Response to Reviewer #2 (Inglis et al. Climate of the Past)

Line numbers refer to the "Track Changes" document
Black: Reviewer comments
Blue: Author response

The manuscript is clearly written. Its structure is logical.

Thank you for the positive feedback!

Line 165: The authors calculate the annual average surface temperature field and the uncertainty in the reanalysis product ERA-5 with the past distribution of geographic samples. It is not clear how the authors proceeded. Does it mean that the closest grid point corresponding to the past position of a site is selected? at the same elevation?

The closest grid point corresponding to the past position of a site is selected. We have now provided all source code and data to reproduce the analysis and the code itself makes it very apparent what assumptions and decisions have been made.

Line 177 and Lines 191-193: The global mean temperature changes from one climate model to the other. Thus, the authors should test other models (available through DeepMIP project)?

In the original manuscript, we utilised Community Earth System Model version 1.0 (8x and 16x CO_2). However, recent work has shown that Community Earth System Model version 1.2 offers a major improvement over earlier models (e.g. better representation of the meridional SST gradient; Zhu et al., 2019; Science Advances). As such, we have performed an additional analysis using CESM1.2 (6x CO_2)

Both CESM1 and CESM1.2 yield similar GMST estimates during the PETM, EECO and latest Paleocene (see table below). This indicates that the final result is not sensitive to the choice of reference simulation, at least within the CESM model family **(see lines 217-230)**

		GMST					
Experiment	**Model simulation**	**EECO**	**SD**	**LP**	**SD**	**PETM**	**SD**
D_{surf}-Default	CESM1 (8x CO2)	24.5	0.8	26.9	1.3	33.9	1.4
D_{surf}-Default	CESM1 (16x CO2)	24.6	0.8	26.4	1.3	33.8	1.4
D_{surf}-Default	CESM1.2 (6x CO2)	25.2	0.9	25.0	1.2	31.8	1.2

Note that we only employ CESM1 simulations in our 'combined' GMST estimates to avoid circularity if the results from this paper are used to evaluate more recent simulations (e.g. CESM1.2; Lunt et al., 2020).

Lines 205-206: Two assumptions are considered: "global temperatures scale linearly with local temperatures, and a climate model can represent this scaling correctly". These assumptions need to be tested. In addition, the two pairs of simulations have been obtained with two different climate models (and different boundary conditions). The influence of the type of model and the boundary conditions should be investigated (a table indicating the model and the boundary conditions used should be added).

In the original manuscript, we calculated transfer functions using **two** climate model simulations: 1) HadCM3L (2x and 4x CO_2) and 2) CESM1 (4x and 8x CO_2). We have now performed the same analysis using two additional model simulations (CESM1.2 and GFDL) at two different CO_2 levels (x3 and x6 CO_2). Both simulations were carried out within the DeepMIP framework (www.deepmip.org).

We find that all four simulations (i.e. HadCM3L, CCSM3, CESM1.2 and GFDL) yield similar GMST estimates. This demonstrates that D_{surf}-2 is not overly sensitive to the climate model simulation **(see lines 262 to 273)**. However, we only employ CESM1 and HadCM3L simulations in our 'combined' GMST estimates to avoid circularity if the results from this paper are used to evaluate more recent simulations (e.g. CESM1.2; GFDL; Lunt et al., 2020).

			GMST					
Experiment	**Model simulation**	**CO_2**	**EECO**	**SD**	**LP**	**SD**	**PETM**	**SD**
D_{surf}-Default	CESM1	4x, 8x	25.86	8.96	26.10	5.81	32.26	6.66
D_{surf}-Default	HadCM3L	2x, 4x	27.51	8.88	27.56	8.05	34.49	13.95
D_{surf}-Default	CESM1.2	3x, 6x	25.82	9.70	27.05	5.68	32.70	6.58
D_{surf}-Defaull	GFDL	3x, 6x	26.21	8.73	27.32	6.39	33.15	6.75

