Paleoclimate data provide constraints on climate models' large-scale response to past CO_2 changes

Daniel J. Lunt[1*], Bette L. Otto-Bliesner[2], Chris Brierley[3], Alan Haywood[4], Gordon N. Inglis[5], Kenji Izumi[1], Masa Kageyama[6], Darrell Kaufman[7], Thorsten Mauritsen[8], Erin L. McClymont[9], Ulrich Salzmann[10], Sebastian Steinig[1], Jessica E. Tierney[11], Anni Zhao[3], Jiang Zhu[2]

[1*]School of Geographical Sciences, University of Bristol, Bristol, UK.
[2]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research (NCAR), Boulder, USA.
[3]Department of Geography, University College London (UCL), London, UK.
[4]School of Earth and Environment, University of Leeds, Leeds, UK.
[5]School of Ocean and Earth Science, University of Southampton, Southampton, UK.
[6]Laboratoire des Sciences du Climat et de l'Environnement / Institut Pierre-Simon Laplace (LSCE/IPSL), CEA-CNRS-UVSQ, Université Paris Saclay, Gif sur Yvette, France.
[7]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA.
[8]Meteorologiska institutionen (MISU), Stockholm University, Stockholm, Sweden.
[9]Department of Geography, Durham University, Durham, UK.
[10]Geography and Environmental Science, Northumbria University, Newcastle, UK.
[11]Department of Geosciences, University of Arizona, Tucson, USA.

*Corresponding author(s). E-mail(s): d.j.lunt@bristol.ac.uk;

Abstract

The paleoclimate record provides a test-bed in which climate models can be evaluated under conditions of substantial CO_2 change; however, these data are typically under-used in the process of model development and evaluation. Here, we use a set of metrics based on paleoclimate proxy observations to evaluate climate models under three past time periods. We find that the latest CMIP6/PMIP4 ensemble mean does a remarkably good job of simulating the global mean surface air temperatures of these past periods, and is improved on CMIP5/PMIP3, implying that the modern climate sensitivity of the CMIP6/PMIP4 model ensemble mean is consistent with the paleoclimate record. However, some models, in particular those with very high or very low climate sensitivity, simulate paleo temperatures that are outside the uncertainty range of the paleo proxy temperature data; in this regard, the paleo data can provide a more stringent constraint than data from the historical record. There is also consistency between models and data in terms of polar amplification, with amplification increasing with increasing global mean temperature across all three time periods. The work highlights the benefits of using the paleoclimate record in the model development and evaluation cycle, in particular for screening models with too-high or too-low climate sensitivity across a range of CO_2 concentrations.

Keywords: climate modelling, paleoclimate

1 Introduction

Climate models are routinely applied to situations outside of the regimes in which they have been evaluated during their development cycle. For example, in the framework of the Coupled Model Intercomparison Project Phase 6 (CMIP6) and the Intergovernmental Panel on Climate Change (IPCC), models are used to project future climates under CO_2 concentrations substantially higher than those of the recent observational period.

However, there is potential for traditional model evaluation and development to be expanded to utilise proxy data associated with paleoclimate states [e.g. 1–6]. In particular, paleoclimate model simulations test model behaviour under a wide range of forcings, which encompass those expected in the timescale of the next few centuries and beyond [7, 8]. The underlying philosophy is that we would expect to have more confidence in future predictions from a model which has successfully simulated both

past and modern climate states, than future predictions from a model which has only successfully simulated the modern climate state.

Here we focus on three time periods, chosen firstly because they were subject to substantial CO_2 forcing relative to preindustrial, so are of most direct relevance to future projections, and secondly because they have been part of ongoing international modelling efforts in the framework of the Paleoclimate Modelling Intercomparison Project (PMIP) [9], so have simulations available from a variety of different climate models. The time periods are (i) the Last Glacial Maximum (LGM, 21,000 years ago), with a CO_2 concentration of ∼180 ppmv [e.g. 10, compared to ∼280 ppmv prior to industrialisation, and ∼420 ppmv today], and an increase in ice sheet area and volume compared to today, in particular in the Northern Hemisphere [e.g. 11], (ii) interglacial KM5c within the mid Pliocene warm period (MPWP; ∼3.2 million years ago), with a CO_2 concentration of ∼400 ppmv [e.g. 12], and reduced Greenland and Antarctic ice sheets compared with today [e.g. 13], and (iii) the early Eocene climatic optimum (EECO; ∼53.3-49.1 million years ago), with CO_2 concentrations of ∼1500 ppmv [e.g. 14], and no ice sheets. In general, older time periods have fewer locations with proxy data, and greater uncertainty in the proxy data that is available.

When evaluating climate models for the purposes of assessing their ability to project the future, the general approach is to focus on properties of the climate system that are routinely used to quantify the magnitude of future climate change, and which are robust inherent features that persist across a range of climate states [15, 16]. It is also useful to evaluate properties that are determined by the combined effect of multiple components of the climate system (e.g. atmosphere, ocean, cryosphere), so that the integrated effect of the whole system can be assessed. Here, we focus on three large-scale properties: global mean surface temperature, polar amplification, and land-sea warming contrast. Global mean surface temperature (GMST) is the most fundamental metric, and is a key focus of international agreements to limit global mean warming

[e.g. 17]. Changes in GMST are determined by processes throughout the atmosphere, ocean, and land surface; changes in GMST forced by CO_2 alone can be quantified by the Equilibrium Climate Sensitivity [ECS; 18]. Polar amplification is also a key component of the climate system; the Arctic is currently warming at between 2 [19] and 4 [20] times that of the global mean, with associated impacts including sea level rise [21]. Polar amplification is determined by a range of processes [22], including changes in heat transport [23], sea ice/snow feedbacks [24], and lapse-rate feedbacks [25]. Land-sea warming contrast has also been observed over the last 150 years, with 1.6°C warming over land areas compared with 0.9°C warming of SSTs, associated with a 1.1°C GMST warming over the same period [26]. Land-sea warming contrast is associated with changes to the hydrological cycle and atmospheric circulation [e.g. 27, 28], and the thermal contrast between land and ocean plays a role in monsoon circulations [29].

Although these metrics are straightforward to define and quantify in a purely modelling or conceptual framework, estimating them from paleoclimate proxy records is challenging given their sparse distribution and large uncertainties [e.g. 30]. This complicates model-data comparison, and means that quantification of model improvements over time is problematic. Here we make use of assessed GMST estimates from the IPCC [26], and additionally provide site-specific definitions for all the metrics, that are straightforward to apply in a paleo context (see Online Methods, Sections 4.2 and 4.3), and apply the metrics to existing simulations from the fourth and third phase of the Paleoclimate Modelling Intercomparison Project (PMIP4, PMIP3). In doing so we provide a benchmark for paleoclimate model simulations, and assess improvements over time, including in some of the very latest CMIP6 models.

