The temperature of the deep ocean is a robust proxy for global mean surface temperature during the Cenozoic

David Evans[1], Julia Brugger[2], Gordon N. Inglis[1], Paul Valdes[3]

[1] School of Ocean and Earth Science, University of Southampton, Southampton, SO14 3ZH, UK
[2] Senckenberg Biodiversity and Climate Research Centre (SBiK-F), 60325 Frankfurt am Main, Germany
[3] School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK

Reconstructing past changes in global mean surface temperature (GMST) is one of the key contributions that palaeoclimate science can make in addressing societally relevant questions and is required to determine equilibrium climate sensitivity (ECS). GMST may be derived from proxy data compilations and/or via simple metrics that can be directly related to global temperature. One such metric is the temperature of the deep ocean (T_d), which has the advantage of being reasonably well-constrained for much of the Cenozoic. Previous work has suggested a simple T_d-GMST scaling factor of 1 prior to the Pliocene. If this were to be accurate, it would enable GMST to be precisely determined for any interval of interest. However, this metric lacks a robust mechanistic basis, and indeed, such a relationship is intuitively difficult to envisage given that polar amplification is a ubiquitous feature of past warm climate states and deep water overwhelmingly forms at high latitudes. Here, we interrogate whether and crucially, why, this relationship exists using a suite of curated data compilations generated for key deep-time climate intervals (Pliocene, early Palaeogene) as well as two independent sets of palaeoclimate model simulations. We show that models and data are in full agreement that a 1:1 relationship is a good approximation. Mechanistically, both sets of climate models suggest that i) increasingly seasonally biased deep water formation, and ii) a faster rate of land versus ocean surface warming are the two processes that act to counterbalance a possible polar amplification-derived bias on T_d-derived GMST. Using this knowledge, we interrogate the quality of the existing deep ocean temperature datasets and provide a new Cenozoic record of GMST. Our estimates are substantially warmer than similar previous efforts for much of the Paleogene and are thus consistent with a substantially higher-than-modern ECS during deep-time high CO_2 climate states.

Key Points:

- Deep ocean temperature changes are used to constrain global mean surface temperature yet the underlying assumptions have not been scrutinised in detail
- Both curated data compilations and climate model simulations demonstrate that deep ocean-derived GMST estimates are robust
- We update the transformation equations and provide a revised estimate of GMST through the Cenozoic

1. Introduction

One of the most important contributions that quantitative reconstructions of Earth's climate can make to society is as an empirical method of constraining key aspects of Earth's climate system (e.g. Gulev et al., 2021; Tierney et al., 2020). Perhaps the most fundamental parameter of interest within this context is Earth's equilibrium climate sensitivity (ECS), which broadly describes the change in global near surface temperature per CO_2 doubling (Sherwood et al., 2020), knowledge of which is required to determine the degree to which our planet will warm over the coming centuries and millennia.

The precise definition of ECS depends on a number of factors such as which long-term feedbacks are taken into account (Rohling et al., 2012; Sherwood et al., 2020), but from a past climate perspective, constraining ECS at the very least requires a reconstruction of radiative forcing (CO_2) and global temperature, and the timescale of interest to be defined. The late Pleistocene, and the last glacial maximum (LGM) in particular have received much attention in terms of deriving ECS from the palaeoclimate record (Osman et al., 2021; Rohling et al., 2012; Schmittner et al., 2011; Sherwood et al., 2020), because direct measurements of atmospheric CO_2 are available from the ice core record, and because an enormous wealth of proxy information exists from both the terrestrial and marine realm.

The use of deep-time palaeoclimate records (pre-Pleistocene) to determine the value of important parameters like ECS has received an increasing amount of attention over the course of the last few decades (Anagnostou et al., 2016; Hansen et al., 2013; Inglis et al., 2020; Martínez-Botí et al., 2015; Zhu et al., 2019). One advantage of this approach is that CO_2 was greater than at present, providing insights into nonlinear features in Earth's climate system such as ice sheet dynamics (Foster and Rohling, 2013; von der Heydt et al., 2014) which cannot be fully determined from the study of cooler-than-modern climate states (Anagnostou et al., 2016; Inglis et al., 2020; Martínez-Botí et al., 2015; Pagani et al., 2010).

