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 217-229)**

		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**).

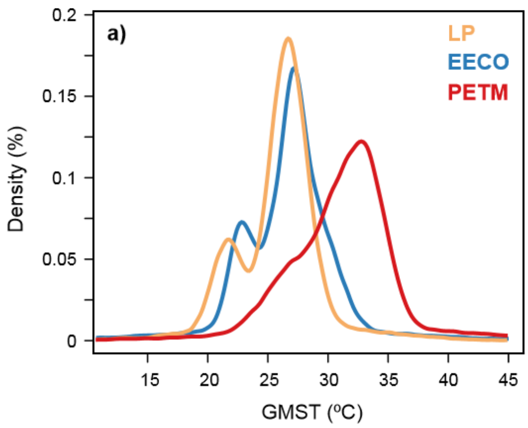
Also, Dcomb-1 is incompatible with Ddeep-1. Dcomb-1 assumes ∆GMST = 0.5 * (∆tropical SST + ∆BWT), while Ddeep-1 assumes ∆GMST = ∆BWT. It is better to keep only one method that has smaller biases.

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 1,000,000 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 6,000,000 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 **(L627-642)**.

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; **L685-686**). This yields more refined ECS estimates.

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 **(L724-734)**

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 (**L190-193)**

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 **(L724-734; 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).

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 $\langle T \rangle^{inferred}$ (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.

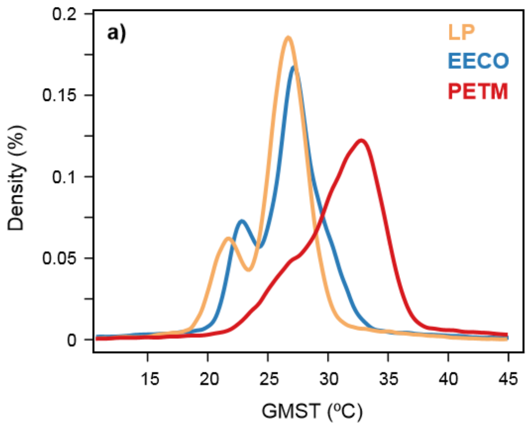
Line 470: how the uncertainties on the best GSMT can be so small.

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 (**L589-607**), to combine GMST and quantify uncertainty.

Specifically, we generate 1,000,000 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 6,000,000 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 figure 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.

Global mean surface temperature and climate sensitivity of the EECO, PETM and latest Paleocene

Gordon N. Inglis[1,2], Fran Bragg[3], Natalie Burls[4], Margot J. Cramwinckel[5], David Evans[6], Gavin L. Foster[2], Matt Huber[7], Daniel J. Lunt[3], Nicholas Siler[8], Sebastian Steinig[3], Jessica E. Tierney[9], Richard Wilkinson[10], Eleni Anagnostou[11], Agatha M. de Boer [12], Tom Dunkley Jones[13], Kirsty Edgar[13], Christopher J. Hollis[14], David K. Hutchinson[12] and Richard D. Pancost[2]

1. School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, UK
2. Organic Geochemistry Unit, School of Chemistry, School of Earth Science, Cabot Institute for the Environment, University of Bristol, UK
3. School of Geographical Sciences, University of Bristol, UK
4. Department of Atmospheric, Oceanic and Earth Sciences, George Mason University, USA
5. Department of Earth Sciences, Utrecht University, Netherlands
6. Institute of Geosciences, Goethe University Frankfurt, Frankfurt am Main, Germany
7. Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, USA
8. College of Earth, Ocean and Atmospheric Sciences, Oregon State University, USA
9. Department of Geosciences, The University of Arizona, 1040 E 4th St Tucson AZ USA
10. School of Mathematics and Statistics, University of Sheffield, UK
11. GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany#
12. Department of Geological Sciences and Bolin Centre for Climate Research, Stockholm University, Sweden.
13. School of Geography, Earth and Environmental Sciences, University of Birmingham, UK
14. GNS Science, Lower Hutt, New Zealand

Corresponding author: Gordon N. Inglis

Email: gordon.inglis@soton.ac.uk. Telephone: +44 (0)117 954 6395

Abstract:

Accurate estimates of past global mean surface temperature (GMST) help to contextualise future climate change and are required to estimate the sensitivity of the climate system to CO_2 forcing through Earth history. Previous GMST estimates for the latest Paleocene and early Eocene (~57 to 48 million years ago) span a wide range (~9 to 23°C higher than pre-industrial) and prevent an accurate assessment of climate sensitivity during this extreme greenhouse climate interval. Using the most recent data compilations, we employ a multi-method experimental framework to calculate GMST during the three DeepMIP target intervals: 1) the latest Paleocene (~57 Ma), 2) the Paleocene-Eocene Thermal Maximum (PETM; 56 Ma) and 3) the early Eocene Climatic Optimum (EECO; 49.1 to 53.3 Ma). Using six different methodologies, we find 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. GMST estimates from the EECO are ~10 to 16°C warmer than pre-industrial, higher than the estimate given by the IPCC 5th Assessment Report (9 to 14°C higher than pre-industrial). Leveraging the large 'signal' associated with these extreme warm climates, we combine estimates of GMST and CO_2 from the latest Paleocene, PETM and EECO to calculate gross estimates of the average climate sensitivity between the early Paleogene and today. We demonstrate that "bulk" equilibriums climate sensitivity (66% confidence) during the latest Paleocene, PETM and EECO is 4.5°C (2.4 to 6.8°C), 3.6°C (2.3 to 4.7°C) and 3.1°C (1.8 to 4.4°C) per doubling of CO_2. These values are generally similar to those assessed by the IPCC (1.5 to 4.5°C per doubling CO_2), but appear incompatible with low ECS values (< 1.5 per doubling CO_2).

1. Introduction

Under high growth and low mitigation scenarios, atmospheric carbon dioxide (CO_2) could exceed 1000 parts per million (ppm) by the year 2100 (Stocker et al., 2013). The long-term response of the Earth System under such elevated CO_2 concentrations remains uncertain (Stevens et al., 2016; Knutti et al., 2017; Hegerl et al., 2007). One way to better constrain these climate predictions is to examine intervals in the geological past during which greenhouse gas levels were similar to those predicted under future scenarios. This is the rationale behind the Deep-time Model Intercomparison Project (DeepMIP; www.deepmip.org) which aims to investigate the behaviour of the Earth System in three high CO_2 climate states in the latest Paleocene and early Eocene (~ 57–48 Ma) (Lunt et al., 2017; Hollis et al., 2019).

Sea surface temperature (SST) and land air temperature (LAT) proxies indicate that the latest Paleocene and early Eocene were characterised by global mean surface temperatures (GMST) much warmer than those of today (Cramwinckel et al., 2018; Farnsworth et al., 2019; Hansen et al., 2013; Zhu et al., 2019; Caballero and Huber, 2013). Having a robust quantitative estimate of the magnitude of warming at these times relative to modern is useful for two primary reasons: (1) it allows us to contextualise future climate change predictions by comparing the magnitude of future anthropogenic warming with the magnitude of past natural warming; (2) combined with knowledge of the climate forcing ~~CO_2 proxy data~~, it allows us to estimate climate sensitivity, a key metric for understanding how the climate system responds to CO_2 forcing. Using different proxy data compilations (Hollis et al., 2012; Lunt et al., 2012), the Fifth IPCC Assessment Report (AR5) stated that GMST was 9°C to 14°C higher than for pre-industrial conditions (*medium confidence*) during the early Eocene (~52 to 50 Ma) (Masson-Delmotte et al., 2014). However, subsequent studies indicate a wider range of estimates, from 9 to 23°C warmer than pre-industrial (Caballero and Huber, 2013; Cramwinckel et al., 2018; Farnsworth et al., 2019; Zhu et al., 2019; Figure 1 and Table 1). It is an open question whether this range arises from inconsistencies between the methods used to estimate GMST, such as selection of proxy datasets, treatment of uncertainty, and/or analysis of different time intervals. This methodological variability has thwarted ~~a robust assessment of GMST estimates for the latest Paleocene and early Eocene.~~robust comparisons between GMST methodologies for key intervals through the latest Paleocene to early Eocene.

Here we calculate GMST estimates within a consistent experimental framework for the target intervals outlined by DeepMIP: i) the Early Eocene Climatic Optimum (EECO; 53.3 to 49.1Ma), ii) the Paleocene-Eocene Thermal Maximum (PETM, ca. 56 Ma) and iii) the latest Paleocene (LP, ca. 57-56 Ma). We use six different methods to obtain new GMST estimates for these three time intervals, employing previously compiled SST and LAT estimates (Hollis et al., 2019) and bottom water temperature (BWT) estimates (Dunkley Jones et al., 2013; Cramer et al., 2009; Sexton et al., 2011; Littler et al., 2014; Laurentano et al., 2015; Westerhold et al., 2018; Barnet et al., 2019). We also undertake a suite of additional sensitivity studies to explore the influence of particular proxies on each GMST estimate. We then compile GMST estimates from all six methods to generate a 'combined' GMST estimate for each time slice and use these, with existing estimates of CO_2 (Gutjahr et al., 2017; Anagnostou et al., 2016) to develop new estimates of "bulk" equilibrium climate sensitivity (ECS) during the latest Paleocene, PETM and EECO.

2. Methods and Materials

Three different input datasets are used to calculate GMST: 1) dataset D_{surf} which consists of surface temperature estimates, both marine (sea surface temperatures) and terrestrial, 2) dataset D_{deep} which consists of deep-water temperature estimates, and 3) dataset D_{comb} which consists of a combination of surface- and deep-water temperature estimates. Here we make use of six different methodologies, which are described in detail below, ~~make use of these datasets~~ to estimate GMST from these datasets.

2.1. Dataset D_{surf}

Dataset D_{surf} is version 0.1 of the DeepMIP database, as described in Hollis et al (2019) (Supplementary Information). It consists of SSTs and LATs for the latest Paleocene, PETM and EECO. The SSTs are derived from multiple proxies, specifically foraminiferal $\delta^{18}O$ values, foraminiferal Mg/Ca ratios, clumped isotopes (Δ47), and isoprenoid GDGTs (TEX_{86}). Foraminiferal $\delta^{18}O$ values and Mg/Ca ratios are calibrated to SST following Hollis et al., 2019 Bemis et al. and and Evans et al. (2018), respectively. TEX_{86} values are calibrated to SST using BAYSPAR (Tierney and Tingley, 2014). Δ47 values are reported using the parameters and calibrations of the original publications (Evans et al., 2018; Keating-Bitonti et al., 2011). LATs are derived from leaf fossils, pollen assemblages, mammal $\delta^{18}O$, paleosol $\delta^{18}O$, paleosol climofunctions and branched GDGTs. LAT estimates are calculated using the parameters and calibrations of the original publications (see Hollis et al., 2019 and ref. therein). The locations of the proxy datasets are shown in Figure S1 using the paleomagnetic-based reference frame (Hollis et al., 2019). For each dataset, we utilise the uncertainty range of temperature estimates reported in Hollis et al. (2019). Although the level of uncertainty between proxies is vastly different, we do not explore calibration uncertainty. Instead, we focus on the methodologies used to calculate GMST.

Four methods (*D_{surf}-1, D_{surf}-2, D_{surf}-3* and *D_{surf}-4*) are employed to calculate GMST from dataset D_{surf}. These methods employ parametric (*D_{surf}-1, D_{surf}-2, D_{surf}-4*) or non-parametric (*D_{surf}-3*) functions to estimate temperature. We calculate GMST on the mantle-based reference frame and employ the rotations provided in Hollis et al (2019). These differ very slightly from those utilised in the DeepMIP model simulations (Lunt et al, 2020). Each method conducts a 'baseline' calculation that uses the SST and LAT data compiled in accordance with the DeepMIP protocols (i.e. Hollis et al., 2019). Our baseline calculation (*D_{surf}-baseline*; Table 2) excludes $\delta^{18}O$ values from recrystallized planktonic foraminifera because the resulting temperature estimates are biased by diagenesis toward significantly cooler temperatures than those derived from the $\delta^{18}O$ value of similar aged and similarly located well-preserved foraminifera, foraminiferal Mg/Ca ratios and Δ47 values from larger benthic foraminifera (Pearson et al., 2001; Hollis et al., 2019 and ref. therein). For each method, we also conduct a series of illustrative sub-sampling calculations relative to *D_{surf}-baseline*, based on varying assumptions about the robustness of different proxies (Table 2). The first sensitivity experiment (*D_{surf}-Frosty;* Table 2) includes $\delta^{18}O$ values from recrystallized planktonic foraminifera. The second sensitivity experiment (*D_{surf}-NoTEX;* Table 2) removes TEX_{86} values as these give slightly higher SSTs than other proxies, especially in the mid-to-high latitudes (Bijl et al., 2009; Hollis et al., 2012; Inglis et al., 2015). The third sensitivity experiment (*D_{surf}-NoMBT;* Table 2) removes MBT(')/CBT values derived from marine sediment archives as they may suffer from a cool bias (Inglis et al., 2017; Hollis et al., 2019). The fourth sensitivity experiment (*D_{surf}-NoPaleosol;* Table 2) removes mammal/paleosol $\delta^{18}O$ values and paleosol climofunctions as these proxies may suffer from a cool bias (Hyland and Sheldon, 2013; Hollis et al., 2019). For each method, GMST is calculated for: i) the Early Eocene Climatic Optimum (EECO; 53.3 to 49.1 Ma), ii) the Paleocene-Eocene Thermal Maximum (ca. 56 Ma) and iii) the latest Paleocene (LP; ca. 57-56 Ma).