2 Results

The spatial patterns of ensemble-mean (see Online Methods, Section 4.1) modelled surface temperature change (near-surface air temperature and SST) are shown in

Figure 1, along with paleoclimate proxy estimates at the locations for which they are available (see Online Methods, Section 4.2). In general, the sparsity of the proxy data increases further back in time. An exception is the terrestrial MPWP data, which is more sparse than the (earlier) EECO; this is because of the relatively narrow time period that is used in the Pliocene terrestrial reconstruction (a window of 30kyrs in the MPWP [31] compared with 4120kyrs years in the EECO [32]; see discussion in Section 3). Polar amplification (more warming in the polar regions than the tropics under increasing CO_2), and land-sea warming contrast (more warming over land than over ocean under increasing CO_2) are qualitatively apparent for all three time periods. However, in order to quantify these features in proxies and models, and in order to assess model-data comparison, quantitative metrics are required that account for the relative sparsity of the paleo proxy data. Here we define and use two forms of metrics: firstly, 'true' metrics based on the globally-defined fields, and secondly 'site-specific' metrics which are defined according to a particular paleo proxy dataset and calculated according to the locations of the proxies (see Online Methods, Sections 4.2 and 4.3).

2.1 Global Mean Surface Temperature (GMST)

The true GMST metric (${}^{l,p,e}\Delta T^{t}$) is shown in Figure 2, for models and observations (see Online Methods, Sections 4.2 and 4.3), for the three paleo time periods, and also for the Historical (1850-2014) and post 1975 (1975-2014) periods. The paleoclimate observed true GMST metrics are assessed values from the IPCC [26]; the equivalent site-specific global SAT and SST modelled and observed metrics (${}^{l,p,e}\Delta T^{s}$) are shown in Supp Info, Figure S1. First of all, it is interesting to note that in the observations, the ratio of mean temperature change to uncertainty in this change (i.e. the signal-to-noise ratio) is similar across the five time periods (Figure 2, black circles and vertical error bars). The LGM has the largest signal-to-noise ratio for GMST, even larger than

the historical record, indicating that it may be the most stringent target for model-data comparisons. This is associated with the fact that the LGM has a greater density of proxy data sites than the other paleo time periods. It is also important to note that the LGM has less uncertainty in the forcing boundary conditions than the other two paleo time periods (in particular CO_2, for which ice core records [e.g. 10, 33] give more accurate and precise values than is possible for the MPWP or EECO, where only indirect CO_2 proxies are available). As such, the uncertainty in the GMST sensitivity to forcing for the Pliocene and EECO compared to the LGM is greater than would be implied from the uncertainties in GMST alone. However, the 5-7°C IPCC assessment of LGM GMST cooling may be overly narrow; recent work has suggested a central GMST estimate of 4.5 °C of cooling [Figure 2, black open circle and dashed range, 34].

For each paleo time period, the multi-model mean GMST metric sits within the observed range, which is quite remarkable given that from the LGM to EECO this represents a temperature range of about 20°C. However, the spread across the ensemble is relatively large, and many individual models sit outside the observed range (78 %, 65%, 29% for the LGM, MPWP, and EECO respectively).

Previous studies have not always found a clear correlation between modern ECS and paleo GMST [e.g. 35, 36]. Although the ECS of every model in this study is not available, there is some indication that models with an ECS that is known to be greater than the IPCC assessed range of 2–5°C simulate too great a change in the paleo time periods (red dots in Figure 2c-e). Similarly, models with an ECS that is known to be lower than this range simulate too small a change in GMST in the paleo time periods (blue dots in Figure 2c-e). Only one model, CESM2, carried out simulations across all five time periods. Apart from that, CESM1.2 is the only model which carried out simulations across all three paleo time periods. The results from these two models, highlighted in Figure 2, indicate a consistency in relative GMST change across the paleo time periods for a particular model. However, more models

carrying out simulations across multiple paleo time periods would allow this to be explored further, and allow emergent constraints on ECS [37] from multiple time periods to be developed. This would also require all PMIP models to carry out $4\times CO_2$ simulations alongside their paleo simulations in order to calculate their ECS.

It also appears that both high and low ECS models can simulate the Historical period in good agreement with observations (Figure 2b), and low ECS models can simulate the post-1970 warming (Figure 2a). Therefore, paleoclimates may be a better discriminator of high- and low-ECS models than the observational periods (which is consistent with findings from an assessment of ECS that included paleoclimate evidence [38]). This may be due to the fact that the paleoclimate simulations are close to equilibrium with the CO_2 forcing, whereas the Historical simulations are transient, and as such have a GMST that is influenced by a transient pattern effect [e.g. 39], and/or it may be related to uncertainties in the aerosol forcing over the historical period [40]. However, more paleo simulations are required to further confirm this relation. In particular, there is a need for more paleo model simulations to be carried out with the same models that carry out the Historical CMIP simulations (this lack of consistency between the CMIP6 and PMIP4 model ensembles arises, at least in part, due to the long integration lengths required for full equilibrium of paleoclimate simulations).

It is also apparent that for all three past time periods there has been an improvement in the modelled GMST in the PMIP4/CMIP6 paleoclimate model simulations compared with the previous CMIP5/PMIP3 simulations (large versus small dark grey dots in Figure 2c-e). This improvement is likely due to a combination of updated boundary conditions, and improvements to the models themselves. Key changes in boundary conditions in PMIP4 compared with PMIP3 include updated ice sheets for the LGM [41], updated paleogeography and representation of ocean gateways for the Pliocene [42], and a consistent experimental design for the EECO including a new

paleogeography [43]. It is harder to robustly identify particular model improvements that may be relevant, because there is no clear lineage between the models in PMIP3 and PMIP4, but, for some models at least, improvements in model representation of cloud microphysics are playing an important role [e.g. 44, 45].

2.2 Polar Amplification

The site-specific polar amplification metrics (see Online Methods, Section 4.3), (${}^{l,p,e}\Delta P^{t,s}$), are shown in Figure 3a. Because the MPWP and EECO are warmer than the preindustrial whereas the LGM is colder, the observed site-specific metric from proxies is positive for the EECO and MPWP but is negative for the LGM (black circles in Figure 3, see Online Methods, Section 4.2 for a description of how the error bars are calculated). For all three time periods this indicates a polar amplification associated with increasing temperature (i.e. a decrease in meridional temperature gradient with increasing temperature).

For the LGM, the proxies indicate a site-specific SST polar amplification of about -0.4°C, whereas the model ensemble mean indicates a greater amplification of -0.7°C (dark grey circles in Figure 3a). The proxy value sits within the model range, but the model range is large compared with the uncertainty range from the proxies, from 0.1 °C (IPSLCM5A2) to -1.4°C (CESM2). For the MPWP and EECO, the polar amplification indicated by the proxies is greater than in any of the models, although for the MPWP two models do get close to the observed value of 1.7 °C and are within the uncertainty range of the proxy metric. For the EECO, the model-data disagreement is much starker, with nearly double the polar amplification in the proxies (12°C) than in the model with the greatest value (CESM2; 7°C). This discrepancy is primarily because of exceptionally warm proxy temperatures in the southwest Pacific. Many reasons for possible warm biases in the proxy temperatures in this region have been proposed, including a seasonal bias in mid- and high-latitude SST proxies [32, 46],

and/or uncertainties in the functional form of different paleo temperature proxies (e.g., TEX_{86}) in the upper temperature range [47, 48]. Since data from this region represent a large number of the high latitude records available from the EECO, they bias the proxy-based metric towards extremely high values. With the SSTs from the southwest Pacific excluded, the proxy polar amplification decreases from 12°C to 4°C, and the model and data are in closer agreement (see Supp Info, Figure S2a). Note that our site-specific proxy-based metrics are not comparable with previous estimates of Eocene polar amplification [e.g. 44, 49], which were based on Mg/Ca estimates of deep ocean temperatures, and designed to be comparable with true model metrics.