The development of precise and accurate methods of reconstructing past changes in CO_2 throughout the Cenozoic (Anagnostou et al., 2016; Foster and Rae, 2016; Hönisch et al., 2012; Pagani, 2002; Pearson and Palmer, 1999) opened up the study of geologic intervals prior to the ice core CO_2 record for this purpose, which was previously challenging in part because of the large uncertainties associated with the CO_2 data (Covey et al., 1996; Hoffert and Covey, 1992). With the production of high quality CO_2 data for much of the Cenozoic (Rae et al., 2021), the accuracy and precision with which global mean surface temperature (GMST) is known has become an increasingly important source of uncertainty in the derivation of ECS from pre-Pleistocene warm intervals (Inglis et al., 2020; McClymont et al., 2020).

Datasets designed to reconstruct past changes in GMST can be broadly placed into two categories: i) the synthesis of large amounts of spatially-distributed data sufficient to constrain the global climate at the time, or ii) the parameterisation of GMST in terms of a single, well-constrained aspect of Earth's climate. In an ideal world, the first of these is preferable as it avoids any

assumption that goes into indirect approaches, but it requires sufficient palaeotemperature reconstructions to be available within a sufficiently narrow time interval to be able to constrain GMST unbiased by (e.g.) latitudinal and zonal heterogeneities in surface temperature. However, few Cenozoic warm intervals exist with sufficient data density to facilitate the calculation of GMST with sufficient certainty, with possible exceptions being the Pliocene Warm Period (PWP, ~3.2 Ma (Dowsett et al., 2016; Haywood et al., 2013; McClymont et al., 2020)), Miocene climatic optimum (MCO, ~16-14 Ma (Burls et al., 2021)), and perhaps also intervals within the early Palaeogene (Hollis et al., 2019). In this latter case, the DeepMIP project recently produced a curated data compilation for the early Eocene climatic optimum (EECO; ~49.1-53.3 Ma), Paleocene-Eocene thermal maximum (PETM; ~56 Ma), and the latest Paleocene (LP, the interval immediately preceding the PETM; ~57-56 Ma), compiling over 1500 'high confidence' quantitative estimates of terrestrial and ocean surface temperature and CO_2 for these intervals (Hollis et al., 2019), see https://www.deepmip.org/data/. Using multiple methodologies, Inglis et al. (2020) derived GMST for these intervals, constraining ECS to 3.1 to 4.5°C albeit with large uncertainties, while Tierney et al. (2022) constrain ECS to 5.7-7.4°C using LP and PETM data coupled with a paleoclimate data assimilation approach.

The difficulty in compiling and quality checking datasets that are large enough to constrain GMST given the spatial heterogeneity in Earth's surface climate has led to the development of techniques based on a simple, (relatively) easily determinable parameter. The key feature of Earth's climate system that has formed the basis of a transformation relationship to GMST is the temperature of the deep ocean (or rather, some closely related parameter such as the oxygen isotopic composition of foraminifera), because a continuous, high temporal resolution record exists for the entirety of the Cenozoic (Billups and Schrag, 2003; Cramer et al., 2011; Lear et al., 2000; Lisiecki and Raymo, 2005; Westerhold et al., 2020; Zachos et al., 2001, 2008). If records such as these can be reliably related to GMST, then many of the issues with generating and compiling large surface temperature datasets can be avoided. This approach, pioneered by Hansen et al. (2013, 2008), requires that the temperature of the deep ocean be coupled to that of the global surface climate. This is likely a reasonable assumption given that the temperature of the deep ocean is broadly similar to the temperature of the surface ocean in the regions of deep water formation, and the temperature of these regions may, in turn, be expected to relate predictably to GMST. The attraction of this approach is that GMST can be immediately calculated for any time interval of interest, and it is for this reason that the approach of Hansen et al. (2013) has been widely cited and reproduced (Lunt et al., 2016; Mills et al., 2019; Rae et al., 2021; Tierney et al., 2020; Westerhold et al., 2020). However, the underlying rationale for the details of the methodology are complex and have largely not been empirically tested, especially in deep time.

Here, we focus on addressing the question of whether GMST and the temperature of the deep ocean are linearly related with a scaling factor of 1, as suggested by Hansen et al. (2013), given that this may seem intuitively problematic. Deep water formation occurs at high latitudes in the modern ocean with both fully coupled climate models and proxy data suggesting that this was similarly the case for most, if not all, of the Cenozoic, although the locus of deep water formation likely shifted

(Ferreira et al., 2018; Ford et al., 2022; Valdes et al., 2021; Zhang et al., 2022). While much remains to be understood about deep-time climatic variation, a ubiquitous and well-constrained feature of multiple past warm climate states is that these are characterised by polar amplification (Burls et al., 2021; Cramwinckel et al., 2018; Evans et al., 2018; Gaskell et al., 2022; Lunt et al., 2012) and overall increased ocean stratification (Green and Huber, 2013; Winguth et al., 2012). Given that deep ocean temperature records are therefore effectively a record of high latitude surface temperature, we may expect a temperature record derived from the deep ocean to include a polar amplification component. As a result, surface temperature estimated from that of the deep ocean using a linear one-to-one relationship would, to a first-order approximation, be expected to result in a GMST overestimate, with the severity of the inaccuracy directly related to the degree of polar amplification in a given time interval.