2.1.1. D_{surf}-1

Method D_{surf}-1 was first employed by Caballero and Huber (2013) to estimate GMST from early Eocene surface temperature proxies after it was recognised that pervasive recrystallization of foraminiferal $\delta^{18}O$ could overprint the original SST signal (e.g. Pearson et al., 2001; Pearson et al., 2007). That study used data compilations (Huber and Caballero, 2011, Hollis et al., 2012) which were the predecessors to the DeepMIP compilation (Hollis et al., 2019).

Here, the anomalies of individual proxy temperature data points with respect to modern values at the corresponding paleolocation are first calculated. The time period used is between 1979 and 2018 and we used a climatology of the full ERA-interim period (Dee et al., 2011). The calculation involves binning into low, mid, and high latitudes (30°N to 30°S, 30°N/S to

60°N/S, and 60°N/S to 90°N/S), and calculating the unweighted mean anomaly within these bins between the median reconstructed value at a given locality and the temperature in the modern system (from reanalysis). The geographically binned means are then weighted according to relative spherical area to calculate a globally weighted mean temperature anomaly between the paleo-time slice and modern. All samples are treated equally and considered independent. The associated errors are added in quadrature with the inter-sample standard deviation. These two sources of error were combined and normalized by the square root of the number of samples. This method is intended as an unsophisticated, brute force approach to estimating GMST when dealing with many localities with poorly characterized errors in which there is a large difference between the reconstructed temperature at a given location and the modern equivalent. It is not intended to identify small changes in GMST; nor is it expected to work well under conditions in which temperature gradients are stronger than today, continents are far removed from their current configuration, or in situations in which systematic errors are not readily mitigated by large sample size (i.e. when there are correlations in systematic errors between proxies). It is designed to be relatively straightforward to interpret and simple to reproduce without relying overly on climate models or sophisticated statistical models.

Various sanity checks have been performed to determine if the method is likely to produce useful results for a given sampling distribution and what corrections should be applied to optimize it. For example, if the modern temperature field is sampled using a geographic sampling distribution for a given time interval, what would the reconstructed modern temperature be? Sampling the modern global annual average surface temperature field in the reanalysis product ERA-5 yields a mean value of 15.1°C but when resampled at the equivalent geographic distribution of our samples from the latest Paleocene, PETM and EECO yields mean values for the modern of 16.9°C (±1.8°C), 14.2°C (±1.7°C), and 15.2°C (±1.1°C), respectively. Thus, for the sampling densities and spatial structure of the early Paleogene, this method can approach the true value within ~1.5°C and the error propagation adequately characterizes the error, in this 'perfect knowledge' scenario. Seeking precision beyond that range is unwarranted and as indicated above, systematic biases are a serious concern. However, estimating the latest Paleocene and early Eocene GMST may be somewhat easier than estimating the modern GMST because temperature gradients were much reduced from modern. Huber and Caballero (2011) estimate a reduction to less than half the modern temperature gradient whilst Evans et al (2018) constrain the low-to-high latitude SST gradient to at least ~30% (+/- 10%) weaker than modern (Evans et al., 2018).

Alongside modern observations, we can also use paleoclimate model results to characterise how well the existing palaeogeographic sampling network will impact results (Figure 2). Here we utilize two CESM1 simulations, as described in Cramwinckel et al., (2018; EO3 and EO4). The two cases are chosen to minimize the magnitude of the correction to GMST and the final result is not sensitive to the choice of reference simulation between these two (Supplementary Information). For each interval, the difference between reconstructed global temperatures and the true paleoclimate model mean is <1 to 3°C. These comparisons demonstrate that this method produces estimates that are within random error given otherwise perfect knowledge. The ~~systematic~~ errors introduced by limited paleogeographic sampling can be alleviated by incorporating the ~~systematic~~ offset in mean values between the true paleoclimate model GMST and the sampled paleoclimate model GMST outlined above (Figure 2). We utilise this offset to correct for systematic errors, but this is the only component in which paleoclimate model information is included in this GMST estimation methodology. This approach is best applied within the context of studying the random and systematic error structure as described above and caution should be taken in using systematic corrections that are significantly bigger than the estimated random error. The underlying assumption is that the bias in the global mean estimate that exists due to uneven sampling is the same in the ‘proxy’ Eocene world as in the ‘model’ Eocene world, i.e. that the zonal and meridional gradients are well characterised by the model, even if the absolute temperatures are not. ~~The calculations shown here utilize two utilize two CESM1 simulations, as described in Cramwinckel et al., (2018; EO3 and EO4). The two cases are chosen to minimize the~~

~~magnitude of the correction to GMST and the final result is not sensitivity to the choice of reference simulation among these two.~~

We note that the magnitude of the global correction could be sensitive to different models and/or boundary conditions. To explore this further, we performed the same analysis using Community Earth System Model version 1.2 (CESM1.2) at 6x CO_2. This model simulation offers a major improvement over earlier models (Zhu et al., 2019) due to the improved treatment of cloud microphysics and is able to reproduce key features of the early Paleogene (e.g. better meridional SST gradient; Zhu et al., 2019; Lunt et al., 2020). We find that CESM1 (8x and 16x CO_2) and CESM1.2 (6x CO_2) yield similar GMST estimates during the PETM, EECO and latest Paleocene. For example, GMST values (obtained using D_{surf}-baseline) during the EECO average 24.5°C, 24.6°C and 25.2°C for CESM1 (x8 CO_2), CESM1 (x16 CO_2) and CESM1.2 (6x CO_2), respectively. This indicates that the final result is not overly sensitive to the choice of reference simulation, at least within the CESM model family. In the following sections, we only discuss CESM1 simulations to avoid circularity if the results from this paper are used to evaluate more recent simulations (e.g. CESM1.2; Lunt et al., 2020).

2.1.2. D_{surf}-2

GMST estimates are calculated using the method described in Farnsworth et al. (2019), in which a transfer-function is used to calculate global mean temperature from local proxy temperatures. The transfer function is generated from a pair of early Eocene climate model simulations, carried out at two CO_2 concentrations. The first simulations are the same 2x CO_2 and 4x CO_2 HadCM3L Eocene simulations from Farnsworth et al (2019). The second simulations are the x 4CO_2 and 8x CO_2 CCSM3 simulations of Huber and Caballero (2011), also discussed in Lunt et al (2012). The two models are configured for the Eocene with different paleogeographies (Supplementary Table S1). We provide a final estimate based on the mean of our two models. ~~The two models are configured for the Eocene with different paleogeographies.~~

The principal assumption of this approach is that global temperatures scale linearly with local temperatures, and that a climate model can represent this scaling correctly (see below). The resulting GMST estimate is therefore independent of the climate sensitivity of the model but dependent on the modelled spatial distribution of temperature. For a single given proxy location with a local temperature estimate (T^{proxy}), Farnsworth et al. (2019) estimate global GMST ($<T>^{inferred}$) as:

$$<T>^{inferred} = <T^{low}> + (T^{proxy} - T^{low}) \frac{<T^{high}> - <T^{low}>}{T^{high} - T^{low}} \quad (1)$$

where $<T^{low}>$ *and* $<T^{high}>$ are the global means of a low- and high-CO_2 model simulation respectively, and T^{low} *and* T^{high} are the local temperatures (same location as the proxy) from the same simulations. T^{low} and T^{high} represent local modelled SSTs or local modelled near-surface LATs (in contrast to Farnsworth et al. 2019, who only used local modelled near-surface LATs to calculate T^{low} and T^{high}, even if T^{proxy} was SST). If the proxy temperature is greater than T^{high} or cooler than T^{low}, then the inferred global mean is found by extrapolation rather than by interpolation and is therefore more uncertain (Figure 3). This 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. We repeat this process for each proxy data location (Figure 4) and take an average over all proxy locations as our best estimate of global mean temperature.

Recent work has demonstrated that CESM1.2 and GFDL model simulations offer a major improvement over earlier models (Zhu et al., 2019; Lunt et al., 2020). As such, we also calculated GMST using CESM1.2 (3x and 6x CO_2; Zhu et al., 2019) and GFDL (3x and 6x CO_2; Hutchinson et al., 2018; Lunt et al., 2020). We find that all four simulations (i.e. HadCM3L, CCSM3, CESM1.2 and GFDL) yield similar GMST estimates (e.g. GMST during the PETM ranges between 32.3 and 34.5°C; Supplementary Information). This demonstrates that D_{surf}-2 is not overly sensitive to the climate model simulation. However, as CESM1.2 and GFDL have greater polar amplification than other models (e.g. HadCM3L, CESM1), GMST is more likely to be found by interpolation (c.f. extrapolation). In the following sections, we employ CCSM3 and HadCM3 simulations 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). To explore whether GMST scales linearly with local temperatures, we used CESM1.2 to re-calculate GMST using the same method as above but using the 9x CO_2 simulation in place of the 6x CO_2 simulation. We find that GMST estimates are very similar (±0.4°C). This is because, although the relationship between GMST and CO_2 is non-linear (Zhu et al, 2019), the relationship between local and global temperature is relatively constant.

2.1.3. D_{surf}-3

For *D_{surf}-3*, GMST estimates are calculated using Gaussian process regression (Figure 5; Bragg et al., in prep). In this method, temperature is treated as an unknown function of location, *f(x)*. Many possible functions can fit the available proxy dataset. By using a Gaussian process model of the unknown function, we assume that temperature is a continuous and smoothly varying function of location, and once fitted to the data, the posterior mean of the model gives the most likely function form for the temperature. We use a Gaussian process prior and update it using the proxy data to obtain the posterior model which we can then use to predict the surface temperatures on a global grid. Prior specification of the model is via a mean function *E(f(x)) = m(x),* and a covariance function *Cov(f(x), f(x'))* =*k(x,x')* (which tells us how correlated *f(x)* is with *f(x')*). We also specify the standard deviation of the observation uncertainty about each data point (σ_i^2). If $\boldsymbol{f} = \left(f(x_1), \dots f(x_n)\right)^T$ is a vector of temperature observations at each location x_i, then the model is:

$$\boldsymbol{f} \sim \mathcal{N}(\mu, \Sigma) \quad (2)$$

where $\mu_i = m(x_i)$ and $\Sigma_{ij} = k(x_i, x_j) + \mathbb{I}_{i=j}\sigma_i^2$. The proxy temperatures are expressed as anomalies to either the marine or terrestrial present-day zonal mean temperature at the respective paleolatitude. We subtract the mean temperature anomaly (weighted by the paleolatitude) for each time period and core experiment prior to the analysis and therefore fit the model to the residuals. This means the predicted field will relax towards the mean surface warming in areas of no data coverage. The covariance function – which considers the clustering of proxy locations – describes the correlation between $f(x_i)$ and $f(x_j)$ in relation to the distance of x_i and x_j. We use a squared-exponential covariance function with Haversine distances replacing Euclidean distances so that correlation is a function of distance on the sphere.

A heteroscedastic noise model is used to weight the influence of individual proxy data by their associated uncertainty, i.e. the model will better fit reconstructions with a smaller reported error. Proxy uncertainties are taken from Hollis et al., (2019). Standard deviations for TEX_{86}, Mg/Ca and $\delta^{18}O$ records are derived from the reported 90% confidence intervals (Hollis et al., 2019). A minimum value of 2.5°C for the standard deviation is assumed for all other methods. The output variances and length scale of the covariance function are estimated using their maximum likelihood values, obtained with the GPy Python package (GPy, 2012). We apply the method to the marine and terrestrial data separately and combine the masked fields afterwards to prevent mutual interference. We further constrain the lower bound of the lengthscale parameter to 2000 km to always fit a reasonably smooth surface, even in some continental areas with noisy proxy data (e.g. western North America). We note that our choice of the minimum lengthscale and the separation of land and ocean temperatures influence the predicted regional surface temperature patterns but do not significantly change our GMST estimates.