To explore whether **GMST scales linearly with local temperatures**, we calculated GMST using CESM1.2 but with a different factor (3x to 9x CO_2, instead of 3x to 6x CO_2). The results are very similar **(±0.3°C; see below).** This is because, although the relationship between GMST and CO_2 is non-linear (Caballero and Huber, 2013; Zhu et al, 2019), the relationship between local and global temperature is relatively constant. **(L272-278)**

			GMST					
Experiment	**CO2 levels**	**Model simulation**	**EECO**	**SD**	**LP**	**SD**	**PETM**	**SD**
D_{surf}-Default	3x, 6x	CESM1.2	25.82	9.70	27.05	27.05	32.70	6.58
D_{surf}-Default	3x, 9x	CESM1.2	26.23	8.90	27.09	5.87	32.79	6.54

We also include a table in the supplementary information (**Table S1**) with details on different model simulations used.

Lines 218-220: How many proxy temperatures are greater than Thigh or Tlow? How many global mean temperatures are thus obtained by extrapolation?

The number of GMST estimates obtained via interpolation vs. extrapolation will be sensitive to the choice of model simulation; models that simulate less polar amplification (e.g. HadCM3L) are more likely to obtain $<T>^{inferred}$ (i.e. GMST) via extrapolation. This discussion has been added to the text (**L257-259, 264-266**).

Lines 339-344: For DComb-1, how to be sure that the equation 5 can be used in case of warm climates?

We agree that it's important to test these assumptions in hothouse climates.

To test these assumptions, we modelled the shape of the latitudinal temperature gradient using a simple algebraic function (Figure S5). We find that D_{comb}-1 may underestimate GMST by 0.75 to 1.25 °C. We also used CESM1 simulations (EO3 and EO4 from Cramwinckel et al., 2018) to compare the "true" model simulation GMST to that calculated using D_{comb}-1 (Supplementary Information). We find that D_{comb}-1 underestimates GMST by 1°C when the model high latitude SST is used a proxy for the deep-ocean, and 2-3°C when the model deep ocean temperature is used.

As such, D_{comb}-1 may reflect a minimum GMST constraint during past warm climates. We now acknowledge these caveats in the text (**L424-430, 517-527**).

Lines 356-382: GMST should be estimated using other climate models to explore model dependency.

Note that only two methods incorporate model simulations (D_{surf}-1 and D_{surf}-2).

D_{surf}-1 originally employed a single GCM (CESM1) to characterise how well the existing palaeographic sampling network will impact GMST estimates. We expand this to include an additional GCM (CESM1.2 **L217-230**) which has undergone a nearly complete overhaul of physical parameterizations in the atmosphere model (Zhu et al., 2019; Lunt et al., 2020).

D_{surf}-2 originally employed two GCMs to calculate GMST (HadCM3L & CESM1). We expand this to include two additional simulations from the DeepMIP ensemble (GFDL & CESM1.2) (**L262-278**).

Lines 386-387: The influence of proxy datasets is shown for EECO only?

We have subsequently moved this figure (Figure 6) to the Supplementary Information. The supplementary information includes the LP and PETM equivalents for consistency.

Line424: The authors should explain why the land air proxy data can suffer from a cold bias.

Several of these proxies saturate between ~25 and 29 °C (e.g. leaf fossils, pollen assemblages and brGDGTs; see Hollis et al., 2019 and ref. therein) and/or are impacted by non-temperature controls (e.g. paleosol climofunctions; see below). As such, this could skew GMST estimates towards lower values.

To confirm this, we calculated GMST values using LAT proxies only (Supplementary Information) and show that GMST values are up to 6°C lower than our 'baseline' (SST + LAT) calculations. This discussion has been added to the text (**L480-492**)

Line 430: The authors should explain why the inclusion of ïA¿d'18O values from paleosols or mammals leads to a cold bias.