There has been little change in the ensemble mean LGM or EECO SST polar amplification between PMIP4 and PMIP3, although improvements in cloud parameterisations since PMIP3 have been shown to improve simulation of polar amplification in the EECO for individual models [44, 50]. However, for the Pliocene there has been a substantial improvement. At least some of this improvement is likely related to the closure of the Bering Strait in the PMIP4 experimental design, which has been shown to increase Pliocene temperatures in the North Atlantic [51]. However, the proxies still indicate greater amplification than the models (0.8°C for PMIP4 and 0.25°C for PMIP3, compared with 1.7°C in the proxies).

For all three time periods, the site-specific polar amplification metric (${}^{l,p,e}\Delta P^s$) has a similar value to the true metric ${}^{l,p,e}\Delta P^t$ for most models. Across the ensemble, the true metric is greater than the site-specific metric in the MPWP (by 0.05°C), and less than the site-specific metric in the EECO (by 0.4°C); indicating that despite the sparsity of the proxy data, there is enough data for the site-specific polar amplification metric to be meaningful. However, the exception to this is for the CESM2 model at the LGM (red dot in the LGM panel of Figure 3a), where the site-specific metric (-1.4°C) is very different, and even of opposite sign, to the true metric (0.3°C). This because although the CESM2 LGM ΔT metric is greater than any other model (Figure 2),

the LGM polar SSTs can not drop below the freezing point of seawater, resulting in relatively low polar amplification in the true metric (see Supp Info, Figure S3b).

There is not enough proxy SAT data in the tropics to define a SAT polar amplification metric for the MPWP or the EECO, and there is not enough data in the Southern Hemisphere to define a global polar amplification metric for the LGM. However, it is possible to quantify the absolute changes in high-latitude SATs for all three time periods (see Supp Info, Figure S4a,b,c), and for the LGM a Northern Hemisphere-only polar amplification metric can be defined (see Supp Info, Figure S4a). This shows that the Northern Hemisphere LGM polar amplification is very well simulated by the PMIP4 model ensemble mean (-4.1°C) compared with the proxies (-4.2°C). For the Pliocene, the model ensemble is colder than the proxies in general in the Northern Hemisphere high latitudes, related to less warmth in the Eurasian and Northern America continental interiors than indicated by the proxies. It has been suggested that the warm proxy temperatures in this region may be related to seasonal biases and/or the lack of modern analogues for the associated pollen records [52]. For the EECO, the Southern Hemisphere high latitude temperatures are well simulated by the ensemble mean, which further supports that the Southwest Pacific SSTs proxy temperatures are biased too warm. For the Northern Hemisphere, the models simulate a greater polar amplification than the proxies, but this is largely due to a set of proxy temperatures at 45°N in North America which are relatively cold, and may be influenced by the local topography of the Rockies.

2.3 Land-sea warming contrast (LSWC)

The site-specific land-sea warming contrast (LSWC) metrics, $({}^{l,p,e}\Delta L^{t,s})$, are shown in Figure 3b. The proxies indicate a negative (positive) LSWC for the LGM (MPWP), indicating that for both these time periods the land surface SAT warms more than the ocean SST under warming GMST. However, for the EECO the proxies indicate

a negative LSWC under warming GMST. Again, this is related to the super warm southwest Pacific proxy SST temperatures, and discounting SSTs from that region results in a positive LSWC for the EECO (see Supp Info, Figure S2b). The terrestrial proxies for the Eocene are from a wider time window (56.0 to 47.8 Ma) than the marine proxies (53.3 to 49.1 Ma) [32], and in many cases have uncertain paleoaltitude, and so this may also be playing a role. For both the LGM and MPWP, the model ensemble has a lower magnitude LSWC than the proxies, and this discrepancy is greater in the PMIP4/CMIP6 models than in the PMIP3/CMIP5 models. For the MPWP, the proxy SAT locations are all in the mid latitudes of the Northern Hemisphere, and as discussed above, in this region the models simulate colder temperatures than indicated by the proxies (see Supp Info, Figure S4b), and it is this discrepancy which leads to the discrepancy in land-sea warming contrast. The model site-specific and true metrics differ from each other quite considerably, with the true metrics being lower than the site-specific metrics for all time periods, by 70%, 50%, and 40% for the LGM, MPWP, and EECO respectively.

3 Discussion

There is a remarkable relationship between the modelled GMST metric, ΔT, and the polar amplification metric, ΔP, across the three time periods, in both the site-specific and true metrics (Figure 4a). This is also supported in the proxies, in particular when the southwest Pacific sites are excluded from the EECO; in this case both models and proxies point to an approximately linear relationship between the two metrics. The fact that this relationship is so linear is surprising given the greatly reduced (or non-existent) sea ice in the EECO, indicating that other mechanisms of polar amplification (for example related to cloud feedbacks) are compensating for each other across different time periods, resulting in the linear relationship. This relationship is

also seen in proxy estimates of global mean temperature and meridional temperature gradient from across the last 95 million years [53].

In the models, there is also a clear relationship between the GMST metric and the LSWC metric (Figure 4b). In this case there is a non-linear relationship, with LSWC increasing at lower GMST, but then flattening out under the high temperatures of the EECO. This relationship, including a saturation, is consistent with theory based on contrasting surface humidities and lapse rates over land and ocean [28]. The LGM proxy data is consistent with this relationship, but Pliocene LSWC in the proxies is greater than in the models, even accounting for the error bars in the proxy metric. In the EECO, the proxies indicate a complete reversal in this relationship, but when the EECO southwest Pacific sites excluded again, the models and proxies are more consistent, especially accounting for the large error bars of the EECO proxy estimates of GMST and LSWC.

