Response to Reviewer #1 (Inglis et al. Climate of the Past)

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

This study is a step forward in the estimation of the early Eocene GMSTs in its usage of the latest temperature compilations, its exploration of multiple methods, and its quantification of uncertainty from potential biases in proxies.

We thank the reviewer for their kind comments!

1. For multiple occasions, corrections or inferences are based on a single climate model. For example, the correction offset in Dsurf-1 and the inference in Dsurf-2. I suggest that the authors explore potential difference in GMST estimates if other models are used, as they are available in the DeepMIP archive (Lunt et al., 2020).

These are all very useful suggestions. We explore potential differences in GMST estimates as followed:

a) Dsurf-1

Dsurf-1 uses model results to characterise how well the existing palaeographic sampling network will impact GMST estimates. 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 temperature gradient; Zhu et al., 2019; Science Advances) due to the improved treatment of cloud microphysical processes. 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 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 216-229)**

		GMST					
Experiment	**Model simulation**	**EECO**	**SD**	**LP**	**SD**	**PETM**	**SD**
D_{surf}-Baseline	CESM1 (8x CO2)	24.5	0.8	26.9	1.3	33.9	1.4
D_{surf}-Baseline	CESM1 (16x CO2)	24.6	0.8	26.4	1.3	33.8	1.4
D_{surf}-Baseline	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).

b) Dsurf-2

D_{surf}-2 uses a transfer-function to calculate global mean temperature from local proxy temperatures. 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). However, 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 (see below).

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

This demonstrates that D_{surf}-2 is not overly sensitive to the climate model simulation **(see lines 262 to 273**).

Note that 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).

2. Assumptions of some methods should be better explored. For example, Dsurf2 assumes that GMST scales linearly with local temperature. Does this assumption hold in model simulations? This could be tested in DeepMIP simulations, as there are several modeling groups providing more than two simulations with different CO2 levels (Lunt et al., 2020)

To explore whether GMST scales linearly with local temperatures, we calculated GMST using CESM1.2 using a different scaling 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. (**L272-277)**

		GMST					
Experiment	**Model simulation**	**EECO**	**SE**	**LP**	**SE**	**PETM**	**SE**
D_{surf}-Baseline	CESM1.2 (3x to 6x CO2)	25.82	9.70	27.05	5.68	32.70	6.58
D_{surf}-Baseline	CESM1.2 (3x to 9x CO2)	26.23	8.90	27.09	5.87	32.79	6.54

3. Related to 2, the authors should also verify the assumptions made in Dcomb-1, i.e. GMST = 0.5 * (tropical SST + BWT). This is especially necessary when results from this method are consistently lower than other methods. That Dcomb-1 can estimate the modern GMST within an error of~1◦C does not guarantee its consistent performance for a hothouse climate. I suggest the authors test this method in model simulations. I understand that most of the current Eocene simulations are short in length and bottom water temperature has substantial trend, but there are longer runs that are worth exploring (e.g., GFDL runs in Hutchinson et al. (2018); HadCM3 runs; and CESM1 runs in Cramwinckel et al. (2018)).

These are useful suggestions!

We tested D_{comb}-1 by modelling the shape of the latitudinal temperature gradient using a simple algebraic function (**Figure S5**). This suggests 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 (**Figure S5**). We use these simulations because they have a "spun-up" deep ocean. We find that D_{comb}-1 may underestimate 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.

These results suggest that D_{comb}-1 may reflect a minimum GMST constraint during past warm climates. We now acknowledge these caveats in the text (**L423-429, 516-527**).

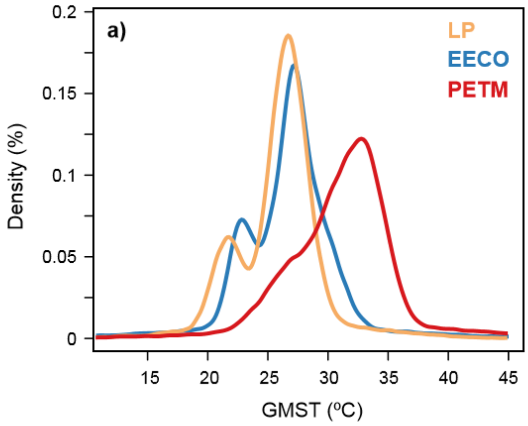
Also, Dcomb-1 is incompatible with Ddeep-1. Dcomb-1 assumes ∆GMST = 0.5 * (∆tropical SST + ∆BWT), while Ddeep-1 assumes ∆GMST = ∆BWT. It is better to keep only one method that has smaller biases.