The Gaussian process approach provides probabilistic predictions of temperature values, i.e., uncertainty estimates of the predicted field. The uncertainty reported for an individual GMST estimate is calculated via random sampling. We generate 10,000 surfaces from a multivariate normal distribution based on the predicted mean and full covariance matrix and calculate the GMST for each sample. Uncertainty of the mean estimate is then defined as the standard deviation of the 10,000 random samples. Regional model uncertainty (expressed as standard deviation fields) is typically highest in areas with sparse data coverage (e.g. the Pacific Ocean and Southern Hemisphere landmasses; Figure S2). The lower uncertainty for the latest Paleocene relative to the PETM and EECO is related to the smaller reported uncertainties in the proxy dataset rather than enhanced data coverage. The large spread in reconstructed terrestrial temperatures for North America during the PETM and EECO (Figure S2) ~~also increases uncertainties for other continental areas during both time intervals.~~propagates through into relatively large uncertainties in the GMSTs estimates for these intervals.

2.1.4. D_{surf}-4

For D_{surf}-4, GMST estimates are calculated using a simple function of latitude (θ), tuned to best fit the proxy data:

$$T(\theta) \approx a + b\theta + c\cos\theta \quad (3)$$

where $T(\theta)$ is the Eocene zonal-mean temperature, and the coefficients a, b, and c are chosen to minimize the sum of the squared residuals relative to D_{surf} (i.e. the SST and LAT data from Hollis et al. 2019). This new model represents $T(\theta)$ well in the modern climate (Figure S3) when supplied with similar number of data points as are in the Hollis et al (2019) dataset, and it ensures a global solution that is consistent with the physical expectation that temperature should decrease - and the meridional gradient in temperature should increase - from the tropics toward the poles (Figure S3).

For each data point, we account for three types of uncertainty (i.e. temperature, elevation, latitude). For temperature, we assume a skew-normal probability distribution based on the stated 90% confidence intervals. Where uncertainty estimates are not given, we assume a (symmetric) normal distribution with a 90% confidence interval of ±5K. For elevation, we assume a skew-normal distribution with a 90% confidence interval equal to the lowest and highest elevations of adjacent grid points in the paleotopography data set of Herold et al. (2014), with a lower bound of zero.

T(θ) was estimated by sampling temperature, elevation, and latitude from their respective distributions at each location (Figure S4) and a lapse-rate adjustment of 6°K/km was applied. Then, using a standard Monte Carlo bootstrapping method, the same number of data points were resampled via replacement, and the coefficients in Equation 3 were found that best fit the sub-sampled data. This procedure was repeated 10,000 times to find a probability distribution of *T(θ)*. The uncertainty associated with an individual GMST estimate is the standard deviation.

2.2. Dataset D_{deep}

Dataset D_{deep} consists of benthic foraminiferal δ^{18}O-derived bottom water temperatures (BWTs) for the latest Paleocene, PETM and EECO. The benthic foraminiferal δ^{18}O dataset is based on previous compilations (Dunkley Jones et al., 2013; Cramer et al., 2009), updated to include more recently published datasets (Sexton et al., 2011; Littler et al., 2014; Laurentano et al., 2015; Westerhold et al., 2018; Barnet et al., 2019). ~~latest Paleocene, PETM and EECO come from previous studies (Westerhold et al., 2018;Barnet et al., 2019;Cramer et al., 2009).~~ The EECO dataset is sourced from eleven sites, providing spatial coverage of both the Pacific, Atlantic and Indian Oceans (DSDP/ODP Sites 401, 550, 577, 690, 702, 738, 865, 1209, 1258, 1262, & 1263). The PETM and latest Paleocene datasets are sourced from a compilation of nine and seven sites, respectively, differing from Dunkley-Jones et al. (2013) in that: i) more recent datasets were added, and ii) PETM sites with a muted CIE magnitude (< 1.5 ‰) were excluded as these datasets may be missing the core PETM interval (Supplementary Table 2). Benthic foraminifera $\delta^{18}O$ values are adjusted to *Cibicidoides* following established methods (Cramer *et al*., 2009), allowing temperature to be calculated using Eq. 9 of Marchitto et al (2014):

$$(\delta_{cp} - \delta_{sw} + 0.27) = -0.245 \pm 0.005t + 0.0011 \pm 0.0002t^2 + 3.58 \pm 0.02 \qquad (4)$$

where t is bottom water temperature in Celsius, δ_{cp} is $\delta^{18}O$ of $CaCO_3$ on the Vienna-Pee Dee Belemnite (VPDB) scale, and δ_{sw} is $\delta^{18}O$ of seawater on the Standard Mean Ocean Water (SMOW). δ_{sw} is defined in accordance with the DeepMIP protocols (−1.00 ‰; see Hollis et al., 2019). ~~A single method (*D_{deep}-1*) is used to calculate GMST from *D_{deep}* following the methodology outlined in Hansen *et al*. (2013). For this method, GMST is calculated for: i) the Early Eocene Climatic Optimum (EECO; 53.3 to 49.4 Ma), ii) the Paleocene-Eocene Thermal Maximum (ca. 56 Ma) and iii) the latest Paleocene (LP; ca. 57-56 Ma).~~

2.2.1. D_{deep}-1

For *D_{deep}-1*, GMST estimates are calculated following the method of Hansen et al. (2013), which utilises only the deep ocean benthic foraminifera $\delta^{18}O$ dataset, and we refer the reader to that study for a detailed justification of the approach. Briefly, for time periods prior to the Pliocene, GMST is scaled directly to deep ocean temperature. Specifically, $\Delta GMST = \Delta BWT$ prior to ~5.3 Ma, where early Pliocene BWT and GMST was calculated following Eq. 3.5, 3.6, and 4.2 of Hansen et al. (2013). As such, the calculations presented here differ from those of Hansen et al. (2013) only in that: i) we use the revised benthic $\delta^{18}O$ compilation described above rather than that of Zachos et al. (2008), and ii) a different equation (Eq. 4) to convert $\delta^{18}O$ to temperature.

2.3. Dataset D_{comb}

Dataset D_{comb} uses a combination of (tropical) surface- and deep-water temperature estimates. The deep ocean dataset (D_{deep}) is identical to that described in Section 2.2. The tropical SST dataset utilises all relevant surface ocean proxy data from the DeepMIP database, i.e. those with a palaeolatitude in the magnetic reference frame within 30° of the equator. An expanded (relative to modern) definition of the tropics is used because tropical SST reconstructions are relatively sparse; 30° was chosen because it retains tropical SST data from several proxies for all three intervals whilst SST seasonality remains relatively low within these latitudinal bounds.

2.3.1. D_{comb}-1

For D_{comb}-*1*, GMST estimates are calculated for each time interval based on the difference between tropical SSTs and deep-ocean BWTs (Evans et al., 2018), such that:

$$GMST = 0.5(\overline{tropical\ SST} + \overline{BWT}) \qquad (5)$$

The fundamental assumptions of this approach are that: 1) GMST can be approximated by global mean SST, 2) global mean SST is equivalent to the mean of the tropical and high latitude regions, 3) benthic temperatures are representative of high latitude surface temperatures and 4) that the temperature gradient between the abyss and high latitude SST is fixed through time (c.f. Sijp et al., 2011). ~~Applying these assumptions to the modern ocean would generate a GMST estimate within ~1°C of measured and a modern latitudinal SST gradient within ~1°C of the surface ocean dataset (as discussed in Evans *et al*., 2018 and Supplementary Information).~~ To test these assumptions from a theoretical perspective, we modelled the shape of the latitudinal temperature gradient using a simple algebraic function (Figure S5). These results suggest that D_{comb}-1 may underestimate GMST by 0.75 to 1.25 °C.

We also compared GMST from the EO3 and EO4 model simulations of Cramwinckel et al. (2018) to that calculated using D_{comb}-1 (Figure S5) and find a similar cold bias (~1 to 3°C). However, we note that these findings depend on the accuracy of the modelled deep ocean temperatures.

Probability distributions for each time interval were computed as follows. In the case of the tropical SST data, 1000 subsamples were taken, following which a random normally distributed error was added to each data point in the DeepMIP compilation, including both calibration uncertainty and variance in the data where multiple reconstructions are available for a given site and time interval. Mean tropical SST was calculated for each of these subsamples. To provide a BWT dataset of the same size as the subsampled tropical SST data, 1000 normally distributed values were calculated for each time interval, based on the mean ±1SD variation of the pooled benthic $\delta^{18}O$ data from all sites including calibration uncertainty.

3. Results and Discussion

~~3.1. D_{surf}-1 to -4~~

~~GMST estimates (*D_{surf} default*) of the latest Paleocene (n =4) range between 26.6 and 27.6°C (Table 3). GMST estimates (*D_{surf} default*) of the PETM (n = 4) range between 30.6 and 33.9°C (Table 3). GMST estimates (*D_{surf} default*) of the EECO (n = 4) range between 24.5 and 29.8°C (Table 3). All four methods indicate that: 1) the PETM is warmer than the latest Paleocene (by ~4 to 7°C) and: 2) the PETM is warmer than the EECO (by ~3 to 9°C). GMST estimates derived using *D_{surf}-Frosty* (i.e. which include recrystallized planktonic foraminifera $\delta^{18}O$ values) are consistently lower (up to 3.5°C) than those derived using *D_{surf} default.* GMST estimates derived using *D_{surf}-NoTEX* (i.e. which exclude TEX_{86} estimates) are also consistently lower (up to ~2°C) than those derived using *D_{surf} default*. GMST estimates derived using *D_{surf}-NoMBT* (i.e. which exclude MBT'/CBT values from marine sediments) are higher than GMST estimates~~

~~derived using D_{surf} *default* (up to 1°C). GMST estimates derived using D_{surf}-*NoPaleosol*(i.e. which exclude $\delta^{18}O$ mammal/paleosol values and paleosol climofunctions) are similar to GMST estimates derived using D_{surf} *default* (±0.5°C), with the exception of D_{surf}-*1* during the EECO which is ~3°C higher when $\delta^{18}O$ mammal/paleosol values and paleosol climofunctions are excluded.~~

~~3.2. D_{deep}-1~~

~~For the D_{deep} methodology, GMST estimates (D_{deep}) of the latest Paleocene, PETM and EECO average 25.8°C (±1.4°C), 31.1°C (±2.9°C) and 28.0°C (±1.3°C), respectively (Table 3). This method indicates that: 1) the PETM is warmer than the latest Paleocene (by ~5°C) and, 2) the PETM is warmer than the EECO (by ~3 °C).~~

~~3.3. D_{comb}-1~~

~~For the D_{comb} methodology GMST estimates (D_{comb}) of the latest Paleocene, PETM and EECO average 21.6°C (±1.2°C), 26.6°C (±2.1°C) and 22.7°C (±1.0°C), respectively (Table 3). This method indicates that: 1) the PETM is warmer than the latest Paleocene (by ~5°C) and, 2) the PETM is warmer than the EECO (by ~4°C).~~

3.1. ~~Intercomparison of methods for calculating GMST~~ Comparison of surface- and bottom water temperature-derived GMST estimates

The following section discusses our 'baseline' GMST estimates on the mantle-based reference frame only. During the latest Paleocene and PETM, GMST estimates derived from D_{surf}-*baseline* average ~27 and 33°C, respectively (Table 3; Figure 6). These values are consistent with previous studies analysing the latest Paleocene (~27°C; Zhu et al., 2019) and PETM (~32°C; Zhu et al., 2019). During the EECO, GMST estimates calculated using D_{surf} average ~27°C (Figure 6). These values are up to 3°C lower compared to previous estimates from similar time intervals (ca. 29 to 30°C; Huber and Caballero, 2011; Caballero and Huber, 2013; Zhu et al., 2019). This is likely because we use an expanded LAT dataset (n = 80) compared to previous studies (n = 51; Huber and Caballero, 2011); 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).This and could skew GMST estimates towards lower values. This is less pronounced in previous studies (i.e. Zhu et al. 2019) because they utilise a different compilation with fewer LAT estimates (n = 51; Huber and Caballero, 2011). Secondly, the inclusion of $\delta^{18}O$ values from paleosols/mammals and paleosol climofunctions leads to a cold bias in GMST estimates. This suggests that more investigation of the systematic cold bias introduced by paleosol is warranted. To confirm this, we calculated GMST values using LAT proxies only (Supplementary Information). We show that LAT-only GMST estimates are up to 6°C lower than our 'baseline' (SST + LAT) calculations, suggesting that EECO GMST estimates (D_{surf}-*baseline*) may represent a minimum temperature constraint.

