Responses to comments of Referee #1

The study by Gao et al. performed a 3000-yr 128-ka simulation to highlight the importance of the meltwater release into the North Atlantic in reducing the model–data discrepancies over the Southern Ocean. The manuscript is clearly written, and the topic fits the CP. However, there are several concerns that should be addressed.

Response: We appreciate the time of Referee #1 in reviewing this manuscript, thank you. These are useful discussions. Please check our detailed responses below.

Major comments:

1. In section 2.2, the authors declared that they selected the most recent surface temperature data syntheses and SIC dataset. I am wondering the reason for ignoring the data from Turney et al. (2020), as they also provide annual SST data during the early LIG. If there is no special reason, comparisons between the SST data from Turney et al. (2020) and models should be added.

[Reference]

Turney, C. S. M., et al. (2020). A global mean sea-surface temperature dataset for the Last Interglacial (129–116 kyr) and contribution of thermal expansion to sea-level change. Earth System Science Data 12(4): 3341-3356.

Response: We thank the reviewer 1 for their suggestion and we will explain below why we think that we would favour not presenting a model-data comparison exercise at 127 ka that include the Turney et al. (2020) dataset

The major issue of the Turney et al. (2020) is related to the fact that the authors present a data compilation based on paleoclimatic records kept on the original chronologies. They do not attempt to build a coherent temporal framework between the different paleoclimatic records nor to provide an assessment of the large chronology errors that are associated with marine sediment records across the LIG. This is hugely problematic considering that using different dating strategies for paleorecords across the LIG could lead to age discrepancies of up to 6 ka, as detailed in Govin et al. (2015). Also, it goes against the large efforts put together over the past years to guide the community towards being careful with harmonizing age models for paleoclimatic archives during the Last Interglacial. Indeed, it is widely recognized now how key it is to harmonize paleorecord chronologies when building data compilations in order to provide (1) a realistic representation of the LIG climate and (2) appropriate benchmarks to evaluate the LIG model simulations (e.g. Capron et al. 2014, Govin et al. 2015, Stone et al. 2016, Hoffman et al. 2017, Capron et al. 2017, Otto-Bliesner et al. 2017, Menviel et al. 2019).

In other words, we would like to strengthen the fact for the purpose of our work there is no added value of the Turney et al. (2020) dataset relative to the existing global SST compilation proposed by Hoffman et al. (2017) based on consistent timescales. In addition, the Turney et al. (2020) synthesis was not intended to be a time slice. We think showing a model-data comparison for the Turney et al. (2020) syntheses would be a step-back for the data communities working on improving the spatio-temporal representation of the LIG climate as it ignores the current understanding of chronological uncertainties for the LIG and it would lead to misinterpretation and misuse of this synthesis by other scientists part of the model community who might not necessary familiar with chronology-related subtleties, when performing model-data comparisons.

Finally, we would like to refer to Capron et al. (2017) and Otto-Bliesner et al. (2021) for further discussion on why peak-warmth climate synthesis over the LIG should be avoided for the model-data comparisons taking place in the framework of the PMIP4 lig simulations. At the time, we discussed the Turney and Jones 2010 and McKay et al. 2010 dataset and our points are still valid and apply to the peak-warmth centered LIG dataset from Turney et al. (2020). We are now clearer on the limitations of the peak-warmth-centered data syntheses in the revised manuscript.

We added the following sentence to line 35:

"Subsequently, Turney et al. (2020) compiled the maximum annual SST between 129-124 ka from 189 marine sediment and coral records. However, they still used the original age models and assumed global synchronous peak warming conditions, which, as mentioned, limit its applicability for the evaluation of equilibrium model simulations at a specific date across the LIG."

For completeness, we gathered the maximum annual SST estimates during the early LIG (129-124ka) from Turney et al. (2020) as in following Fig 1. There are 28 records south of 40 degrees south. There are more records in the Turney dataset compared to the ones we used as less strict criteria were applied. The anomalies are relative to HadISST 1981-2010 and the average anomalies are 2.5°C. The RMSE between LIG climate anomalies in the Turney dataset and the Null Scenario, HadCM3_127k, and HadCM3_128k_H11 are 3.8°C, 3.6°C, and 3.0°C, respectively (though the simulated anomalies are relative to preindustrial values). Since HadCM3_127k performs similarly as the Null Scenario and HadCM3_128k_H11 has the smallest RMSE, it complies with the results obtained when performing the model-data comparison with the other data syntheses. However, as mentioned above, the use of the Turney et al. (2020) synthesis is associated with large problems with chronology and prevents us from drawing any robust conclusions.