In the following sections, we: identify the potential issues with this simple transformation of deep ocean temperature (inferred via benthic foraminifera $\delta^{18}O$) into GMST in Section 2, critically evaluate the quality of the Cenozoic deep ocean temperature datasets (Sec. 3.1), and then test whether the methodology can usefully approximate GMST by comparing to a combination of curated proxy data compilations and two sets of climate model simulations, including the DeepMIP model database (Sec. 3.2 and 3.3). In comparison to previous work, especially that of Goudsmit-Harzevoort et al. (2023) and Valdes et al. (2021), our analysis differs in several key respects in that i) we do not limit our analysis to climate model output, ii) we explicitly set out to test what relationship between GMST and deep ocean temperature is expected, for the entirety of the Cenozoic, and iii) present several advances in the way in which climate model data are evaluated for these purposes, detailed below (Sec. 3.3).

2. Relating global surface temperature to the deep ocean

The geochemistry of deep-dwelling benthic foraminifera forms the basis of our long term records of deep ocean temperature change, because of the near-continuous nature of the fossil record of these organisms in sediments of Cenozoic age and beyond (e.g. Westerhold et al., 2020; Zachos et al., 2008). At least three proxy methods exist for reconstructing the temperature of the deep ocean all of which are based on the geochemistry of benthic foraminiferal (Evans, 2021), namely, the stable isotope (the oxygen isotopic ($\delta^{18}O_b$) and clumped isotope composition (Δ_{47}) of their shells (e.g. Leutert et al., 2019; Marchitto et al., 2014)), and trace incorporation of metal impurities (Mg/Ca (Rosenthal et al., 1997)), each with their own advantages and disadvantages, discussed in supplementary text S2. For the purposes of this introductory discussion, we focus on benthic foraminifera $\delta^{18}O$ and the transformation equations of Hansen et al. (2013), hereafter abbreviated H13, because this approach is by far the most widely utilised, and because the benthic oxygen isotope stack has a far higher temporal resolution than any other deep ocean proxy.

H13 first calculate the temperature of the deep ocean (T_d) from $\delta^{18}O$, and then transform this into GMST. In the first step, the Cenozoic benthic $\delta^{18}O$ stack is divided into three portions on the basis

that i) all change in $\delta^{18}O_b$ can be ascribed to temperature prior to the growth of a major ice sheet on Antarctica at ~34 Ma, ii) that ice volume changes increasingly contribute to $\delta^{18}O_b$ when Earth is cooler, because there is a lower limit of the temperature of seawater. Specifically:

$$T_d = -4\ \delta^{18}O + 12 \quad \text{(Eq. 1.1)}$$

prior to 35 Ma (equation 3.1 of H13), m = 4°C ‰$^{-1}$

$$T_d = 5 - 8\ (\delta^{18}O - 1.75)/3 \quad \text{(Eq. 1.2)}$$

since 35 Ma and when $\delta^{18}O_b$ <3.25‰ (equation 3.5 of H13), m = 2.67°C ‰$^{-1}$, and

$$T_d = 1 - 4.4\ (\delta^{18}O - 3.25)/3 \quad \text{(Eq. 1.3)}$$

when $\delta^{18}O_b$ >3.25‰ (equation 3.6 of H13), m = 1.47°C ‰$^{-1}$. This latter equation in effect covers all of the Pleistocene and the glacial intervals of the late Miocene and Pliocene, and a slope of 1.47 is broadly similar to the canonical view that two thirds of the Pleistocene $\delta^{18}O_b$ signal can be ascribed to ice growth and decay (Raymo et al., 1989 and references therein). The result of applying Eqs 1.1-1.3 to the most recent version of the benthic foraminiferal $\delta^{18}O$ stack (Westerhold et al., 2020) is shown in Fig. 1A, alongside independent temperature proxies based on the Mg/Ca and clumped isotopic composition of benthic foraminifera.