GMST estimates for the latest Paleocene, PETM and EECO, calculated using D_{deep}, are 25.8°C (±1.4°C), 31.1 (± 2.9°C) and 28.0°C (±1.3°C) respectively (Table 3; Figure 6). These estimates are comparable to those derived from surface temperature proxies alone (Table 3). GMST estimates from the EECO are also comparable to previous estimates based on globally distributed benthic foraminifera data (~28°C; Hansen et al., 2013). As benthic foraminifera are less susceptible to diagenetic alteration than planktonic foraminifera (e.g. Edgar et al., 2013), this implies that benthic foraminiferal $\delta^{18}O$ values could be used to provide the 'fine temporal structure' of Cenozoic temperature change (e.g. Lunt et al., 2016;Hansen et al., 2013). However, we also urge caution as this approach scales GMST directly to BWT prior to the Pliocene and assumes that the characteristics of polar amplification are constant through time (c.f. Evans et al., 2018; Cramwinckel et al., 2018). Changes in ice volume may also influence the benthic foraminiferal $\delta^{18}O$ signal (see Hansen et al., 2013) and additional corrections are required before applying this method to other time intervals (e.g. the Eocene-

Oligocene transition). D_{deep} also implies that vertical ocean stratification is fixed, even though vertical ocean stratification has been proposed to change dramatically in the past (e.g. Sijp et al., 2013; Goldner et al., 2014) and may shift the slope and/or intercept of the relationship between BWT and GMST. In addition, the D_{deep} GMST estimate for the PETM is associated with a large uncertainty. This is due to differences in $\delta^{18}O$ values between sites and an overall lack of PETM benthic data (n = 38 from 9 sites) rather than an inherent uncertainty in the proxy or method of calculating GMST.

GMST estimates for the latest Paleocene, PETM and EECO, calculated using D_{comb}, are 21.6°C (±1.2°C), 26.6 (± 2.1°C) and 22.8°C (± 1.0°C), respectively (Figure 6). These estimates are consistently lower (up to 5°C) than GMST estimates derived using D_{surf} and D_{deep}. Although D_{comb}-1 can estimate modern GMST within ~1 °C of measured values, whether this approach can be applied in greenhouse climates remains to be confirmed. As described above, we 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, we suggest that D_{comb}-1 may reflect a minimum GMST constraint.ggest this mismatch could be related to two factors. First, if deep water formation preferentially takes place during the winter months, GMST estimates will be biased towards lower values. Secondly, there are relatively few tropical SST estimates during the EECO (n = 10 sites). As such, D_{comb} may not be fully representative of actual tropical warmth. Secondly, this method We suggest that variable weighting of the deep ocean and tropics could improve the D_{comb} method in future studies (Eq. 5 gives an equal weighting to each).

3.2. Influence of different proxy datasets upon D_{surf}-derived GMST estimates

To explore the importance of the proxies themselves upon our reconstructed latest Paleocene, PETM and EECO D_{surf}-derived GMST estimates, we conducted a series of illustrative subsampling experiments relative to D_{surf}-baseline (Table 2). This was performed for methods *D_{surf}-1, -2, -3* and *-4*. In the first subsampling experiment (D_{surf}-Frosty; Table 2), we include $\delta^{18}O$ SST estimates from recrystallized planktonic foraminifera. This yields lower GMST estimates (<1 to 4°C; e.g. Figure S6-8) and is consistent amongst all four methods. This agrees with previous studies which indicate that $\delta^{18}O$ values from recrystallized planktonic foraminifera are significantly colder than estimates derived from the $\delta^{18}O$ value of well-preserved foraminifera (Pearson et al., 2001; Sexton et al., 2006; Edgar et al., 2015), foraminiferal Mg/Ca ratios (Creech et al., 2010; Hollis et al., 2012) and clumped isotope values from larger benthic foraminifera (Evans et al., 2018).

The removal of TEX_{86} results in lower GMST estimates (~1 to 4 °C; e.g. Figure S6-8) across all methodologies (D_{surf}-NoTEX; Table 2). This is consistent with previous studies which indicate that TEX_{86} gives slightly higher SSTs than other proxies, especially in the mid-to-high latitudes (e.g. Hollis et al., 2012; Inglis et al. 2015. The functional response of TEX_{86} at higher than modern SSTs remains relatively uncertain, which may explain why TEX_{86} gives slightly higher SSTs than other proxies (see discussion in Hollis et al., 2019). New indices or calibrations could help to reduce the uncertainty associated with TEX_{86}-derived SST estimates beyond the modern calibration range ~~(e.g. Bayesian regression models; Tierney and Tingley, 2014)~~. TEX_{86} values can also be complicated by the input of isoGDGTs from other sources (see discussion in Hollis et al., 2019). The DeepMIP database excludes samples with anomalous GDGT distributions (Hollis et al., 2019). However, a Gaussian process regression (GPR) model may help to better identify anomalous GDGT distributions in the sedimentary record using a nearest neighbour distance metric (Eley et al., 2019). This methodology could be employed in future studies to further refine GDGT-based SST datasets, but this methodology is currently under review and is not further considered here. Despite the caveats and concerns raised in previous work, the exclusion of TEX_{86} data shifts GMST by a relatively small amount.

The input of brGDGTs from archives other than mineral soils or peat can bias LAT estimates towards lower values (Inglis et al., 2017; Hollis et al., 2019) and the exclusion of MBT(')/CBT-derived LAT estimates could yield higher GMST values. Excluding MBT(')/CBT in marine sediments does yield slightly warmer GMST estimates (0.5 to 1.0°C). However, the impact of excluding MBT(')/CBT values is relatively minor because there are other proxies (e.g. pollen assemblages, leaf floral) which yield comparable LAT estimates in the regions where MBT(')/CBT values are removed (e.g. the SW Pacific).

The removal of $\delta^{18}O$ values from paleosols/mammals and paleosol climofunctions (D_{surf}-NoPaleosol; Table 2) also leads to slightly warmer GMST estimates (~0.5°C). This may be related to additional controls on paleosol and mammal $\delta^{18}O$ values. This includes variations in the isotopic composition of rainfall (i.e. meteoric $\delta^{18}O$; Hyland and Sheldon, 2013), variations in soil water $\delta^{18}O$ values (Hyland and Sheldon, 2013) and/or $\delta^{18}O$ heterogeneity within nodules (e.g. Dworkin et al. 2005). 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. Intriguingly*,* D_{surf}-1 method yields much higher GMST estimates during the EECO when $\delta^{18}O$ values from paleosols/mammals and paleosol climofunctions are excluded (~3°C higher than *D_{surf}-baseline).* This is attributed to the inclusion of two "cold" LAT estimates from the Salta Basin, NW Argentina (Hyland et al., 2017) which overly influence GMST (e.g. Figure 2). ~~These estimates are derived from paleosol climofunctions. These include the salinization index (SAL) (Sheldon et al., 2002) and the paleosol weathering index (PWI) (Gallagher and Sheldon, 2013), both of which yield a cold bias in the original DeepMIP database (Hollis et al. 2019).~~ For *D_{surf}-1*, a direct comparison of new and prior estimates (Caballero and Huber, 2013) can be made in which the only change has been the use of a newer data compilation. For our new estimate, the EECO is ~4.5°C colder than previous estimates (29.75°C; Caballero and Huber, 2013). Given that the floristic LAT estimates are identical between the DeepMIP compilation and the older compilation, the lower GMST estimates are largely due to the incorporation of additional LAT datasets (e.g. paleosol climofunctions).

3.3. A combined estimate of GMST during the DeepMIP target intervals

To derive a combined estimate of GMST during the latest Paleocene, PETM and EECO, we employ a probabilistic approach, using Monte Carlo resampling with full propagation of errors. Our combined estimates employs GMST estimates from each ‘baseline’ experiment (except *D_{surf}-1* for the EECO for which we use *D_{surf}-NoPaleosol;* see discussion above). ~~and calculate a weighter average. This approach is useful because it assigns lower confidence to GMST estimates associated with larger uncertainties (e.g. Ddeep-1 during the PETM). The reported uncertainty is the reciprocal square root of the sum of all the individual weights.~~ We generated 1,000,000 iterations for each of the six methods, for each time interval (latest Paleocene, 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 6,000,000 GMST iterations for each time period are thus simply added into a single probability density function, in order to fully represent uncertainty (Figure 7). ~~and calculate a weighted average (Figure 8). This approach is useful because it assigns lower confidence to GMST estimates associated with larger uncertainties (e.g. *D_{deep}-1* during the PETM). The reported uncertainty is the reciprocal square root of the sum of all the individual weights.~~ From this probability distribution, the median value and the upper and lower limits corresponding to 66 and 90% confidence limits were identified.

Sequential removal of one GMST method at a time (jackknife resampling) was performed 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; Supplementary

Information and Figure S9). However, the removal of D_{surf}-2 (which has relative large error bars; Figure 6) does reduce the 90% confidence interval. We also show that removing D_{comb}-1 removes the bimodality of the temperature distribution (Figure S9). This is because D_{comb}-1 is associated with consistently lower GMST estimates compared to other methods (Figure 6).

~~We find that the average GMST estimate for the latest PaleocenHe, PETM and EECO are 25.7°C (± 0.6°C), 32.7°C (± 0.8°C) and 27.3°C (± 0.5°C), respectively (Figure 7). Assuming a preindustrial GMST of 14°C, our average GMST estimates indicate that the latest Paleocene, PETM and EECO are +11.7°C, +18.7°C and +13.3°C warmer than pre-industrial, respectively. The GMST anomaly for the EECO cooler than previous studies (~15°C warmer than pre-industrial; Caballero and Huber, 2013; Zhu et al., 2019) but consistent with the range quoted previously in the IPCC AR5 (9 to 14°C warmer than pre-industrial). On average, GMST increases by ~6 to 7°C between the latest Paleocene and PETM, in keeping with previous estimates (Frieling et al., 2019; Dunkley Jones, 2013). The PETM is approximately 5°C warmer than the EECO. This is higher than previously suggested (~3°C; Zhu et al., 2019) and may related to a cold bias in EECO GMST estimates (see Section 4.2).~~

During the latest Paleocene, the average GMST estimate is 26.3°C and ranges between 22.3 and 28.3°C (66% confidence interval; Table 4; Figure 7). During the PETM, the average GMST is higher (31.6°C) and ranges between 27.2 and 34.5°C (66% confidence interval; Table 4; Figure 7). Assuming a preindustrial GMST of 14°C, our average GMST estimates indicate that the latest Paleocene, and PETM are +12.3°C and +17.6°C warmer than pre-industrial, respectively. Our results indicate that GMST likely increased by ~4 to 6°C between the latest Paleocene and PETM (66% confidence), in keeping with previous estimates (Frieling et al., 2019; Dunkley Jones, 2013). During the EECO, the average GMST estimate is 27.0°C and likely ranges between 23.2 and 29.7°C (66% confidence interval; Table 4; Figure 7). Assuming a preindustrial GMST of 14°C, our average GMST estimate indicates that the EECO is +13.0°C warmer than pre-industrial. The GMST anomaly for the EECO is ~2°C lower than previous studies (~15°C warmer than pre-industrial; Caballero and Huber, 2013; Zhu et al., 2019) but falls within the range quoted previously in the IPCC AR5 (9 to 14°C warmer than pre-industrial). The EECO is approximately 4 to 5°C colder than the PETM (66% confidence). This is larger than previously suggested (~3°C; Zhu et al., 2019) and may related to a cold bias in EECO GMST estimates (see Section 3.1).