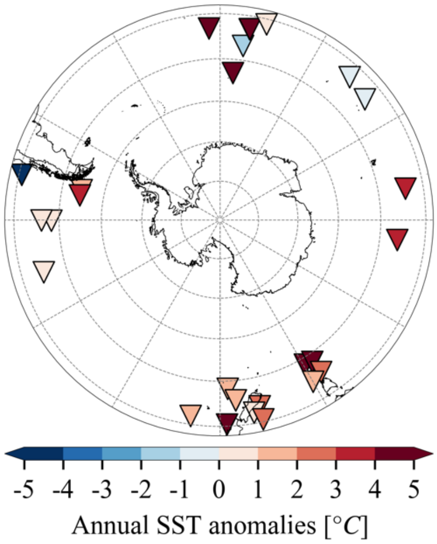

Figure 1: Reconstructed maximum annual SST anomalies during 129-124 ka relative to 1981-2010 over the Southern Ocean from Turney et al. (2020).

2. In section 2.3, the experimental setups are not clear. The boundary conditions and external forcings (e.g. greenhouse gases, orbital parameters, ice sheet, vegetation, land-sea mask) at 128 ka need to be displayed or indicated directly.

Response: We added in the revised manuscript the following description to section 2.3, Line 159:

The greenhouse gas concentrations in this simulation are close to those set by the PMIP4 lig127k guideline: carbon dioxide at 275 parts per million (ppm), methane at 706.8 parts per billion (ppb), and nitrous oxide at 266 ppb. The vegetation, aerosol, and ice sheets were set identical to the corresponding preindustrial simulation (Tindall et al., 2009).

3. In section 3.1, the authors evaluated the performance of PMIP4 models in simulating SST over the Southern Ocean, but they only provide the RMSE. The spatial correlation coefficients between PMIP4 models and HadISST1 dataset should also be provided. I suggest providing a Taylor diagram to show the performance of PMIP4 models more clearly.

Response: Thank you for the suggestion on a Taylor diagram. It does provide more information on correlation coefficients and standard deviations as shown in the following figure 2 for annual mean SST south of 40 degree south, but we chose to display area-weighted RMSE as it is for three reasons:

1) Compared to the RMSE in the Taylor diagram, we consider our area-weighted RMSE to be more robust, as it takes into account meridional variations in the area of grid cells.

2) As meridional gradients in annual mean SST are well captured by all models, all correlation coefficients are larger than 0.93. This means the Taylor diagram provides little insight into any model bias.

3) Except for MIROC-ES2L, all other models demonstrate similar RMSE and standard deviation. Since this can be expected from the large warm bias shown in Fig 2j, additional presentation of this similar information may be redundant.

However, if the referee and/or the editor still would like us to include it in the revised manuscript, we can do so.

