

We appreciate it that you help us to improve our paper by giving your suggestions. We list our point-to-point replies as below.

The original review comments are shown in blue, and our responses in **black**.

This paper takes a careful, nuanced analysis of the potential biases in PlioMIP2 models' representation of mean climate state and whether this is related to bias in their simulation of Pliocene climate. This is important, since we normally justify the use of certain models to simulate past climates in a given region based on their ability to reproduce the pre-industrial climatology. However, if this is not the case, then we must be more careful of our choice of models. The results presented required a huge amount of data compilation and analysis, and it is a nice demonstration of the utility of the multi-cluster mean approach to understanding model disagreement and consensus.

Thank you for your kind words.

- ✓ Figure 3: make the markers larger, more bold, it's a little hard to see all of them directly. Would suggest using different colors rather than distinct symbols, or increase line weight or something. Make sure each panel is labeled (a) and (b) do not appear on the figure

A: As there are some overlapping, we agree that it does take some time to distinguish every marker. As such, we will make adjustments to this figure to make it clearer. Regarding to the panel label, we will check every figure and add labels to sub-figures.

- ✓ Figure 4: is it possible to propagate through the uncertainty (e.g. full error envelope of Pliocene proxy values) into the EOI400 discrepancy calculations? Or at least add some 95% error bounds.

A: Thank you for this useful comment - we agree that it is important to consider the uncertainties in the discrepancy (shown in Figure 3) when interpreting Figure 4. As such, we will add this to the figure caption and/or discuss in the text.

- ✓ Figure 5, and generally all figures with continental outlines: make the outlines of the land bold.

A: We will adjust the outlines of these figures.

- ✓ Figure 6 is very interesting. The relationship between ECS and SSTA over the maritime continent - I wonder if this can be formally connected to the 'pattern effect' literature. See for instance: Dong, Y., Armour, K.C., Zelinka, M.D., Proistosescu, C., Battisti, D.S., Zhou, C. and Andrews, T., 2020. Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate*, 33(18), pp.7755-7775.

This literature points to the fact that long-term changes in climate feedbacks seems to depend on the relative warming in the western Pacific warm pool region.

A: Thank you - we will add appropriate discussion about the pattern effect in our discussion section 4.2.

- ✓ While the text does a good job of articulating the differences between the MMM and the MCM, in the plots themselves they look quite similar for SSTA and precipitation. This might be a feature of the small size of the plots. Can the individual panels be made larger, and the dendrogram made much smaller, so that it is easier to compare and contrast the map panels?

A: We will adjust the size of each figure.

- ✓ Line 475: I would say more about the potential sources of proxy uncertainty for ODP 214. As of now the analysis of this potential source of model disagreement is quite cursory.

A: We will expand the discussion in section 3.1.2, Line 202-209, of the underestimate of temperature reconstruction for the site ODP 214 which is Mg/Ca-derived and not be supported by other proxies such as alkenones.