In this paper we have used metrics derived from paleo proxy data to evaluate climate model simulations of the LGM, MPWP, and EECO. We find that model ensemble mean GMSTs are in exceptionally good agreement with the proxy data for all three paleo time periods, and that this agreement has improved in CMIP6/PMIP4 compared to CMIP5/PMIP3. The LGM is shown to be a very stringent target for model evaluation and development due to its large signal-noise ratio, and well-defined boundary conditions. There are indications that model evaluation using the paleo proxy record can be a better discriminator of models with very high or very low climate sensitivity than using the Historical observational period. Models also simulate polar amplification, and the relationship between GMST and polar amplification in reasonable agreement with proxies. However, there are uncertainties associated with the proxy records in: i) the MPWP within the northern hemisphere continental interiors, and ii) during the EECO, particularly in the southwest Pacific. In addition, some proxy terrestrial sites are from high elevation regions that are not resolved in the models,

or, for the EECO, are from regions for which the paleoelevation is uncertain. Furthermore, the relatively wide temporal window of the EECO (4.12 Myr) means that the proxy signal is affected by orbital forcing, and temporal variations in CO_2. All of these proxy uncertainties should be further explored in future work in order to maximise the utility of the paleoclimate proxy record for model development. Land-sea warming contrast is reasonably well simulated at the LGM, but less so at the MPWP and EECO. The models indicate an increasing but saturating relationship between GMST and LSWC, consistent with theory.

Overall, the paper provides a framework for paleo model evaluation that can be used for future model development in the framework of CMIP7 and beyond [6, 8, 54]. The framework also provides a traceability to previous model generations, allowing a robust assessment of model improvements over time, through successive model development cycles.

4 Online Methods

4.1 Model simulations

The most recent experimental designs for the three time periods above are described in detail in Ref. [41] for the LGM, Ref. [42] for the MPWP, and Ref. [43] for the EECO. These experimental designs describe standard boundary conditions (e.g. CO_2, non-CO_2 greenhouse gases, ice sheets, and vegetation) to be implemented in models, and protocols for the simulations themselves (e.g. run length and initial conditions). Simulations carried out using these experimental designs are all classified here as PMIP4/CMIP6 simulations. The models that carried out these PMIP4 simulations are of varying complexity, and include models developed for use in CMIP6, as well as earlier iterations of CMIP. The large-scale features of these PMIP4 simulation results are discussed in Ref. [4] for the LGM, Ref. [1] for the MPWP (as part of the PlioMIP2 project), and Ref. [3] for the EECO (as part of the DeepMIP project).

Simulation results are also presented for previous model simulations in the framework of PMIP3/CMIP5, described in Ref. [4] for the LGM, [31] for the MPWP, and Ref. [55] for the EECO. Tables listing all the simulations used in this paper are given in Supp Info, Tables S1-S5.

Note that for the EECO, the NorESM1_F model uses a paleogeography with a different reference frame than the other models, and as such is only included in the GMST metric and not in the polar amplification or land-sea warming contrast metrics, which are reference frame-specific. Also for the EECO, there are fewer models presented here than in Ref. [3]. This is because here we only include those models which carried out simulations in the range $\times 4$ to $\times 8$ preindustrial levels of CO_2, in accordance with CO_2 proxy estimates for the EECO [3]. The exception is CESM2.1slab, which we include for context and which was run at $\times 3$.

4.2 Proxy datasets

In order to evaluate the model simulations, we use existing syntheses and compilations of paleo proxy data for all three time periods.

For the GMST metric, we make use of the IPCC AR6 assessments of GMST change for the three paleo time periods [26]. These are based on a thorough review of the literature, and are designed to be global metrics directly comparable with the global mean output from models (i.e, they are 'true' metrics, see Metrics section 4.3). For the LGM, we also include the GMST metric of [34].

For the polar amplification and land-sea warming contrast metric, we use site-based data; for the LGM we use Ref. [56] for the sea surface temperatures (SSTs) and Ref. [57] (at the locations defined in Ref. [58], which are the actual proxy locations that inform the global assimmilated dataset of [57]) for the land air temperatures (LATs). For the MPWP we use Ref. [59] for the SSTs and Ref. [60] for the LATs. For the EECO we use Ref. [61] for the SSTs and LATs.

4.3 Definition of metrics

For changes in GMST, polar amplification, and land-sea warming contrast, we can define two types of quantitative metrics. Firstly, 'true' quantities, Q^t, which in theory require SST, and LAT and near-surface air temperature (SAT) values to be defined over the entire ocean and globe respectively (i.e. at all gridcells of a model or global gridded observational dataset). SST^t is the ocean-only true global mean SST, LAT^t is the land-only true global mean SAT, and SAT^t is the true global mean SAT. Secondly, 'site-specific' means; SST^s, LAT^s, and SAT^s. These are similar to the true quantities, but rather than averaging over all gridcells, they are defined according to a particular paleo proxy dataset, and are averaged only over those cells/locations that include at least one proxy data point in that dataset. True quantities, Q^t, can in theory only be defined for globally gridded output, whereas site-specific quantities, Q^s can be defined either for global model output or for proxy datasets. In practice, the IPCC-assessed paleoclimate GMST metrics are also considered to be 'true' metrics, as discussed in Section 4.2. Site-specific quantities are simply the average of the temperatures at each site in the proxy dataset. All quantities can be defined for a particular time period (x, where x can be e for EECO, p for MPWP, l for LGM, or pi for preindustrial) and can also be defined for selected latitude ranges (r), ${}^{x}_{r}Q$, so that, for example the site-specific mean SST in the range 90S to 30S during the EECO, is written ${}^{e}_{-90:-30}SST^s$.

We then define 3 key metrics as a function of these quantities. In particular, the change in true or site-specific (t,s) mean temperature relative to the preindustrial (ΔT), for the LGM (l), MPWP (p), or EECO (e) is:

$$ {}^{l,p,e}\Delta T^{t,s} = {}^{l,p,e}SAT^{t,s} - {}^{pi}SAT^{t,s} \quad (1) $$

for SAT, and similarly for SST and LAT. The polar amplification metric (ΔP) is

$$ {}^{l,p,e}\Delta P^{t,s} = {}^{l,p,e}_{-30:+30}SST^{t,s} - {}^{l,p,e}_{\pm 60:\pm 90}SST^{t,s} - {}^{pi}_{-30:+30}SST^{t,s} + {}^{pi}_{\pm 60:\pm 90}SST^{t,s} \quad (2) $$

for SST, and similarly for LAT. The land-sea warming contrast metric (ΔL) is

$$ {}^{l,p,e}\Delta L^{t,s} = {}^{l,p,e}LAT^{t,s} - {}^{l,p,e}SST^{t,s} - {}^{pi}LAT^{t,s} + {}^{pi}SST^{t,s}. \quad (3) $$

The proxy compilations that we use are published with associated uncertainties in temperature for each individual site. However, the meaning of these uncertainty ranges is unclear in some cases, and inconsistent across different time periods. Here we interpret all published uncertainties as representing a range of uniformly distributed uncertainty. In order to estimate the associated uncertainty in the polar amplification and land-sea warming contrast site-specific proxy metrics that we present, we use monte carlo sampling to generate 100 proxy datasets, and use these to generate 100 associated metrics, from which we report a mean and a 90% uncertainty range (consistent with the IPCC 'very likely' range).