3.4. Equilibrium climate sensitivity during the latest Palaeocene, PETM and EECO

Equilibrium climate sensitivity (ECS) can be defined as the equilibrium change in global near surface air temperature, resulting from a doubling in atmospheric CO_2. Various "flavours" of ECS exist, some of which specifically exclude various feedback processes not always included in climate models, such as those associated with ice sheets, vegetation, or aerosols (Rohling et al., 2012). ECS may also be state-dependent (Caballero and Huber, 2013) and there is no reason to expect that it has not changed with time or as a function of climate state (Farnsworth et al, 2019; Zhu et al., 2020). Therefore, direct comparison of ECS in the past to modern conditions is a fraught enterprise. For our purposes we define a "bulk" ECS (ECS_{bulk}) as being a gross estimate of ECS, between our three intervals and preindustrial. i.e.

$$ECS_{bulk} = (\Delta T_{CO2\text{-}vs\text{-}PI}) / (\Delta F_{CO2\text{-}vs\text{-}PI}) \quad [6]$$

where $\Delta T_{CO2\text{-}vs\text{-}PI}$ is the temperature difference between pre-industrial and the time period of interest that can be attributed to CO_2 forcing, and $\Delta F_{CO2\text{-}vs\text{-}PI}$ is the CO_2 forcing relative to preindustrial. The result is then normalised to a CO_2 forcing equal to a doubling of CO_2. Such calculations have been performed previously (e.g. Anagnostou et al., 2016) and they provide some constraint on the range of climate sensitivity values that are relevant for near-modern prediction (Rohling et al., 2012). For example, Anagnostou et al. (2016) indicated that early Eocene ECS (excluding ice sheet feedbacks) falls within the range 2.1–4.6 °C per CO_2 doubling with maximum probability for the EECO of 3.8 °C. These values (2.1–4.6 °C per CO_2 doubling) are similar to the IPCC ECS range (1.5–4.5 °C at 66% confidence). Here we calculate bulk ECS estimates using the change in GMST and CO_2 in the latest Paleocene, PETM and EECO intervals with reference to the pre-industrial. Following the approach of Anagnostou et al. (2016) and using the forcing equation of Byrne and Goldblatt (2014), we first determine the relative change in climate forcing relative to pre-industrial ($\Delta F_{CO2\text{-}vs\text{-}PI}$):

$$\Delta F_{CO2\text{-}vs\text{-}PI} = 5.32\ln(C_t/C_{PI}) + (0.39[\ln(C_t/\ C_{PI})]^2 \quad [7]$$

where C_{PI} is the atmospheric CO_2 concentration during pre-industrial (278 ppm) and C_t refers to the CO_2 reconstruction at a particular time in the Eocene. The mean proxy estimate of CO_2 for the PETM is ~2200 ppmv (+1904/-699 ppmv; 95% CI) (Gutjahr et al., 2017). The mean proxy estimate of CO_2 for the LP is ~870 ppmv (Gutjahr et al., 2017). The uncertainty of latest Paleocene CO_2 represents two standard deviations of pre-PETM CO_2 (Gutjahr et al. 2017), equal to ±400 ppm. The mean proxy estimate of CO_2 for the EECO is ~1625 ppmv (±750 ppmv; 95% CI) (Anagnostou et al., 2016; Hollis et al., 2019). To calculate bulk ECS, we then use radiative forcing from a doubling of CO_2 from Byrne and Goldblatt (2014) to translate CO_2 into forcing relative to preindustrial (ΔF_{CO2}):

$$ECS = (\Delta T_{CO2\text{-}vs\text{-}PI})\ /\Delta F_{CO2\text{-}vs\text{-}PI} * 3.875 \quad [8]$$

, where GMST (ΔT) distributions are based on output generated via our Monte Carlo simulations (see Section 3.3). Some of the temperature anomaly of the latest Paleocene, PETM, and EECO is caused not by CO_2 but by the different paleotopography, paleobathymetry, and solar constant compared with preindustrial. Furthermore, we choose here to calculate an ECS that explicitly excludes feedbacks associated with vegetation, ice sheets, and aerosols, i.e. $S_{[CO2,LI,VG,AE]}$ in the nomenclature of Rohling et al (2012). To account for these effects, we subtract a value of 4.5°C (Caballero and Huber, 2013; Zhu et al. 2019) from GMST; i.e.

$$\Delta T_{CO2\text{-}vs\text{-}PI} = \Delta GMST - 4.5°C \qquad [9]$$

Following Anagnostou et al. (2016), the uncertainty on the slow-feedback correction on ΔGMST follows a uniform 'flat' probability (±1.5°C). This value of 4.5°C is based upon a comparison of preindustrial and Eocene simulations (both 1x CO_2) conducted with CESM1.2 (Zhu et al., 2019), which incorporates the paleogeographic, solar constant, ice sheet, vegetation, aerosol, and ice sheet changes from preindustrial to Eocene. Our value is similar to previous studies which attribute ~4 to 6°C to the non-CO_2 and non-aerosol forcings and feedbacks (Anagnostou et al., 2016; Caballero and Huber, 2013, Lunt et al., 2012). However, the sensitivity to these Eocene boundary conditions is likely model-dependant and this value will likely differ between model simulations. The uncertainties in our estimated ECS are the products of 10,000 realizations of the latest Paleocene, PETM and EECO CO_2 values and the respective ΔGMST estimate (the mean estimate and propagated uncertainty) based on randomly sampling each variable within its 66% and 90% confidence interval uncertainty envelope

$S_{[CO2,LI,VG,AE]}$ values (66% confidence) for the EECO and PETM average 0.80 (0.46 to 1.15) and 0.92 (0.60 to 1.20), respectively. This yields ECS estimates (66% confidence) for the EECO and PETM compared to modern which average 3.1°C (1.8 to 4.4°C) and 3.6°C (2.3 to 4.7°C), respectively (Figure 8). These are broadly comparable to previous estimates from the early Eocene which account for paleogeography and other feedbacks (~2.1 to 4.6°C; Anagnostou et al., 2016) They are also similar to those predicted by the IPCC (1.5 to 4.5°C per doubling CO_2). ~~care must be exercised when relating geological estimates to modern climate predictions (e.g. Rohling et al., 2012).~~ $S_{[CO2,LI,VG,AE]}$ values (66% confidence) during the latest Paleocene average 1.16 (0.61 to 1.75), which is somewhat higher than the other DeepMIP intervals. This yields ECS estimates (66% confidence) for the latest Paleocene which average 4.5°C (2.4 to 6.8°C) (Figure 8). Higher ECS values are attributed to relatively high GMST estimates (~26°C) and relatively low CO_2 values (~870ppm) during the latest Paleocene. As latest Paleocene CO_2 estimates remain highly uncertain (Gutjahr et al., 2017; see above), new high-fidelity CO_2 records are required to accurately constrain ECS during this time.

ECS may be strongly state-dependant and model simulations indicate a non-linear increase in ECS at higher temperatures (Caballero and Huber, 2013; Zhu et al., 2019) due to changes in cloud feedbacks (Abbot et al., 2009; Caballero and Huber, 2010; Arnold et al., 2012; Zhu et al., 2019). This implies caution when relating geological estimates to modern climate predictions (e.g. Rohling et al., 2012; Zhu et al., 2020) and it may be more appropriate to calculate ECS between different time intervals (e.g. latest Paleocene to PETM; Shaffer et al., 2016). To this end, we also calculate ECS 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. 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.

4. Conclusions

Using six different methods, we have quantified global mean surface temperatures (GMST) during the latest Paleocene, PETM and EECO. GMST was calculated within a coordinated, experimental framework and utilised six methodologies including three different input datasets. After evaluating the impact of different proxy datasets upon GMST estimates, we combined all six methodologies to derive an average GMST value during the latest Paleocene, PETM and EECO. We show that the 'average' GMST estimate (66% confidence) during the latest Paleocene, PETM and EECO is 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. Assuming a preindustrial GMST of 14°C, the latest Paleocene, PETM and EECO are 12.3°C, 17.6°C and 13.0°C higher than modern, respectively. Using our 'combined' GMST estimate, we demonstrate that "bulk" ECS (66%

confidence) during the latest Paleocene, PETM and EECO is 4.5°C (2.4 to 6.8°C), 3.6°C (2.3 to 4.7°C) and 3.1°C (1.8 to 4.4°C) per doubling of CO_2. Taken together, our study improves our characterisation of the global mean temperature of these key time intervals, allowing future climate change to be put into the context of past changes, and allowing us to provide a refined estimate of ECS.

Data availability

Data can be accessed via the online supporting information, via www.pangaea.de/, or from the author (email: gordon.inglis@soton.ac.uk).

Authorship tiers and contributions

Authorship of this manuscript is organized into three tiers according to the contributions of each individual author. Inglis (Tier I) organized the structure and writing of the manuscript, contributed to all sections of the text and designed the figures. Tier II authors (listed alphabetically following Inglis) assumed a leading role by contributing methodologies used in the text. Tier III authors (listed alphabetically following Wilkinson) contributed intellectually to the text and figure design.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank two anonymous reviewers whose thoughtful comments significantly improved the manuscript. This research was funded from NERC through NE/P01903X/1 and NE/N006828/1, both of which supported GNI, DJL, SS and RDP. The authors also thank Chris

Poulsen and Jiang Zhu for assistance with the CESM1.2 model simulations. GNI was supported by a GCRF Royal Society Dorothy Hodgkin Fellowship. N.J.B. is supported by NSF AGS-1844380. FB, DL, and RDW were funded by the EPSRC 'Past Earth Network'. MH was funded by NSF OPP 1842059. TDJ, KME and GLF were supported by NERC grant NE/P013112/1. AdB and DKH acknowledges support from the Swedish Research Council Project 2016-03912. GFDL numerical simulations were performed using resources provided by the Swedish National Infrastructure for Computing (SNIC) at NSC, Linköping. DKH was also supported by FORMAS project 2018-01621.

References

Abbot, D.S., Huber, M., Bousquet, G. and Walker, C.C.: High-CO2 cloud radiative forcing feedback over both land and ocean in a global climate model, Geophysical Research Letters, 36, 2009.

Arnold, N.P., Tziperman, E. and Farrell, B.: Abrupt transition to strong superrotation driven by equatorial wave resonance in an idealized GCM, Journal of the Atmospheric Sciences, 69, 626-640, 2012.

Anagnostou, E., John, E. H., Edgar, K. M., Foster, G. L., Ridgwell, A., Inglis, G. N., Pancost, R. D., Lunt, D. J., and Pearson, P. N.: Changing atmospheric CO2 concentration was the primary driver of early Cenozoic climate, Nature, 533, 380-384, 10.1038/nature17423, 2016.

Barnet, J. S., Littler, K., Westerhold, T., Kroon, D., Leng, M. J., Bailey, I., Röhl, U., and Zachos, J. C.: A high-Fidelity benthic stable isotope record of late Cretaceous–early Eocene climate change and carbon-cycling, Paleoceanography & Paleoclimatology, 34, 672-691, 2019.

Bemis, B. E., Spero, H. J., Bijma, J., and Lea, D. W.: Reevaluation of the oxygen isotopic composition of planktonic foraminifera: Experimental results and revised paleotemperature equations, Paleoceanography & Paleoclimatology, 13, 150-160, 10.1029/98pa00070, 1998.

Bijl, P. K., Schouten, S., Sluijs, A., Reichart, G.-J., Zachos, J. C., and Brinkhuis, H.: Early Palaeogene temperature evolution of the southwest Pacific Ocean, Nature, 461, 776-779, 2009.

Bragg, F. J., Paine, P., Saul, A., Lunt, D. J., Wilkinson, R., and Zammit-Mangion, A.: A Statistical Algorithm for Evaluating Palaeoclimate Simulations Against Geological Observations, Geoscientific Model Development, In preperation.

Byrne, B., and Goldblatt, C.: Radiative forcing at high concentrations of well-mixed greenhouse gases, Geophysical Research Letters, 41, 152-160, 2014.

Caballero, R. and Huber, M.: Spontaneous transition to superrotation in warm climates simulated by CAM3, Geophysical Research Letters, 37, 2010

Caballero, R., and Huber, M.: State-dependent climate sensitivity in past warm climates and its implications for future climate projections, Proceedings of the National Academy of Sciences, 110, 14162-14167, 2013.

Cramer, B. S., Toggweiler, J. R., Wright, J. D., Katz, M. E., and Miller, K. G.: Ocean overturning since the Late Cretaceous: Inferences from a new benthic foraminiferal isotope compilation, Paleoceanography & Paleoclimatology, 24, 10.1029/2008pa001683, 2009.

Cramwinckel, M. J., Huber, M., Kocken, I. J., Agnini, C., Bijl, P. K., Bohaty, S. M., Frieling, J., Goldner, A., Hilgen, F. J., Kip, E. L., Peterse, F., van der Ploeg, R., Rohl, U., Schouten, S., and Sluijs, A.: Synchronous tropical and polar temperature evolution in the Eocene, Nature, 559, 382, 2018.

Creech, J.B., Baker, J.A., Hollis, C.J., Morgans, H.E. and Smith, E.G.: Eocene sea temperatures for the mid-latitude southwest Pacific from Mg/Ca ratios in planktonic and benthic foraminifera, Earth and Planetary Science Letters, 299, 483-495, 2010.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de

Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society 137, 553-597, 2011.

Dunkley Jones, T., Lunt, D.J., Schmidt, D.N., Ridgwell, A., Sluijs, A., Valdes, P.J. and Maslin, M.: Climate model and proxy data constraints on ocean warming across the Paleocene–Eocene Thermal Maximum, Earth Science Reviews, 125, 123-145, 2013.

Dworkin, S.I., Nordt, L. and Atchley, S.: Determining terrestrial paleotemperatures using the oxygen isotopic composition of pedogenic carbonate, Earth and Planetary Science Letters, 237,56-68, 2005.

Edgar, K.M., Anagnostou, E., Pearson, P.N. and Foster, G.L: Assessing the impact of diagenesis on δ11B, δ13C, δ18O, Sr/Ca and B/Ca values in fossil planktic foraminiferal calcite, Geochimica et Cosmochimica Acta, 166, 89-209, 2015.

Edgar, K.M., Pälike, H. and Wilson, P.A.: Testing the impact of diagenesis on the δ18O and δ13C of benthic foraminiferal calcite from a sediment burial depth transect in the equatorial Pacific, Paleoceanography, 28, 468-480. 2013.

