specific comments

L52-53: 'of lower temperatures... that cooled the climate'. This is more or less a repetition of the same element.

Response: "Cooled the climate" was deleted.

L58: 'While many studies'. Proxy studies or modelling studies? Please specify.

Response: "modelling" added before "studies".

L58: 'globally and in the Northern Hemisphere'. Is this an 'and/or'? This sounds a bit contradictory.

Response: We are highlighting the lack of studies which focus on regional changes in the Southern Hemisphere, e.g. over Australia or other SH continents. This is not contradictory to other studies focusing on global changes e.g. Kageyama et al. 2021 and lots of NH regional studies.

L73: 'have begun to explore' → 'have explored'

Response: Changed as suggested.

L75: A link word and a recapitulation of the originality of the study would be welcomed to contrast this study with previous ones. Something along the lines of: 'Hence, the present study used the most recent PMIP3 and PMIP4 simulations to investigate climate changes over the Australian region specifically'.

Response: Edited as suggested.

L137: Mentioning the Gray et al., 2023 paper (and contrasting its conclusions to Kohfeld's) would be welcomed here.

Response: This has been added as suggested.

L193-194 and L205-207: It is a bit confusing for the reader to switch abruptly (without link words) from the description of previous studies to the aims of the present study.

Response: The sentences were refined to improve consistency.

L199-204: Mentioning the changes in simulated SH westerly winds in the Gray et al., 2023 paper would also be welcomed here, as it is the most recent study.

Response: This has been added as suggested in Section 1.2.

Table 1 legend: The sentence related to length of simulation should be deleted (as the column was also deleted).

Response: Deleted.

L252: Please delete 'minor'.

Response: Deleted.

L257: 'PMIP3 ice-sheet configurations' → 'the PMIP3 ice-sheet configuration'

Response: Changed.

L262: 'with CMIP5 CNRM-CM5 having the smallest expander land area'. It is difficult to see why this statement is relevant here. Possibly, quantifying the range of the difference of surface area between PI and LGM (in km²), from the model with the smallest difference to the one with the largest would be more meaningful.

Response: The statement was deleted.

L296 'According to the PMIP protocols' and L297 'see Kageyama et al., 2017 for details of the spin-up protocol'. Why are you mentioning the recommendations (which are not always thoroughly followed) instead of the spin-up duration that was actually done for PMIP4 simulations (in Kageyama et al., 2021)?

Response: It is relevant to mention the PMIP4 protocol here. We now add a reference to Kageyama et al. (2021) for the actual spin-up durations as suggested.

L246: 'Otherwise' → 'Hence' or synonym

Response: Changed to "Overall" as this more clearly indicates the lack of an overall correlation between global and regional temperature anomalies despite some matches.