4.4 Developments since IPCC AR6

IPCC AR6 includes a Figure showing ensemble mean maps, and zonal means, of the SST and SAT data analysed in this paper ([18, ; Figure 7.13 therein]. Compared with the IPCC Figure, here we have carried out some developments, and incorporated these into our overall analysis: (1) Here, in Figure S3, the horizontal lines showing the banded mean SSTs, and the values given in the plot for the values of the polar amplification associated with these bands, are calculated using the ensemble mean SSTs only for those gridboxes where all models have an ocean grid ocean (cdo operator 'ensaver'). In the equivalent IPCC plot, the values given are the same as in Figure S1, but the

horizontal lines were calculated using the mean of the models for all gridboxes for which at least one model had ocean (cdo operator 'ensmean'). (2) Here, for extracting the modelled SST at the location of a proxy, for SST proxy locations which were defined as land in the models, the nearest ocean gridcell was used to define the model value. In the IPCC, due to a coding error, the nearest-but-one ocean gridcell was used. (3) Here, we assigned an uncertainty of ±5°C for any proxy data that did not have an associated uncertainty in the original reference. In the IPCC, due to a coding error, an error of zero was assigned. (4) Here, with the exception of NorESM stated above, all models are used to calculate all three metrics. In the IPCC, the EECO CESM2.1slab simulation was not included in the map of the ensemble mean map or in the plot of the zonal mean.

Supplementary information. Supplementary material is available, consisting of Figures S1-S5, and Tables S1-S5.

Data availability statement. All model outputs and proxy data used in this study are available from the IPCC AR6 Data Distribution Centre (https://www.ipcc-data.org/), in the archive for Figure 7.13 of WG1 (https://ipcc-browser.ipcc-data.org/browser/dataset/7509/0 ; https://dx.doi.org/10.5285/4dbd3ccb85d747188586735133f1d3d9).

Code availability statement. The code for carrying out the analysis and making the plots is available from https://github.com/danlunt1976/ipcc_ar6/blob/master/patterns/fgd/plot_all_metrics.pro, version fb09c5e.

Competing Interests. The authors declare no competing interests.

Author contributions. DJL carried out the analysis and wrote the first draft of the paper. BLOB, CB, AH, GNI, KI, MK, DK, TM, ELMcC, US, SS, JET, AZ, and JZ discussed the paper and provided edits.

Acknowledgments. DJL acknowledges NERC grants NE/P01903X/1 (SWEET:Super-Warm Early Eocene Temperatures and climate: understanding the response of the Earth to high CO_2 through integrated modelling and data) and NE/X000222/1 (PaleoGradPhan: Paleoclimate meridional and zonal Gradients in the Phanerozoic). GNI is supported by a GCRF Royal Society Dorothy Hodgkin Fellowship (DHF/R1/191178) with additional support via the Royal Society (RF/ERE/231019, RF/ERE/210068). US acknowledges NERC grant NE/P019137/1. The CESM project is supported primarily by the National Science Foundation (NSF). This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. All authors acknowledge CMIP6/PMIP4 and the associated infrastructure that makes model intercomparisons possible, and all the modelling groups that contributed simulations that have been included in this study.

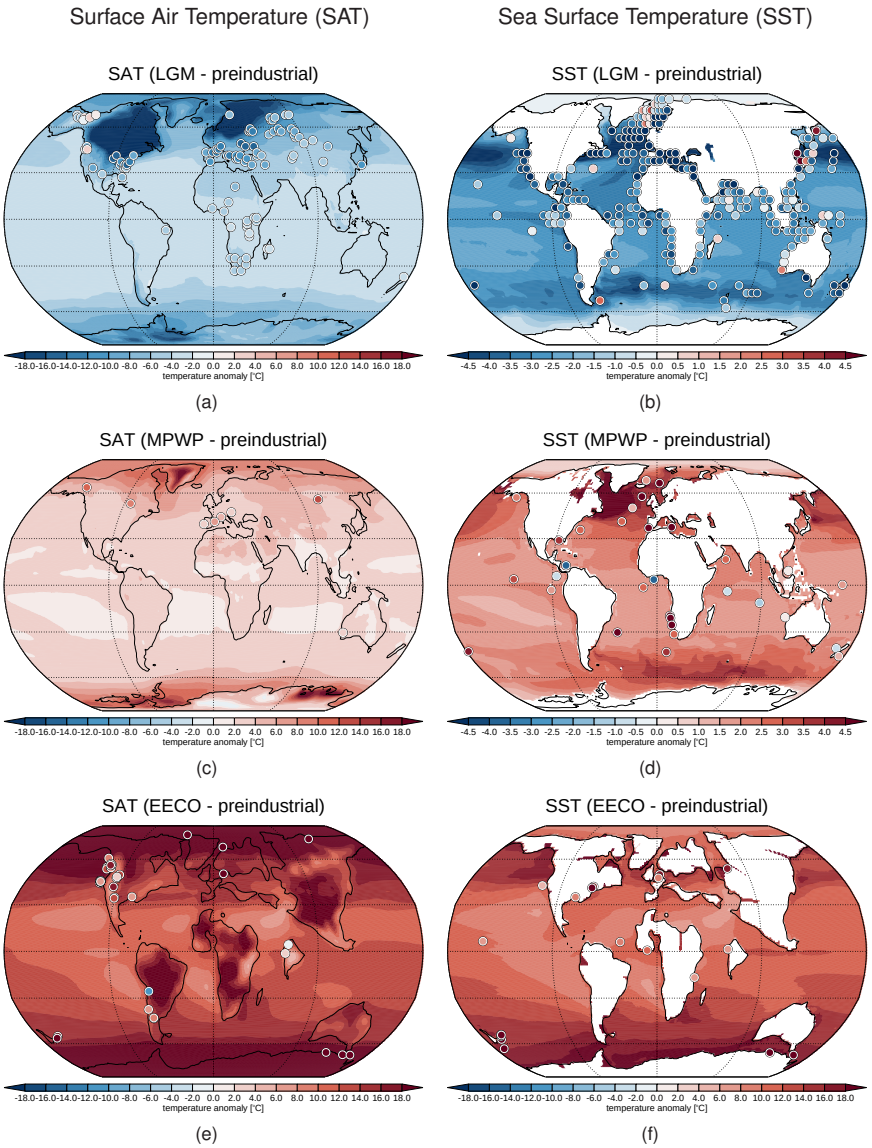

Fig. 1 Patterns of model and proxy temperature change relative to preindustrial. Patterns of (a,c,e) near surface air temperature (SAT), and (b,d,f) sea surface temperature (SST), in paleo proxies and models of the (a,b) Last Glacial Maximum (LGM), (c,d) the Mid-Pliocene Warm Period (MPWP), and the (e,f) Early Eocene Climatic Optimum (EECO). Modelled ensemble-mean temperature anomalies compared with pre-industrial are shown in the background colours. Proxy near-surface air temperatures and SST anomalies are shown as coloured circles (see Online Methods, Section 4.2). Note the differing colour scales for each map.