Hollis et al (2019) (who compiled the SST + LAT dataset employed in this paper) state that "…paleosol or mammal $\delta^{18}O$ are anomalously cold at several sites, notably, Salta Basin, Argentina (Hyland et al., 2017); Wind River Basin, Wyoming, USA (Hyland et al., 2013) and Ellesmere Island, Canada (Fricke and Wing, 2004)".

We suggest this could be because paleosol and/or mammal $\delta^{18}O$ values are impacted by other controls (e.g. variations in the isotope composition of rainfall and soil water (e.g. Hyland and Sheldon, 2013; Dworkin et al., 20015)). Paleosol $\delta^{18}O$ values also have issues with error estimation due to $\delta^{18}O$ heterogeneity within nodules (e.g. Dworkin et al. 2005). Such uncertainties could lead to unreliable temperature estimates.

Temperature estimates from paleosol climofunctions may also be prone to underestimation (e.g. Sheldon et al., 2009) and Hyland and Sheldon (2013) suggest that paleosol climofunctions are only applied as an indicator of relative temperature change. This discussion has been added to the text (**L565-572**)

Lines 438-458: Curiously GMST estimates using Ddeep and Dcomb did not yield a similar cold bias.

GMST estimates derived from D_{deep} and D_{comb} do not utilise LAT estimates (c.f. D_{surf}-1 to -4). As such, it is unsurprising that these methods fail to yield a similar cold bias.

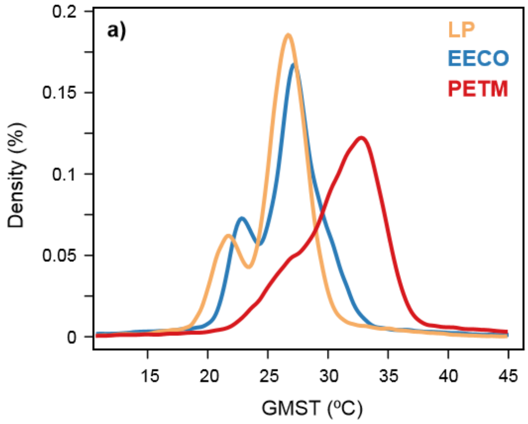
Line 470: how the uncertainties on the best GSMT can be so small.

The original method employed a weighted average to estimate GMST and the uncertainty was calculated using the reciprocal square root of the sum of all the individual weights. This led to unrealistically low uncertainty estimates.

We now employ a probabilistic approach, using Monte Carlo resampling with full propagation of errors (**L589-605**), to combine GMST and quantify uncertainty.

Specifically, we generate 10,000 iterations for each of the six methods for the LP, PETM and EECO. In these iterations, the GMST estimates were randomly sampled with replacement within their full uncertainty envelopes, assuming Gaussian distribution of errors. As the different GMST estimates ultimately derive from the same proxy dataset, we do not consider them to be independent. The resulting 60,000 GMST iterations for each time period are thus simply added into a single probability density function, in order to fully represent uncertainty (**L589-605; see below**). From this this probability distribution, the median value and the upper and lower limits corresponding to 66 and 90% confidence limits were identified.

Our new results indicate that the average GMST estimate (66% confidence) during the latest Paleocene, PETM and EECO was 26.3°C (22.3 to 28.3°C), 31.6°C (27.2 to 34.5°C) and 27.0°C (23.2 to 29.7°C), respectively (summarised below).



Line 75: Figure 1 and Table 1

Amended accordingly.

Figure 2a: a site located to the north of South America is unnamed

Amended accordingly. Also note that Figure 2 has been moved to the Supplementary Information.

Line 89: ECS is used before being defined (line 483)

Amended accordingly.

Line108: define GDGT (ie glycerol dialkyl glycerol tetraethers)

Amended accordingly.

Line122: replace Table 1 by Table 2

Amended accordingly.

Line 127: define MBT(')/CBT

Amended accordingly.

Line 385: subsampling case must be explicitly indicated in the caption of figure 6

Amended accordingly.