Eley, Y., Thompson, W., Greene, S.E., Mandel, I., Edgar, K., Bendle, J.A and Dunkley Jones, T.: OPTiMAL: A new machine learning approach for GDGT-based palaeothermometry, Climate of the Past Discussions, 2019.

Evans, D., Sagoo, N., Renema, W., Cotton, L. J., Müller, W., Todd, J. A., Saraswati, P. K., Stassen, P., Ziegler, M., Pearson, P. N., Valdes, P. J., and Affek, H. P.: Eocene greenhouse climate revealed by coupled clumped isotope-Mg/Ca thermometry, Proceedings of the National Academy of Sciences, 115, 1174-1179, 10.1073/pnas.1714744115 %J Proceedings of the National Academy of Sciences, 2018.

Farnsworth, A., Lunt, D., O'Brien, C., Foster, G., Inglis, G., Markwick, P., Pancost, R., and Robinson, S.: Climate sensitivity on geological timescales controlled by non-linear feedbacks and ocean circulation, Geophysical Research Letters, 2019.

Gallagher, T. M., and Sheldon, N. D.: A new paleothermometer for forest paleosols and its implications for Cenozoic climate, Geology, 41, 647-650, 10.1130/G34074.1, 2013.

Goldner, A., Herold, N. and Huber, M.: Antarctic glaciation caused ocean circulation changes at the Eocene–Oligocene transition, Nature, 511,574-577, 2014.

Gutjahr, M., Ridgwell, A., Sexton, P. F., Anagnostou, E., Pearson, P. N., Pälike, H., Norris, R. D., Thomas, E., and Foster, G. L.: Very large release of mostly volcanic carbon during the Palaeocene–Eocene Thermal Maximum, Nature, 548, 573-577, 10.1038/nature23646, 2017.

Hansen, J., Sato, M., Russell, G., and Kharecha, P.: Climate sensitivity, sea level and atmospheric carbon dioxide, Philosophical Transactions of the Royal Society A: Mathematical, Physical Engineering Sciences, 371, 20120294, 2013.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J., Nicholls, N., Penner, J. E., and Stott, P. A.: Understanding and attributing climate change, IPCC, 2007: Climate Change 2007: the physical science basis. contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, 2007.

Herold, N., Buzan, J., Seton, M., Goldner, A., Green, J. A. M., Müller, R. D., Markwick, P., and Huber, M.: A suite of early Eocene (~ 55 Ma) climate model boundary conditions, Geoscientific Model Development, 7, 2077-2090, 10.5194/gmd-7-2077-2014, 2014.

Hollis, C. J., Taylor, K. W. R., Handley, L., Pancost, R. D., Huber, M., Creech, J. B., Hines, B. R., Crouch, E. M., Morgans, H. E. G., Crampton, J. S., Gibbs, S., Pearson, P. N., and Zachos, J. C.: Early Paleogene temperature history of the Southwest Pacific Ocean: Reconciling proxies and models, Earth and Planetary Science Letters, 349–350, 53-66, http://dx.doi.org/10.1016/j.epsl.2012.06.024, 2012.

Hollis, C. J., Dunkley Jones, T., Anagnostou, E., Bijl, P. K., Cramwinckel, M. J., Cui, Y., Dickens, G. R., Edgar, K. M., Eley, Y., Evans, D., Foster, G. L., Frieling, J., Inglis, G. N., Kennedy, E. M., Kozdon, R., Lauretano, V., Lear, C. H., Littler, K., Lourens, L., Meckler, A. N., Naafs, B. D. A., Pälike, H., Pancost, R. D., Pearson, P. N., Röhl, U., Royer, D. L., Salzmann, U., Schubert, B. A., Seebeck, H., Sluijs, A., Speijer, R. P., Stassen, P., Tierney, J., Tripati, A., Wade, B., Westerhold, T., Witkowski, C., Zachos, J. C., Zhang, Y. G., Huber, M., and Lunt, D. J.: The DeepMIP contribution to PMIP4: methodologies for selection, compilation and analysis of latest Paleocene and early Eocene climate proxy data, incorporating version 0.1 of the DeepMIP database, Geoscientific Model Development, 12, 3149-3206, 10.5194/gmd-12-3149-2019, 2019.

Huber, M., and Caballero, R.: The early Eocene equable climate problem revisited, Clim. Past, 7, 603-633, 10.5194/cp-7-603-2011, 2011.

Hutchinson, D. K., de Boer, A. M., Coxall, H. K., Caballero, R., Nilsson, J. and Baatsen, M.: Climate sensitivity and meridional overturning circulation in the late Eocene using GFDL CM2.1, *Climate of the Past,* 14**,** 789-810, 2018.

Hyland, E. G., Sheldon, N. D., and Cotton, J. M.: Constraining the early Eocene climatic optimum: A terrestrial interhemispheric comparison, GSA Bulletin, 129, 244-252, 10.1130/B31493.1, 2017.

Hyland, E.G. and Sheldon, N.D.: Coupled CO2-climate response during the early Eocene climatic optimum, *P*alaeogeography, Palaeoclimatology, Palaeoecology, 369, 125-135, 2013.

Inglis, G. N., Farnsworth, A., Lunt, D., Foster, G. L., Hollis, C. J., Pagani, M., Jardine, P. E., Pearson, P. N., Markwick, P., Galsworthy, A. M. J., Raynham, L., Taylor, K. W. R., and Pancost, R. D.: Descent toward the Icehouse: Eocene sea surface cooling inferred from GDGT distributions, Paleoceanography, 30, 1000-1020, 10.1002/2014pa002723, 2015.

Inglis, G. N., Collinson, M. E., Riegel, W., Wilde, V., Farnsworth, A., Lunt, D. J., Valdes, P., Robson, B. E., Scott, A. C., Lenz, O. K., Naafs, B. D. A., and Pancost, R. D.: Mid-latitude continental temperatures through the early Eocene in western Europe, Earth and Planetary Science Letters, 460, 86-96, https://doi.org/10.1016/j.epsl.2016.12.009, 2017.

Keating-Bitonti, C. R., Ivany, L. C., Affek, H. P., Douglas, P., and Samson, S. D.: Warm, not super-hot, temperatures in the early Eocene subtropics, Geology, 39, 771-774, 2011.

Knutti, R., Rugenstein, M. A., and Hegerl, G. C.: Beyond equilibrium climate sensitivity, Nature Geoscience, 10, 727-736, 2017.

Lauretano, V., K. Littler, M. Polling, J. C. Zachos, and L. J. Lourens.: Frequency, magnitude and character of hyperthermal events at the onset of the Early Eocene Climatic Optimum, Climate of the Past, 11, 1313-132, 2015.

Littler, K., Röhl, U., Westerhold, T. and Zachos, J.C.: A high-resolution benthic stable-isotope record for the South Atlantic: Implications for orbital-scale changes in Late Paleocene–Early Eocene climate and carbon cycling, Earth and Planetary Science Letters, 401, 18-30, 2014.

Lunt, D. J., Jones, T. D., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., Loptson, C., Marotzke, J., Roberts, C., and Tindall, J.: A model-data comparison for a multi-model ensemble of early Eocene atmosphere-ocean simulations: EoMIP, Climate of the Past, 8, 2012.

Lunt, D. J., Farnsworth, A., Loptson, C., Foster, G. L., Markwick, P., O'Brien, C. L., Pancost, R. D., Robinson, S. A., and Wrobel, N.: Palaeogeographic controls on climate and proxy interpretation, Climate of the Past, 12, 1181-1198, 2016.

Lunt, D. J., Huber, M., Anagnostou, E., Baatsen, M. L. J., Caballero, R., DeConto, R., Dijkstra, H. A., Donnadieu, Y., Evans, D., Feng, R., Foster, G. L., Gasson, E., von der Heydt, A. S., Hollis, C. J., Inglis, G. N., Jones, S. M., Kiehl, J., Kirtland Turner, S., Korty, R. L., Kozdon, R., Krishnan, S., Ladant, J. B., Langebroek, P., Lear, C. H., LeGrande, A.

N., Littler, K., Markwick, P., Otto-Bliesner, B., Pearson, P., Poulsen, C. J., Salzmann, U., Shields, C., Snell, K., Stärz, M., Super, J., Tabor, C., Tierney, J. E., Tourte, G. J. L., Tripati, A., Upchurch, G. R., Wade, B. S., Wing, S. L., Winguth, A. M. E., Wright, N. M., Zachos, J. C., and Zeebe, R. E.: The DeepMIP contribution to PMIP4: experimental design for model simulations of the EECO, PETM, and pre-PETM (version 1.0), Geoscientific Model Development, 10, 889-901, 10.5194/gmd-10-889-2017, 2017.

Lunt, D.J., Bragg, F., Chan, W.L., Hutchinson, D.K., Ladant, J.B., Niezgodzki, I., Steinig, S., Zhang, Z., Zhu, J., Abe-Ouchi, A., de Boer, A.M., Coxall, H.K., Donnadieu, Y., Knorr, G., Langebroek, P.M., Lohmann, G., Poulsen, C.J., Sepulchre, P., Tierney, J., Valdes, P.J., Dunkley Jones, T., Hollis, C.J., Huber, M. and Otto-Bliesner, B.L.: DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data, Climate of the Past Discussions, 1-20, 2020.

Marchitto, T., Curry, W., Lynch-Stieglitz, J., Bryan, S., Cobb, K., and Lund, D.: Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera, Geochimica et Cosmochimica Acta, 130, 1-11, 2014.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez Rouco, J. F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from Paleoclimate Archives, in: Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 383-464, 2014.

Müller, R.D., Sdrolias, M., Gaina, C., Steinberger, B. and Heine, C.: Long-term sea-level fluctuations driven by ocean basin dynamics, Science, 319, 1357-1362, 2008.

Pearson, P. N., Ditchfield, P. W., Singano, J., Harcourt-Brown, K. G., Nicholas, C. J., Olsson, R. K., Shackleton, N. J., and Hall, M. A.: Warm tropical sea surface temperatures in the Late Cretaceous and Eocene epochs, Nature, 413, 481-487, 2001.

Pearson, P. N., van Dongen, B. E., Nicholas, C. J., Pancost, R. D., Schouten, S., Singano, J. M., and Wade, B. S.: Stable warm tropical climate through the Eocene Epoch, Geology, 35, 211-214, 10.1130/g23175a.1, 2007.

Rohling, E. J., Sluijs, A., Dijkstra, H. A., Köhler, P., van de Wal, R. S. W., von der Heydt, A. S., Beerling, D. J., Berger, A., Bijl, P. K., Crucifix, M., DeConto, R., Drijfhout, S. S., Fedorov, A., Foster, G. L., Ganopolski, A., Hansen, J., Hönisch, B., Hooghiemstra, H., Huber, M., Huybers, P., Knutti, R., Lea, D. W., Lourens, L. J., Lunt, D., Masson-Delmotte, V., Medina-Elizalde, M., Otto-Bliesner, B., Pagani, M., Pälike, H., Renssen, H., Royer, D. L., Siddall, M., Valdes, P., Zachos, J. C., Zeebe, R. E., and Members, P. P.: Making sense of palaeoclimate sensitivity, Nature, 491, 683-691, 10.1038/nature11574, 2012.

Sexton, P.F., Norris, R.D., Wilson, P.A., Pälike, H., Westerhold, T., Röhl, U., Bolton, C.T. and Gibbs, S.: Eocene global warming events driven by ventilation of oceanic dissolved organic carbon, Nature, 471, 349-352, 2011.

Sexton, P.F., Wilson, P.A. and Pearson, P.N.: Microstructural and geochemical perspectives on planktic foraminiferal preservation: “Glassy” versus “Frosty”, Geochemistry, Geophysics, Geosystems, 7, 2006.

Sijp, W.P., England, M.H. and Huber, M.: Effect of the deepening of the Tasman Gateway on the global ocean, Paleoceanography, 26, 2011.

Shaffer, G., Huber, M., Rondanelli, R., and Pedersen, J. O. P. J. G. R. L.: Deep time evidence for climate sensitivity increase with warming, 43, 6538-6545, 2016.

Sheldon, Nathan D., Retallack, Gregory J., and Tanaka, S.: Geochemical Climofunctions from North American Soils and Application to Paleosols across the Eocene-Oligocene Boundary in Oregon, The Journal of Geology, 110, 687-696, 10.1086/342865, 2002.

Sheldon, N.D.: Non-marine records of climatic change across the Eocene-Oligocene transition, The Late Eocene Earth—Hothouse, Icehouse, and Impacts: Geological Society of America Special Paper, 452, 241-248, 2009.

Stevens, B., Sherwood, S. C., Bony, S., and Webb, M. J.: Prospects for narrowing bounds on Earth's equilibrium climate sensitivity, Earth's Future, 4, 512-522, 2016.