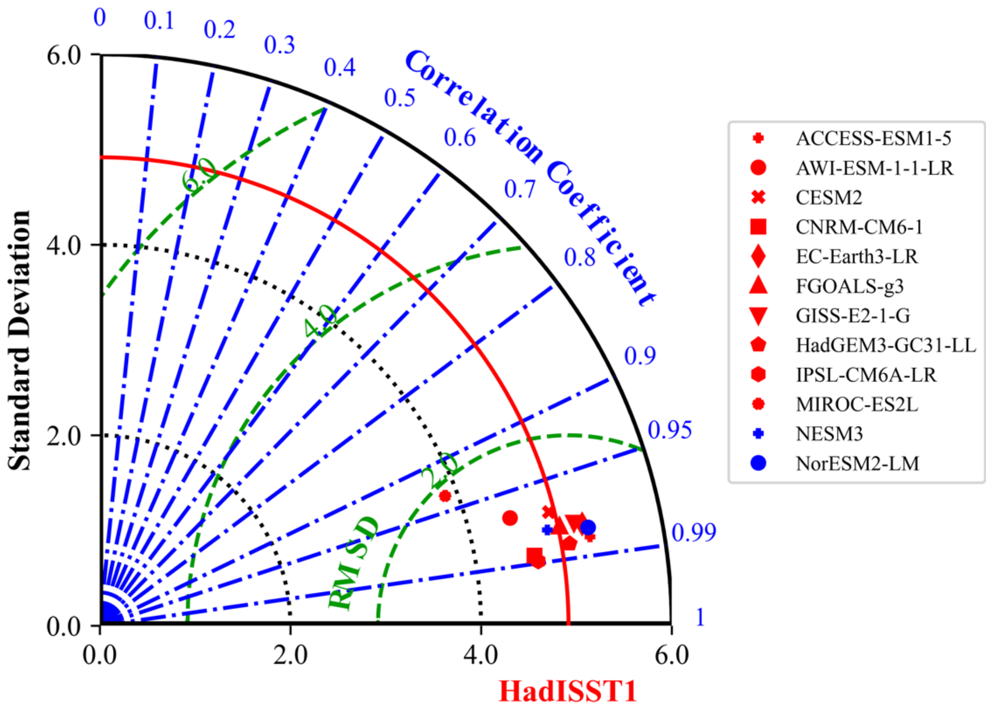

Figure 2: A Taylor diagram showing standard deviation, root mean squared errors, and correlation coefficients between annual mean SST south of 40 degree south from 12 models and the HadISST1 dataset.

4. Across the manuscript, the authors suggested that the orbital parameters, greenhouse gases and Antarctic ice sheet played a limited role in the warming of the Southern Ocean during the early LIG, and attributed the warming to the meltwater release into the North Atlantic. Zhang et al. (2023) indicated that the global sea level rising during the LIG (at 126 ka) can also reduce the model-data discrepancies over the Southern Ocean. I suggest that more discussions about this point need to be added.

[Reference]

Zhang, Z., et al. (2023). Atmospheric and oceanic circulation altered by global mean sea-level rise. Nature Geoscience 16(4): 321-327.

Response: We are grateful for being pointed to this nice article. The following sentence is added into the Introduction at Line 51.

"While Zhang et al. (2023) found that increased global mean sea level warms southern mid-to-high latitudes at 126 ka using a climate model NorESM1-F, the root mean squared errors (RMSE) between temperature anomalies at 126 ka relative to preindustrial from the simulations and the Chandler and Langebroek (2021a) dataset were only reduced by ~10% while applying a 5-m or 10-m sea level rise."

5. The authors indicated that long-time simulation (i.e. the 3000-yr simulation) of H11 is likely required to capture the full magnitude of the Southern Ocean and Antarctic warming, following the guideline of the modeled linear SST trend in a 1600-yr simulation by Holloway et al. (2018). In Figure A1, the linear trend of the SST from 0 to 1600 years is significant. However, from 1600 to 3000 years, the SST seems to fluctuate near a mean state, rather than shows a long-term increasing trend. So I am wondering the necessity of running a long-time simulation in reconciling the model-data mismatch based on the current results. I suggest providing the difference between the "short" simulation and "long" simulation at 128 ka to highlight the novelty of long-time simulation performed in this study.

Response: We agree that the increasing rate of Southern Ocean SST slows down after ~1600 model years. This indeed contrasts the expectation of Holloway et al. (2018) who assumed a linear trend in temperature changes and proposed that 3-to-4 thousand years simulation may reconcile the model-data mismatch in ice core records. Our preliminary analysis using a 2000-year simulation indeed provides similar results as the 3000-year simulation. Though, we still consider it valuable to reach a quasi-equilibrium state, which facilitates the investigation of post-hosing abrupt climate changes. This is actually a main research topic in our group now. We added the following sentences in Line 312:

"In addition, we note that the rate of Southern Ocean SST increase slows down after ~1600 model years (Fig. A1), and our preliminary analysis using a 2000-year HadCM3_128k_H11 simulation gives similar results as the 3000-year one. This contrasts with the expectation of Holloway et al. (2018) that the Southern Ocean and Antarctica would exhibit a linear warming trend throughout the period of meltwater input. However, we consider it valuable to run our model long enough to reach a quasi-equilibrium state to test their hypothesis. It also facilitates the investigation of post-hosing abrupt climate changes."

Minor comments:

Line 87: Four most recent surface temperature data syntheses, may be three?

Response: It is four since Chadwick et al. (2021) also provided temperature reconstructions.