With a deep ocean temperature record, H13 separate the Cenozoic into the Pleistocene, Pliocene, and pre-Pliocene. For the Pleistocene, H13 relate surface temperature (T_s) to the deep ocean by the following relationship:

$$T_s = 2\ T_d + 12.25 \quad \text{(Eq. 2.1)}$$

This is equation 4.1 of H13, which was rooted in a Holocene-LGM ΔT of 4.5°C, and a Holocene GMST of 14.15°C, requiring that GMST changed twice as quickly as the deep ocean given a deep ocean LGM-Holocene $\Delta\delta^{18}O_b$ of ~1.7‰ of which approximately two thirds is assigned to ice sheet decay. Then, in the Pliocene:

$$T_s = 2.5\ T_d + 12.15 \quad \text{(Eq. 2.2)}$$

Equation 4.2 of H13, which is the sensitivity required to match an "Early Pliocene mean temperature 3°C warmer than the Holocene" (Hansen et al., 2013). Beyond the Pliocene, H13 relate T_s to T_d in a 1:1 relationship, i.e.

$$T_s^{t=t} - T_s^{t=5.3} = T_d^{t=t} - T_d^{t=5.3} \quad \text{(Eq. 3)}$$

Where t=t is the time interval of interest, and t=5.3 is the base of the Pliocene. The result of applying the transformation Eqs 2.1, 2.2, and 3 to both the $\delta^{18}O$ and Mg/Ca-derived deep ocean temperature records (Fig. 1A) is displayed in Fig. 1B. The focus of the analysis presented here is on the warm intervals of the early Cenozoic, and therefore it is Eq. 3 which we most closely scrutinise.

A simple illustration of whether we might expect Eq. 3 to be accurate is shown in Fig. 2A, wherein surface area-weighted global ocean temperature is calculated for an arbitrary degree of tropical

and high latitude SST increase for an ocean planet. Specifically, a value for SST is assigned to the equator and poles, varying linearly as a function of latitude, from which GMST is computed (see supplementary text S1 for a comprehensive description). The contours in Fig. 2A show the ratio of the increase in GMST compared to the temperature of the deep ocean, assuming that deep ocean temperature is equal to surface area-weighted high latitude SST between 65-80°. Note that the position of these contour lines is independent of absolute temperature, depending only on the choice of latitude representative of deep water formation. The theoretical case of exactly no polar amplification is represented by the line with m = 1 (equal warming at 0° and 90°), which naturally coincides with the $\Delta GMST/\Delta T_{deep}$ contour equal to 1 as the high latitudes warm at an identical rate to the rest of the planet. All scenarios in which the high latitudes warm at a faster rate than the tropics are characterised by $\Delta GMST/\Delta T_{deep}$ relationships <1.

In order to determine the extent to which GMST could be overestimated by Eq. 3 within the context of this simple calculation, surface ocean proxy data from three key intervals within the early Palaeogene are overlain on Fig. 2A (the latest Paleocene (LP), Paleocene-Eocene Thermal Maximum (PETM) and the early Eocene climatic optimum (EECO)). These average values for each interval represent the mean of all data in the DeepMIP database (Hollis et al., 2019), available at https://www.deepmip.org/data, conservatively excluding planktonic foraminifera $\delta^{18}O$ data points impacted by diagenesis following Inglis et al. (2020). The change in tropical and high latitude temperature is calculated relative to modern (Locarnini et al., 2018), averaged over 0-30° and >60°N/S respectively. Two estimates of high latitude warming are shown, based on the high latitude SST proxy data and the temperature of the deep ocean, the latter calculated using the benthic foraminiferal oxygen isotope stack (Westerhold et al., 2020) and Eqs 1.1-1.3. We include this alternative indirect assessment of high latitude SST to determine the extent to which the analysis could be impacted by a potential summer bias in the high latitude SST proxy data (Hollis et al., 2012), which can be avoided via the assumption that the temperature of the deep ocean may more closely represent mean annual high latitude SST than the direct surface ocean proxy data (Evans et al., 2018). The results of this simple analysis confirm that, whichever dataset is used to constrain high latitude SST, all three intervals are characterised by high latitude warming ~2-5 times greater than in the tropics, as previously described (e.g. Hollis et al., 2019; Tierney et al., 2022). While we stress the simplicity of the approach used to calculate the position of the contours in Fig. 2A (e.g. ignoring for the moment that this is a water world), the data suggest that, in the absence of other processes driving the system in the opposite direction, the temperature of the deep ocean would be expected to rise at a rate of ~0.6-0.8 times that of the global surface ocean for this degree of polar amplification. If this simple analysis is applicable to Earth, then the corollary of this is that H13 likely overestimate Palaeogene GMST by approximately ~15-30% (less than ~0.2-0.4 times (1 – 0.6-0.8) as the 1:1 relationship (Eq. 3) is anchored to the base of the Pliocene).