Tierney, J. E., and Tingley, M. P.: A Bayesian, spatially-varying calibration model for the TEX86 proxy, Geochimica et Cosmochimica Acta, 127, 83-106, 2014.

Westerhold, T., Röhl, U., Donner, B., and Zachos, J. C.: Global extent of early Eocene hyperthermal events: A new Pacific benthic foraminiferal isotope record from Shatsky Rise (ODP Site 1209), Paleoceanography & Paleoclimatology, 33, 626-642, 2018.

von der Heydt, A.S., Dijkstra, H.A., van de Wal, R.S., Caballero, R., Crucifix, M., Foster, G.L., Huber, M., Köhler, P., Rohling, E., Valdes, P.J. and Ashwin, P.: Lessons on climate sensitivity from past climate changes, Current Climate Change Reports, *2*, 148-158, 2016.

Zachos, J.C., Dickens, G.R. and Zeebe, R.E.; An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics, Nature, 451, 279-283, 2008.

Zhu, J., Poulsen, C. J., and Tierney, J. E.: Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks, Science Advances, 5, 2019.

Zhu, J., Poulsen, C.J. and Otto-Bliesner, B.L.: High climate sensitivity in CMIP6 model not supported by paleoclimate, Nature Climate Change, 10, 378-379, 2020,

Label in Fig. 1	Source	Time window	GMST (°C)	Uncertainty (°C)	Proxy system
1a	Farnsworth et al. (2019)	EE	23.4	±3.2	$\delta^{18}O$ planktonic
1b	Farnsworth et al. (2019)	EE	37.1	±1.4	$\delta^{18}O$ planktonic + TEX_{86}
2a	Zhu et al. (2019)	LP	27	n/a	Multiple
2b	Zhu et al. (2019)	EECO	29	±3	Multiple
2c	Zhu et al. (2019)	PETM	32	n/a	Multiple
3	Caballero and Huber (2013)	EE	29.5	±2.6	Multiple
4	Hansen et al (2013)	EE	28	n/a	$\delta^{18}O$ benthic
5	Cramwinckel et al. (2018)	EE	29.3	n/a	Multiple

Table 1: Previous studies that have determined GMST for the early Eocene (EE), EECO, PETM or latest Paleocene (LP). n/a indicates that no error bars were reported in the original publications.

Experiment	Description
D_{surf}-*Baseline*	All SST and LAT data compiled in Hollis et al. (2019) but excluding recrystallized planktonic foraminifera $\delta^{18}O$ values
D_{surf}-*Frosty*	D_{surf}-*baseline* but including recrystallized planktonic foraminifera $\delta^{18}O$ values
D_{surf}-*NoTEX*	D_{surf}-*baseline* but excluding TEX_{86} values
D_{surf}-*NoMBT*	D_{surf}-*baseline* but excluding MBT(')/CBT values from marine sediments
D_{surf}-*NoPaleosol*	D_{surf}-*baseline* but excluding mammal/paleosol $\delta^{18}O$ values and paleosol climofunctions

Table 2: Baseline and optional subsampling experiments applied to D_{surf}

	GMST (°C)					
	D_{surf}-1	D_{surf}-2	D_{surf}-3	D_{surf}-4	D_{deep}-1	D_{comb}-1
LP	26.6 (±1.3)	26.8 (±6.9)	27.6 (±1.5)	26.8 (±1.3)	25.8 (±1.4)	21.6 (±1.2)
PETM	33.9 (±1.4)	33.4 (±10.3)	32.6 (±1.5)	30.7 (±1.6)	31.1 (±2.9)	26.6 (±2.1)
EECO	27.2 (±0.7)	26.7 (±8.9)	29.8 (±1.5)	25.7 (±1.1)	28.0 (±1.3)	22.8 (±1.0)

Table 3: Individual GMST estimates for latest Paleocene (LP), PETM and EECO. Reported GMST estimates utilise 'baseline' experiments except D_{surf}*-1* during the EECO which uses D_{surf}*-NoPaleosol.* GMST estimates are based on the mantle-based reference frame. Error bars on each individual method are the standard deviation (1σ), except D_{surf-1} and D_{surf-2} which use the standard error ($1\sigma_{\bar{x}}$).

	GMST (°C) (Average)	GMST (°C) (66% CI)	GMST (°C) (90% CI)
LP	26.3	22.3 – 28.3	21.3 – 29.1
PETM	31.6	27.3 - 34.5	25.9 – 35.6
EECO	27.0	23.2 – 29.6	22.2 – 30.7

Table 4: 'Combined' GMST estimates (66% and 90% confidence intervals) during the: i) latest Paleocene, ii) PETM, and iii) EECO.

	ECS (°C) (Average)	ECS (°C) (66% confidence)	ECS (°C) (90% confidence)
LP	4.5	2.4 – 6.8	1.6 – 8.0
PETM	3.6	2.3– 4.7	1.9 – 5.2
EECO	3.1	1.8 – 4.4	1.3 – 5.0

Table 5: Estimates of ECS (66% and 90% confidence) during the: i) latest Paleocene, ii) PETM and iii) EECO.

Figure captions:

Figure 1: Published GMST estimates during the early Paleogene (57 to 48 Ma). Dots represent average values. The horizontal limits on the individual dots represent the reported error. *y*-Axis labels refer to previous estimates (see Table 1).

Figure 2: An illustration of Method D_{surf}*-1* during the EECO. (a) Modelled early Eocene temperatures utilising CESM1.2 at 6x pre-industrial CO_2, (b) Interpolated absolute SST reconstructions, (c) Data-model difference between (a) and (b).

~~**Figure 2:** Location of proxies within the surface temperature dataset (D_{surf}). A) SST proxies with time intervals indicated as followed: black circles, all three-time intervals represented. Red circles: PETM ± latest Paleocene intervals; orange circles, EECO interval (b) Terrestrial sites with time intervals indicated as in (a) and green circles, LP only.~~

Figure 3: An illustration of Method D_{surf}*-2* for 2 sites: (a) Big Bend LAT in the EECO as diagnosed using HadCM3L, and (b) DSDP401 SST in the PETM as diagnosed using CCSM3. The vertical dashed line shows $< T >^{inferred}$ and the horizontal dashed line shows T^{proxy}, which intercept at the orange dot. The dark blue dots show the intercept of T^{low} with $< T^{low} >$, and the red dots show the intercept of T^{high} with $< T^{high} >$.

Figure 4: Inferred global mean temperature ($< T >^{inferred}$) using D_{surf}*-2*, for (a) each EECO-aged LAT proxy as diagnosed using HadCM3, and (b) each PETM-aged SST proxy as diagnosed using CCSM3. For (a) and (b), the final estimate of global mean temperature is the average of all the individual sites. The solid line shows the continental outline in each model, and the dashed line shows the continental outline.

Figure 5: Predicted surface warming by Gaussian process regression using D_{surf}-*3* for the (a) latest Paleocene, (b) PETM and (c) EECO. Anomalies are relative to the present-day zonal mean surface temperature. Circles (triangles) indicate all available SST (LAT) proxy data for the respective time slice that were used to train the model. Symbols for locations where multiple proxy reconstructions are available are slightly shifted in latitude for improved visibility.

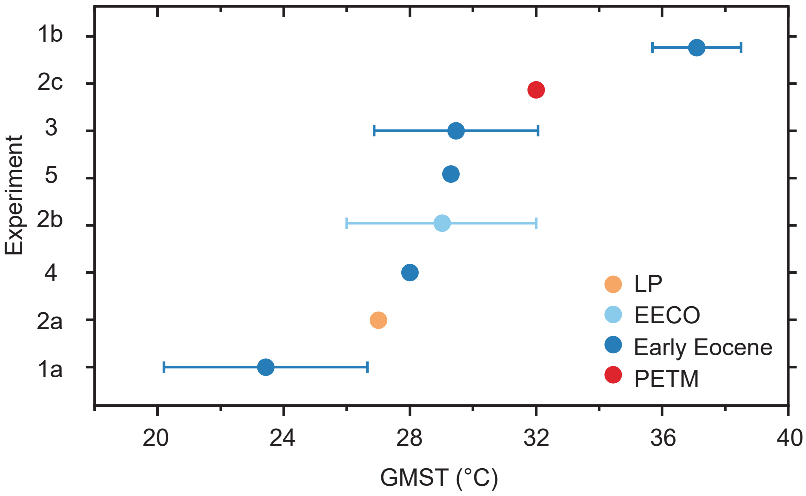
~~**Figure 5:** Predicted surface warming by Gaussian process regression using D_{surf}-*3* for the EECO for the five core experiments (see Table 2). Anomalies are relative to the present-day zonal mean surface temperature. Circles indicate all available SST and LAT proxy data for the respective time slice and experiment that were used to train the model. Circles for locations where multiple proxy reconstructions are available are slightly shifted in latitude for improved visibility.~~

Figure 6: GMST estimates during the (a) PETM, (b) EECO and (c) latest Paleocene for each methodology. GMST estimates utilise 'baseline' experiments except D_{surf}-*1* during the EECO which uses D_{surf}-*NoPaleosol.* GMST estimates are based on the mantle-based reference frame. Error bars on each individual method are the standard deviation (1σ), except $D_{surf\text{-}1}$ and $D_{surf\text{-}2}$ which use the standard error (1σ).

Figure 7: Probability density function of our 'combined' GMST estimate during the DeepMIP intervals with full propagation of errors. GMST estimates are based on the mantle-based reference frame.

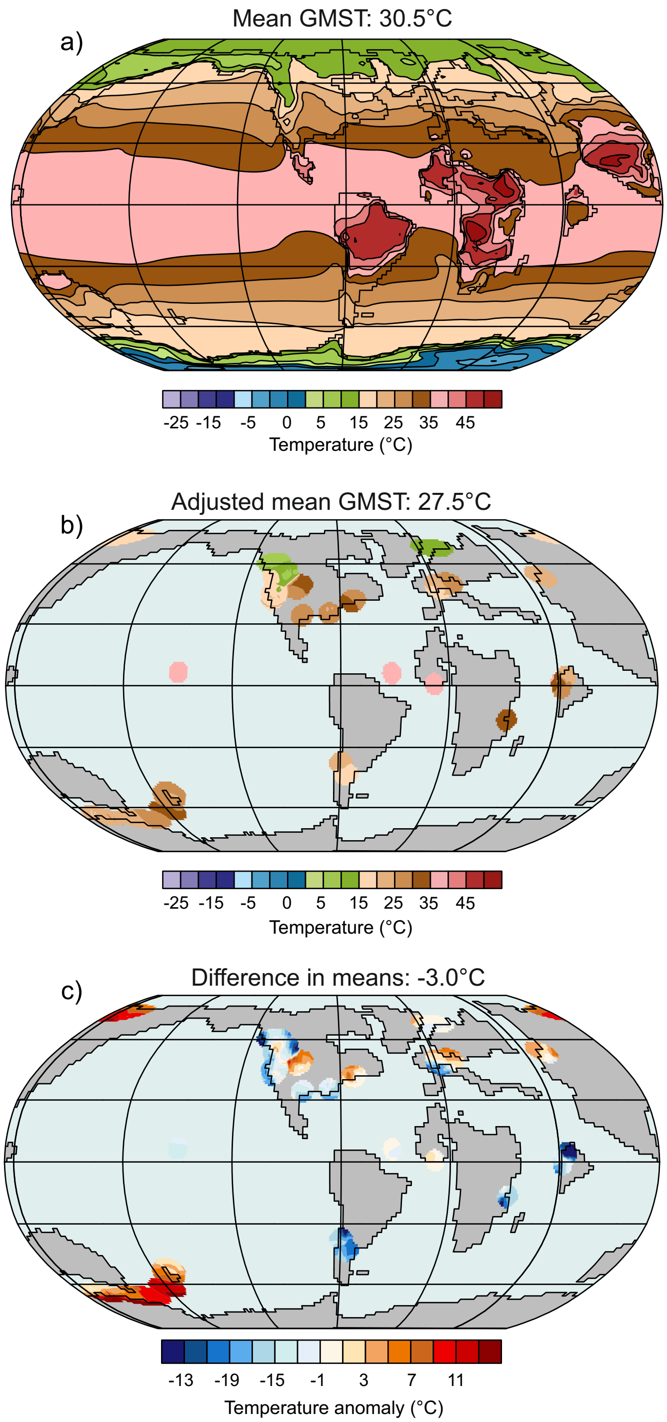
Figure 8: Probability density function of 'bulk' ECS during the latest Paleocene, PETM and EECO that explicitly accounts for non-CO_2 forcings of palaeography and solar constant, and feedbacks associated with land ice, vegetation, and aerosols (Zhu et al., 2019), i.e. $S_{[CO2,LI,VG,AE]}$ in the nomenclature of Rohling et al (2012).