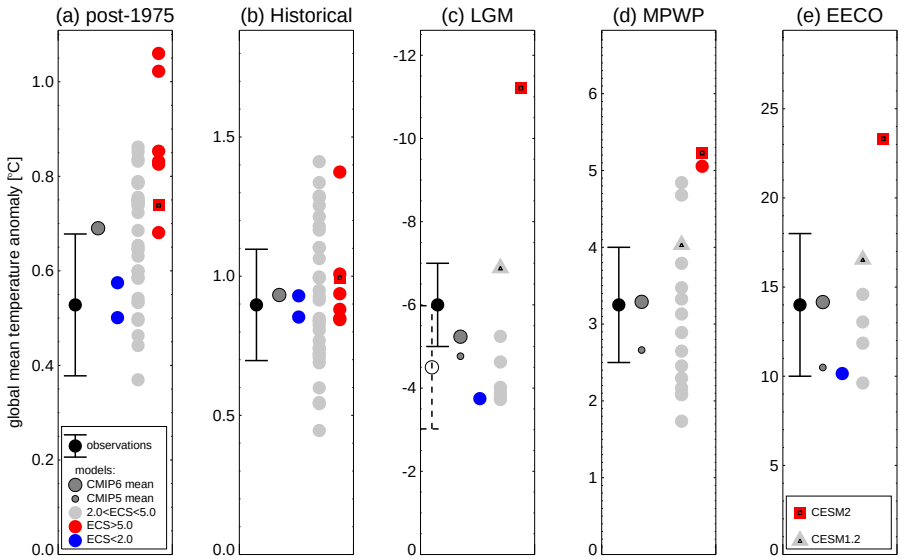

Fig. 2 Model and proxy global mean temperature change relative to preindustrial. Global mean true surface temperature (GMST) anomaly, ${}^{l,p,e}\Delta T^{t}$ in models and observations from five time periods. (a) post-1975; (b) Historical; (c) Last Glacial Maximum (LGM, ^{l}); (d) mid-Pliocene Warm Period (MPWP, ^{p}); (e) Early Eocene Climatic Optimum (EECO, ^{e}). Light grey circles show CMIP6/PMIP4 models with ECS in the very likely range as assessed by [18]; models in red have an ECS greater than the assessed very likely range ($> 5°$C); models in blue have an ECS lower than the assessed very likely range ($< 2°$C). Dark grey large circles show the multi-model ensemble mean for CMIP6/PMIP4. Dark grey large circles show the multi-model ensemble mean for CMIP5/PMIP3. Black circles and very likely ranges show the IPCC assessed temperature anomaly derived from observations [26]. For the LGM, the black open circle with dashed very likely uncertainty range shows the GMST anomaly estimate from [34]. The Historical anomaly in models and observations is calculated as the difference between 2005–2014 and 1850–1900, and the post-1975 anomaly is calculated as the difference between 2005–2014 and 1975–1984. For the LGM, MPWP and EECO, modelled temperature anomalies are compared with pre-industrial. The square symbol denotes the five simulations carried out by CESM2, and the triangle symbol denotes the three simulations carried out by CESM1.2. A version of this figure with all models labelled is in the Supp Info, Figure S5, and all the models in this plot are listed in order of GMST in the Supp Info, Tables S1-S5. A similar plot of the paleo time periods, but for the site-specific metric, ${}^{l,p,e}\Delta T^{s}$, is shown in Supp Info, Figure S1.

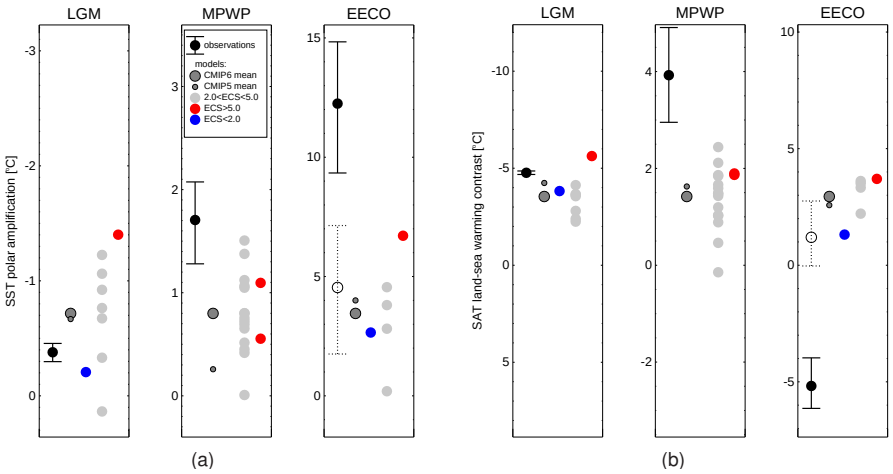

Fig. 3 Metrics of polar amplification and land-sea warming contrast. Metrics for (a) SST polar amplification (${}^{l,p,e}\Delta P^{t,s}$) and (b) land-sea warming contrast (${}^{l,p,e}\Delta L^{t,s}$), for Last Glacial Maximum (LGM, ^{l}), mid-Pliocene Warm Period (MPWP, ^{p}), and early Eocene Climatic Optimum (EECO, ^{e}). Black circles and very likely ranges show the observed site-specific metric (^{s}), dark grey circles show the model ensemble mean site-specific metric (large circles for CMIP6/PMIP4 and small circles for CMIP5/PMIP3), and light-grey/red/blue circles show the individual CMIP6/PMIP4 model site-specific metric. The EECO observed metric shown with an open circle and dotted error bar excludes SST data from the southwest Pacific. All metrics are calculated relative to the preindustrial.

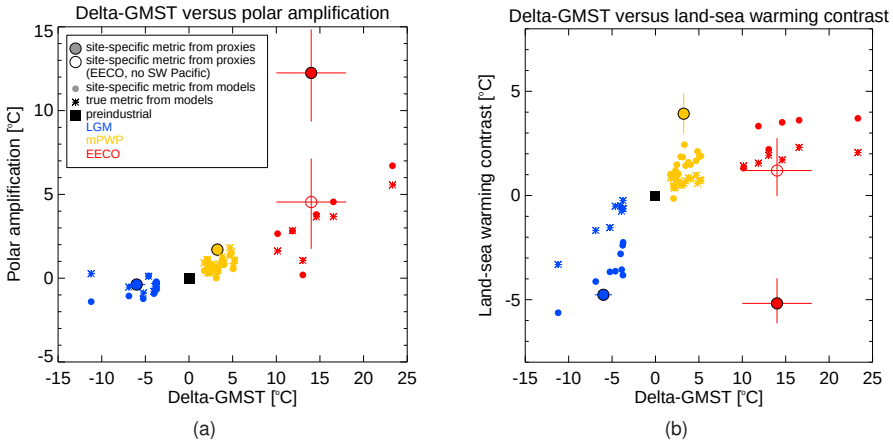

Fig. 4 Relationship between global mean surface temperature, polar amplification, and land-sea warming contrast. Relationship between metrics for (a) GMST (${}^{l,p,e}\Delta T^{t,s}$) and polar amplification (${}^{l,p,e}\Delta P^{t,s}$)), and (b) GMST and land-sea warming contrast (${}^{l,p,e}\Delta L^{t,s}$), for the Last Glacial Maximum (LGM; blue, ^{l}), mid-Pliocene Warm Period (MPWP; orange, ^{p}), and early Eocene Climatic Optimum (EECO; red, ^{e}). Large circles and very likely ranges show the observed site-specific metric (^{s}), small circles show the model site-specific metric for all CMIP6/PMIP4 models, and stars show the true model metric (^{t}) for all CMIP6/PMIP4 models. The square shows the preindustrial. The EECO observed metric shown with an open circle excludes SST data from the southwest Pacific.

