Final response to reviewer 1

Regarding the main comment relating to the choice of 30ka interval for the PRISM4 data set, we now additionally mention results generated from the tighter 10ka PRISM4 interval. However, we also modified our wording to clarify that these values are the full widths of the two PRISM4 intervals, ie they are defined by the ranges 3.205+-0.015Ma and 3.205+-0.005Ma respectively, and there is no PRISM4 data set using an interval of 3.205+-0.01Ma that would be directly equivalent to that of McClymont et al. In any case, the two versions of the PRISM4 sets are very similar and generate near-identical results.

Individual notes:

Page 4, line 2:

This appears to be a simple misunderstanding, but Bragg (2014) could not have used PRISM4 data since those data were not available prior to 2016 and SST estimates shown in Foley and Dowsett (referred to here as PRISM4), were not produced until 2019.

The reference made to the data set used by Bragg has been corrected.

Page 4, lines 18-19:

Why use the PRISM4 community sourced verification data with a 30K window to compare to PLIOVAR's 20K window when PRISM also, in the same release, produced a version with a ±10K window?

Page 4, lines 28-29:

The PlioVAR interval is slightly narrower only due to your choice of the 30 kyr window rather than the identical 20 kyr window provided in Foley and Dowsett (2019).

As above, neither of the PRISM4 intervals coincides with that of PlioVAR. We have tested both versions of PRISM4 and report this in the paper.

Page 5, lines 26-27:

You should probably cite a couple of the many available references that previously documented differences between Mg/Ca and alkenone based SST estimates in Pliocene and Pleistocene sequences.

Additional comments and references added

Page 5, lines 30-34:

This is an interesting point and it would be helpful if it was addressed in this paper. Foley and Dowsett (2019) is a compilation of previously published alkenone data and it would be useful to know whether the sites not in common with PlioVAR are from a particular region, particular lab, etc.

Page 7, lines 33-35:

As in one of the comments above, comparing alkenone and Mg/Ca based SST estimates is like apples and oranges. They are measuring different things and while both are calibrated to mean annual SST, the literature is ripe with examples of the two providing discordant estimates. On page 8 of this manuscript you indicate some possible reasons for Mg/Ca data being less reliable, the same reasons that have been stated by many authors in the past. Maybe move those up to page 7 and provide citations to earlier works?

Comments moved as requested

Page 9, line 17:

I think you just made a simple typo with the citations. If you are referring to the PRISM4 compilation you must mean Haywood et al. 2020 (not 2010).

This has been corrected.

Thank you also for the various corrections to references including DOIs.

Final response to reviewer 2.

Where we have disagreed with your suggestions in our previous response, we have not changed the paper.

In regards to your point about cutoff radius, we clarify that the cutoff radius is actually double the length parameter that we've quoted. we now make specific reference to the tests we performed with a greater length parameter of 10,000km which demonstrate that this is not a major factor in our results regardless of recentering the prior.

We have also emphasised more strongly that the prior is inevitably a

significant factor in the results, and added a reference to a newly published LGM/deglaciation reconstruction which further demonstrates this point. Reconstructions based on a single model ensemble cannot reasonably be considered robust.

We have explained that leave one out cross validation isn't a meaningful exercise in this situation due to the extremely low number of data points. However we also observe that the tests of various parameters in the reconstruction process have relatively little influence especially in contrast to the choice of data set and recentering (or not) the model prior.

We do include the global surface air temperature anomaly for each model, in Table 1.