Figure 1



1b
2c
3
5
2b
4
2a
1a
Experiment
LP
EECO
Early Eocene
PETM
20
24
28
32
36
40
GMST (°C)

Figure 2



a)
Mean GMST: 30.5°C
-25 -15 -5 0 5 15 25 35 45
Temperature (°C)
b)
Adjusted mean GMST: 27.5°C
-25 -15 -5 0 5 15 25 35 45
Temperature (°C)
c)
Difference in means: -3.0°C
-13 -19 -15 -1 3 7 11
Temperature anomaly (°C)

Figure 3

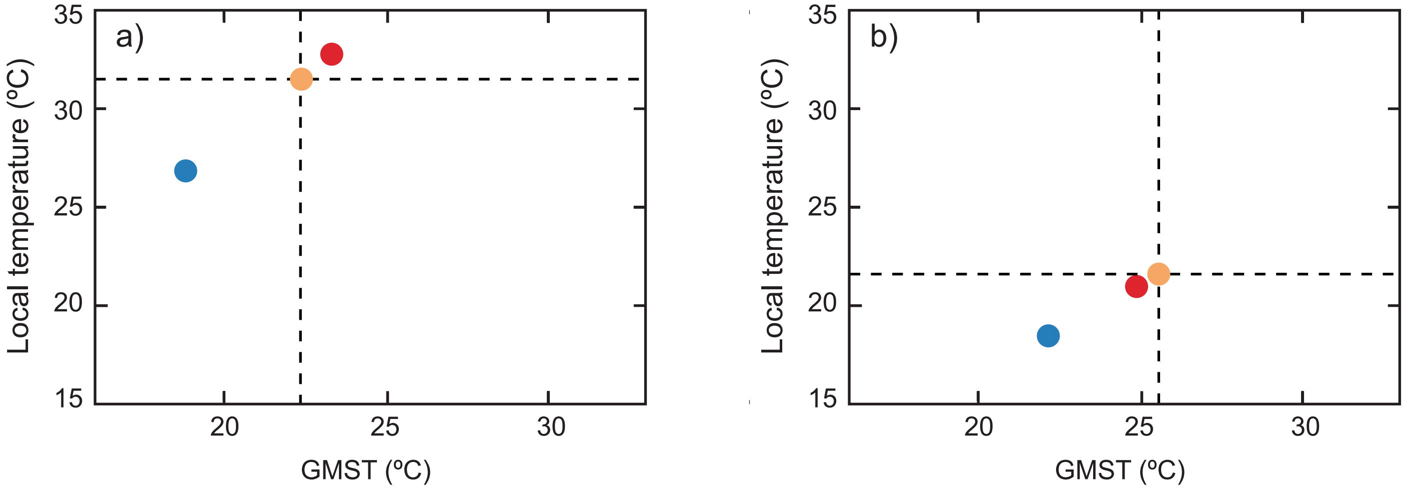


Figure 4

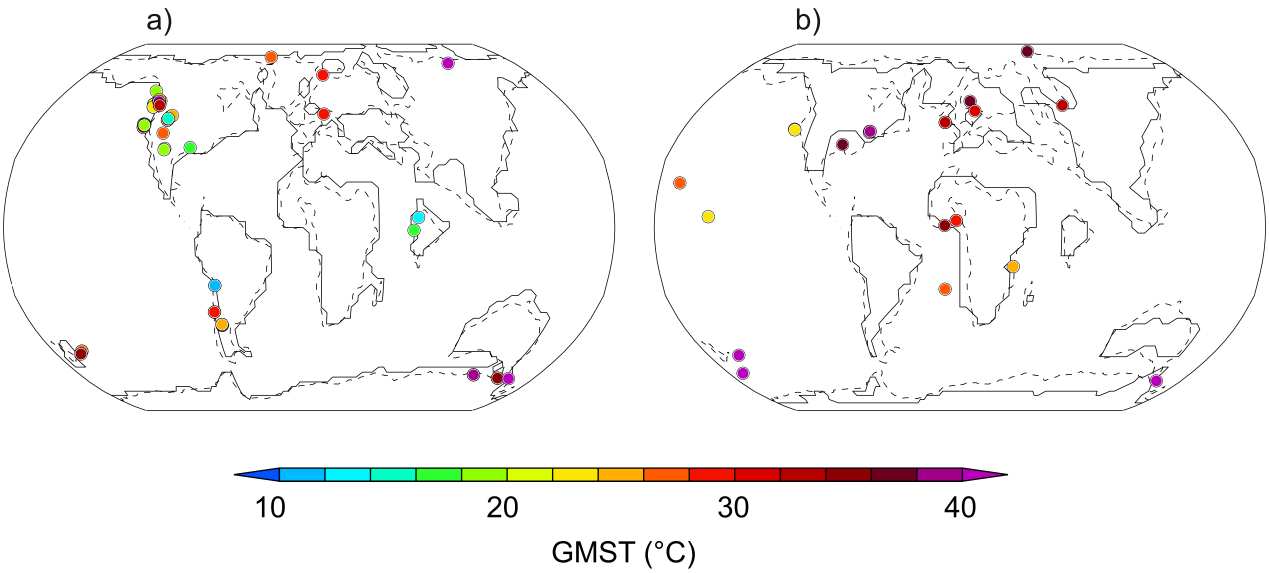


Figure 5

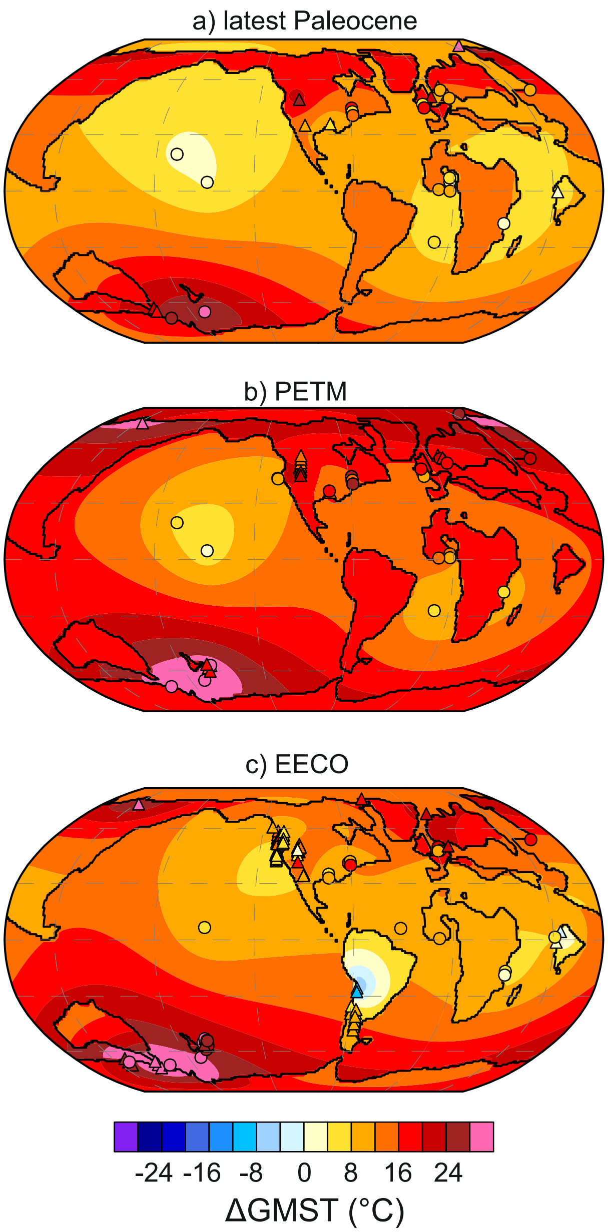
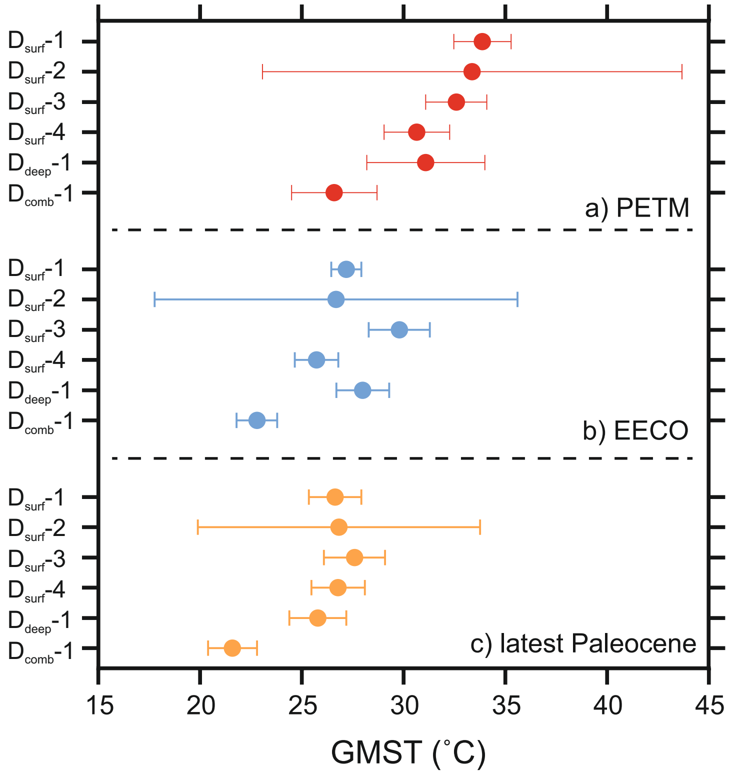
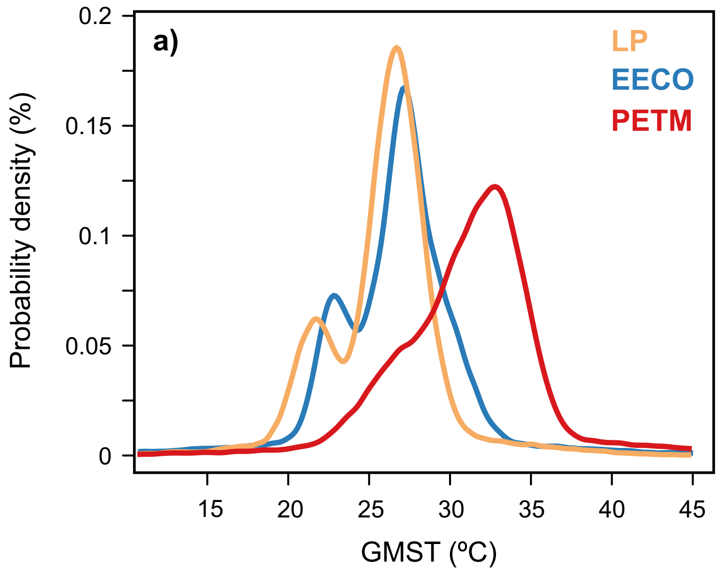


Figure 6



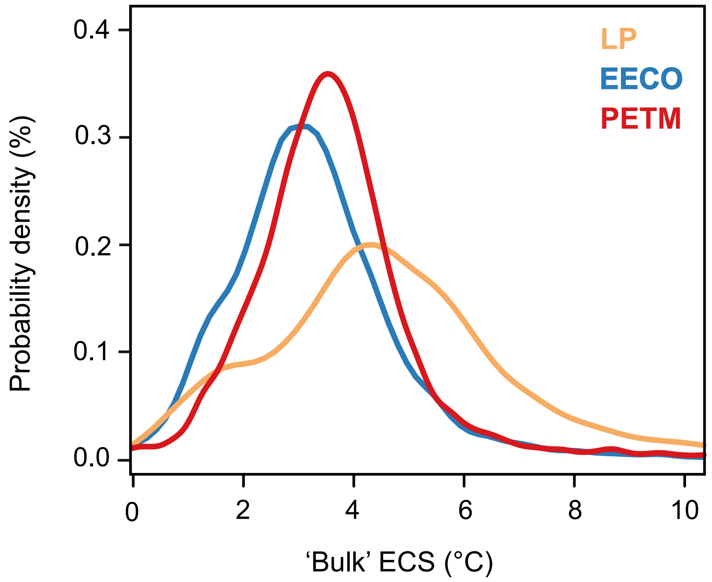
Dsurf-1
Dsurf-2
Dsurf-3
Dsurf-4
Ddeep-1
Dcomb-1
a) PETM
b) EECO
c) latest Paleocene
15
20
25
30
35
40
45
GMST (˚C)

Figure 7



a)
LP
EECO
PETM
Probability density (%)
0.2
0.15
0.1
0.05
0
15
20
25
30
35
40
45
GMST (ºC)

Figure 8



LP
EECO
PETM
Probability density (%)
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
'Bulk' ECS (°C)