In this paper, we aim to put forward multiple approaches to estimate GMST. We do not want to argue which is better or worse. Nonetheless, we fully agree with the reviewer that there are caveats associated with both methods. These are now discussed extensively in the text (e.g. **L423-429, 500-508**)

4. The reported uncertainty of the "best estimate" is meaningless. An estimation uncertainty of 0.5–0.8◦C is impossible for Eocene GMST, given the large uncertainty of individual reconstructions, data scarcity, and the uneven spatial distribution of records. I suggest that a more appropriate method is used to better quantify the uncertainty, e.g., a Monte Carlo bootstrapping method.

We agree that a more appropriate method should be used to combine GMST and quantify uncertainty. As suggested, we now employ a probabilistic approach, using Monte Carlo resampling with full propagation of errors (**L590-606**).

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. From this this probability distribution, the median value and the upper and lower limits corresponding to 66 and 90% confidence limits were identified (**L590-606; see below**).



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 **(L626-641)**.

We also perform sequential removal of one GMST method at a time (jackknife resampling) to examine the influence of a single method upon the average GMST estimate. Jackknifing reveals that that no single method overly influences the mean GMST or 66% confidence intervals during the latest Paleocene, PETM or EECO (±1.5°C; Figure S9) (**L607-614**).

Finally, we also use the GMST output generated from our Monte Carlo simulations in our subsequent calculations of bulk ECS (see Section 3.4; **L684-685**).

5. In addition to the "gross ECS estimate", it would be interesting to calculate an ECS using the GMST and CO2 increases from the LP to PETM (e.g., Shaffer et al., 2016).

This is a good suggestion and we now calculate climate sensitivity between the transition from the latest Palaeocene to the PETM, assuming that non-CO_2 forcings and feedbacks are negligible. This yields an ECS estimate of 3.6°C. This is consistent with previous work (e.g. Shaffer et al., 2016). However, we note that latest Paleocene CO_2 estimates remain uncertain (Gutjahr et al., 2017) and well-synchronised, continuous and high-resolution CO_2 records are required to accurately constrain ECS during the DeepMIP intervals **(L721-731)**

Line69: If we take the modern climate as a baseline, Eocene climate forcings are more than just proxy CO2. For example, several climate forcing agents are discussed in Lunt et al. (2017). Please consider changing "CO2 proxy data" to "knowledge of climate forcing".

Amended as suggested

Line 84: Please define BWT.

BWT = Bottom water temperature. Amended as suggested.

Line 140: Please provide more details of the "modern values". Which dataset is used? What time period is used as modern reference?

The time period used is between 1979 and 2018 and we used a climatology of the full ERA-interim period (Dee et al., 2011). However, we have performed the same analysis with ERA-40 and ERA5 and find that our results are insensitive to the choice of reference period or reanalysis product (**L157-158)**

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 172: "temperature gradients are roughly half modern values or less". Please list references for this.

We have refined this sentence and added appropriate references (**L189-194)**

Line 190: Delete one "utilize two"

Amended as suggested.

Line 202: 4x CO2?

Amended as suggested.

Line 532–541: Please add a discussion of a caveat of this ECS estimate, as ECS depends on the background climate, e.g., it might increase with warming (Caballero and Huber, 2013; Zhu et al., 2019).

We agree with the reviewer and we have added discussion on ECS and its state dependence **(L721-728; 648-650).**

Figure and Table captions: Please specify what the uncertainty range in tables/figures represent (e.g., 1 sd).

Amended as suggested. Error bars on each individual method are the standard deviation (1σ), except D_{surf-1} and D_{surf-2} which use the standard error ($1\sigma_{\bar{x}}$).

Additional references:

Dee, D.P., with 35 co-authors., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. R. Meteorol. Soc., 137, 553-597 (DOI: 10.1002/qj.828).