References

[1] Haywood, A. *et al.* What can palaeoclimate modelling do for you? *Earth Systems and Environment* **3**, 1–18 (2019). URL https://link.springer.com/article/10.1007/s41748-019-00093-1.

[2] Haywood, A. M. *et al.* The Pliocene Model Intercomparison Project Phase 2: large-scale climate features and climate sensitivity. *Climate of the Past* **16**, 2095–2123 (2020). URL https://cp.copernicus.org/articles/16/2095/2020/.

[3] Lunt, D. J. *et al.* DeepMIP: model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. *Climate of the Past* **17**, 203–227 (2021). URL https://cp.copernicus.org/articles/17/203/2021/.

[4] Kageyama, M. *et al.* The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations. *Climate of the Past* **17**, 1065–1089 (2021). URL https://cp.copernicus.org/articles/17/1065/2021/.

[5] Zhu, J. *et al.* LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. *Journal of Advances in Modeling Earth Systems* **14**, e2021MS002776 (2022). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002776. E2021MS002776 2021MS002776.

[6] Burls, N. & Sagoo, N. Increasingly sophisticated climate models need the out-of-sample tests paleoclimates provide. *Journal of Advances in Modeling Earth Systems* **14**, e2022MS003389 (2022). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003389. E2022MS003389 2022MS003389.

[7] Burke, K. D. *et al.* Pliocene and eocene provide best analogs for near-future climates. *Proceedings of the National Academy of Sciences* **115**, 13288–13293

(2018). URL https://www.pnas.org/doi/abs/10.1073/pnas.1809600115.

[8] Tierney, J. E. *et al.* Past climates inform our future. *Science* **370**, eaay3701 (2020). URL https://www.science.org/doi/abs/10.1126/science.aay3701.

[9] Kageyama, M. *et al.* The PMIP4 contribution to CMIP6 – part 1: Overview and over-arching analysis plan. *Geoscientific Model Development* **11**, 1033–1057 (2018). URL https://gmd.copernicus.org/articles/11/1033/2018/.

[10] Luthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000-800,000 years before present. *nature* **453** (2008).

[11] Peltier, W. R., Argus, D. F. & Drummond, R. Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_C (VM5a) model. *Journal of Geophysical Research: Solid Earth* **120**, 450–487 (2015). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JB011176.

[12] Vega, E. D. L., Chalk, T. B., Wilson, P. A., Bysani, R. P. & Foster, G. L. Atmospheric CO_2 during the Mid- Piacenzian Warm Period and the M2 glaciation. *Scientific Reports* **10** (2020). URL https://eprints.soton.ac.uk/444151/.

[13] Dolan, A., de Boer, B., Bernales, J., Hill, D. & Haywood, A. High climate model dependency of pliocene antarctic ice-sheet predictions. *Nature Communications* **9** (2018). URL https://eprints.whiterose.ac.uk/131951/.

[14] Anagnostou, E. *et al.* Proxy evidence for state-dependence of climate sensitivity in the eocene greenhouse. *Nature Communications* **11** (2020). URL https://rdcu.be/b6RYM.

[15] Izumi, K., Bartlein, P. J. & Harrison, S. P. Consistent large-scale temperature responses in warm and cold climates. *Geophysical Research Letters* **40**, 1817–1823 (2013). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50350.

[16] Drost, F., Karoly, D. & Braganza, K. Communicating global climate change using simple indices: an update. *Climate Dynamics* **39**, 989–999 (2012).

[17] United Nations Framework Convention on Climate Change (UNFCCC). The Paris Agreement (2016). URL https://unfccc.int/sites/default/files/resource/parisagreement_publication.pdf.

[18] Forster, P. *et al. The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity*, 923–1054 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021).

[19] Walsh, J. E. Intensified warming of the Arctic: Causes and impacts on middle latitudes. *Global and Planetary Change* **117**, 52–63 (2014). URL https://www.sciencedirect.com/science/article/pii/S0921818114000575.

[20] Rantanen, M. *et al.* The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth and Environment* **3**, 168 (2022).

[21] Fox-Kemper, B. *et al. Ocean, Cryosphere and Sea Level Change*, 1211–1362 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021).

[22] Previdi, M., Smith, K. L. & Polvani, L. M. Arctic amplification of climate change: a review of underlying mechanisms. *Environmental Research Letters* **16**, 093003 (2021). URL https://dx.doi.org/10.1088/1748-9326/ac1c29.

[23] Armour, K. C., Siler, N., Donohoe, A. & Roe, G. H. Meridional atmospheric heat transport constrained by energetics and mediated by large-scale diffusion. *Journal of Climate* **32**, 3655 – 3680 (2019). URL https://journals.ametsoc.org/view/journals/clim/32/12/jcli-d-18-0563.1.xml.

[24] Graversen, R. G., Langen, P. L. & Mauritsen, T. Polar amplification in ccsm4: Contributions from the lapse rate and surface albedo feedbacks. *Journal of Climate* **27**, 4433 – 4450 (2014). URL https://journals.ametsoc.org/view/journals/clim/27/12/jcli-d-13-00551.1.xml.

[25] Pithan, F. & Mauritsen, T. Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience* **7**, 181–184 (2014).

[26] Gulev, S. *et al.* *Changing State of the Climate System*, 287–422 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021).

[27] Joshi, M. & Gregory, J. Dependence of the land-sea contrast in surface climate response on the nature of the forcing. *Geophysical Research Letters* **35** (2008). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GL036234.

[28] Byrne, M. P. & O'Gorman, P. A. Link between land-ocean warming contrast and surface relative humidities in simulations with coupled climate models. *Geophysical Research Letters* **40**, 5223–5227 (2013). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50971.

[29] Zuo, Z. & Zhang, K. Link between the land–sea thermal contrast and the Asian summer monsoon. *Journal of Climate* **36**, 213 – 225 (2023). URL https://journals.ametsoc.org/view/journals/clim/36/1/JCLI-D-21-0944.1.xml.

[30] Gebbie, G., Streletz, G. J. & Spero, H. J. How well would modern-day oceanic property distributions be known with paleoceanographic-like observations? *Paleoceanography* **31**, 472–490 (2016). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015PA002917.

[31] Haywood, A. *et al.* On the identification of a Pliocene time slice for data-model comparison. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.* **371** (2013). EPrint Processing Status: Full text deposited in DRO.

[32] Hollis, C. J. *et al.* Early Paleogene temperature history of the southwest Pacific Ocean: Reconciling proxies and models. *Earth and Planetary Science Letters* **349-350**, 53–66 (2012).

[33] Monnin, E. *et al.* Atmospheric co¡sub¿2¡/sub¿ concentrations over the last glacial termination. *Science* **291**, 112–114 (2001). URL https://www.science.org/doi/abs/10.1126/science.291.5501.112.

[34] Annan, J. D., Hargreaves, J. C. & Mauritsen, T. A new global surface temperature reconstruction for the last glacial maximum. *Climate of the Past* **18**, 1883–1896 (2022). URL https://cp.copernicus.org/articles/18/1883/2022/.

[35] Renoult, M., Sagoo, N., Zhu, J. & Mauritsen, T. Causes of the weak emergent constraint on climate sensitivity at the last glacial maximum. *Climate of the Past* **19**, 323–356 (2023). URL https://cp.copernicus.org/articles/19/323/2023/.

[36] Hargreaves, J. C. & Annan, J. D. Could the pliocene constrain the equilibrium climate sensitivity? *Climate of the Past* **12**, 1591–1599 (2016). URL https://cp.copernicus.org/articles/12/1591/2016/.

[37] Renoult, M. *et al.* A Bayesian framework for emergent constraints: case studies of climate sensitivity with PMIP. *Climate of the Past* **16**, 1715–1735 (2020). URL https://cp.copernicus.org/articles/16/1715/2020/.

[38] Sherwood, S. C. *et al.* An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics* **58**, e2019RG000678 (2020). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000678. E2019RG000678 2019RG000678.

[39] Dong, Y. *et al.* Intermodel spread in the pattern effect and its contribution to climate sensitivity in CMIP5 and CMIP6 models. *Journal of Climate* **33**, 7755 – 7775 (2020). URL https://journals.ametsoc.org/view/journals/clim/33/18/jcliD191011.xml.

[40] Bellouin, N. *et al.* Bounding global aerosol radiative forcing of climate change. *Reviews of Geophysics* **58**, e2019RG000660 (2020). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000660. E2019RG000660 10.1029/2019RG000660.

[41] Kageyama, M. *et al.* The PMIP4 contribution to CMIP6 – Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments. *Geoscientific Model Development* **10**, 4035–4055 (2017). URL https://gmd.copernicus.org/articles/10/4035/2017/.

[42] Haywood, A. M. *et al.* The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: scientific objectives and experimental design. *Climate of the Past* **12**, 663–675 (2016). URL https://cp.copernicus.org/articles/12/663/2016/.

[43] Lunt, D. J. *et al.* 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 (2017). URL https://gmd.copernicus.org/articles/10/889/2017/.

[44] Zhu, J., Poulsen, C. J. & Tierney, J. E. Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Science Advances* **5**, eaax1874 (2019). URL https://www.science.org/doi/abs/10.1126/sciadv.aax1874.

[45] Feng, R., Otto-Bliesner, B. L., Brady, E. C. & Rosenbloom, N. Increased climate response and Earth System Sensitivity from CCSM4 to CESM2 in Mid-Pliocene simulations. *Journal of Advances in Modeling Earth Systems* **12**, e2019MS002033 (2020). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS002033. E2019MS002033 2019MS002033.

[46] Davies, A., Hunter, S. J., Gréselle, B., Haywood, A. M. & Robson, C. Evidence for seasonality in early Eocene high latitude sea-surface temperatures. *Earth and Planetary Science Letters* **519**, 274 – 283 (2019). URL http://www.sciencedirect.com/science/article/pii/S0012821X1930295X.

[47] Cramwinckel, M. J. *et al.* Synchronous tropical and polar temperature evolution in the Eocene. *Nature* **559**, 382–386 (2018). URL https://EconPapers.repec.org/RePEc:nat:nature:v:559:y:2018:i:7714:d:10.1038_s41586-018-0272-2.

[48] Inglis, G. & Tierney, J. E. *The TEX86 paleotemperature proxy* (Cambridge University Press, 2020). URL https://eprints.soton.ac.uk/442791/.

[49] Evans, D. *et al.* Eocene greenhouse climate revealed by coupled clumped isotope-mg/ca thermometry. *Proceedings of the National Academy of Sciences* **115**, 1174–1179 (2018). URL https://www.pnas.org/doi/abs/10.1073/pnas.1714744115.

[50] Kiehl, J. T. & Shields, C. A. Sensitivity of the palaeocene–eocene thermal maximum climate to cloud properties. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371** (2013). URL https://api.semanticscholar.org/CorpusID:19531501.

[51] Otto-Bliesner, B. L. *et al.* Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways. *Geophysical Research Letters* **44**, 957–964 (2017). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071805.

[52] Tindall, J. C., Haywood, A. M., Salzmann, U., Dolan, A. M. & Fletcher, T. The warm winter paradox in the Pliocene northern high latitudes. *Climate of the Past* **18**, 1385–1405 (2022). URL https://cp.copernicus.org/articles/18/1385/2022/.

[53] Gaskell, D. E. *et al.* The latitudinal temperature gradient and its climate dependence as inferred from foraminiferal δ^{18}O over the past 95 million years. *Proceedings of the National Academy of Sciences* **119**, e2111332119 (2022). URL https://www.pnas.org/doi/abs/10.1073/pnas.2111332119.

[54] Zhu, J., Poulsen, C. J. & Otto-Bliesner, B. L. High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change* **10**, 378–379 (2020). URL https://ideas.repec.org/a/nat/natcli/v10y2020i5d10.1038_s41558-020-0764-6.html.

[55] Lunt, D. J. *et al.* A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP. *Climate of the Past* **8**, 1717–1736 (2012). URL https://cp.copernicus.org/articles/8/1717/2012/.

[56] Tierney, J. E. *et al.* Glacial cooling and climate sensitivity revisited. *Nature* **584**, 569–573 (2020).

[57] Cleator, S. F., Harrison, S. P., Nichols, N. K., Prentice, I. C. & Roulstone, I. A new multivariable benchmark for Last Glacial Maximum climate simulations. *Climate of the Past* **16**, 699–712 (2020). URL https://cp.copernicus.org/articles/16/699/2020/.

[58] Bartlein, P. *et al.* Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics* **37**, 775 – 802 (2011).

[59] McClymont, E. L. *et al.* Lessons from a high-Co_2 world: an ocean view from $\sim$ 3million years ago. *Climate of the Past* **16**, 1599–1615 (2020). URL https://cp.copernicus.org/articles/16/1599/2020/.

[60] Salzmann, U. *et al.* Challenges in quantifying Pliocene terrestrial warming revealed by data-model discord. *Nature Climate Change* **3**, 969–974 (2013). URL http://nora.nerc.ac.uk/id/eprint/503649/.

[61] Inglis, G. N. *et al.* Global mean surface temperature and climate sensitivity of the early Eocene Climatic Optimum (EECO), Paleocene–Eocene Thermal Maximum (PETM), and latest Paleocene. *Climate of the Past* **16**, 1953–1968 (2020). URL https://cp.copernicus.org/articles/16/1953/2020/.