There are of course several possible reasons that this analysis may be incorrect, and several processes that may mean that the key assumption of H13 (Eq. 3) is not biased by polar amplification/stratification; exploring whether or not such processes exist is the focus of the remainder of this contribution, and indeed, weakened polar amplification mechanisms in warm

climate states have been posited (Cramwinckel et al., 2018). An immediate indication that the approach of H13 may perform better than might be expected based on the simple reasoning outlined above is that GMST estimates based on the approach of H13 are in overall good agreement with several fully independent approaches based only on surface proxy datasets (Inglis et al., 2020). For example, EECO GMST estimates based on a climate model-derived transfer function (Farnsworth et al., 2019) or Gaussian processes regression (Inglis et al., 2020) agree with H13 to within 1.5°C, suggesting that no large bias results from applying Eq. 3 to the early Cenozoic, although we note that a weakness of comparison derived (in part) from climate model simulations is that most underestimate polar amplification (e.g. Burls et al., 2021; Evans et al., 2018; Lunt et al., 2012).

For H13 to accurately reconstruct GMST requires process(es) that either i) drive GMST increase at a faster rate than the average global ocean, and/or ii) bias the temperature of the deep ocean below that of mean annual high latitude SST. For these processes to result in a world that is well characterised by the assumption of Eq. 3, the magnitude of the effect must be exactly equal to the surface area-weighted difference between the change in temperature of the global relative to the high latitude ocean. This requirement was of course clear to H13, and indeed, alternative assumptions were tested in that study requiring an even greater counteracting force; namely that GMST increases at 1.5 times the rate of the deep ocean.

There are, to our knowledge, two key mechanisms that have the potential to drag GMST and deep ocean temperature onto a 1:1 relationship within the context of polar amplification and increased ocean stratification under past warm climate states:

1) While changes in global mean (near) surface temperature (e.g. Morice et al., 2012) and global mean sea surface temperature (GMSST) may be approximately equated over the instrumental period (Hansen et al., 2010), this is not the case when considering more extreme (Cenozoic) climatic change given the greater specific heat capacity of water compared to most other surface materials and a strongly different land-ocean evaporative flux and different response of that flux to warming (Henry and Vallis, 2022; Roderick et al., 2014). For these reasons, it is important to bear in mind that ΔGMSST and ΔGMST are not equivalent throughout much of the Cenozoic (see e.g. Fig. 1B; Gaskell et al., (2022)).
2) A bias in the season of deep water formation towards the winter (or possibly, an increased seasonal bias as GMST increases), could counteract the effect of amplified high latitude SST increase relative to the global mean. In the modern ocean, there is no strong seasonal trend in meridional overturning circulation (MOC) in the North Atlantic west of Greenland and a strong seasonal variation in MOC between Greenland and Scotland, although with a seasonal timing that varies between years (Wang et al., 2021). In contrast, the mixed layer depth (MLD) close to Antarctica is characterised by a strong seasonal variation (Pellichero et al., 2017), with sea ice-driven densification thought to play an important role in deep overturning (Pellichero et al., 2018). Given a likely

increase in high latitude SST seasonality under past greenhouse climate states (Hollis et al., 2012), it is at least possible that overturning was biased towards winter to a greater degree than at present during these times, although we note that a sea-ice related seasonality in deep water formation is unlikely to have been a relevant process for much of the Cenozoic.

A simple illustration of this from a theoretical point of view is shown in Fig. 2B. Here, the same calculation of the relationship between polar amplification and the ratio of GMST/deep ocean temperature as in Fig. 1A is shown, except with a GMST-dependent bias of SST during the season of deep water formation of 1/5°C (i.e., for each 5°C increase in GMST, the temperature of the surface ocean in the regions of deep water formation is biased by 1°C below that of mean annual high latitude SST). Comparing the recomputed contours to the same estimates of GMST for three key early Cenozoic intervals demonstrates that a relatively modest increasing seasonal bias in deep water formation is sufficient to bring these observations of tropical versus high latitude warming almost exactly in line with a GMST/deep ocean temperature ratio of 1. The implication of this is that, in absolute terms, the necessary seasonal bias in the temperature of the surface ocean in the regions of deep water formation is minor. For example, the EECO was characterised by a GMST ~15°C warmer than pre-industrial, thus requiring only a ~3°C difference between the temperature of subducting deep water and that of mean annual high latitude SST.

Determining the real-world applicability and magnitude of the two mechanisms listed above requires robust observational evidence for the relationship between deep ocean and GMST throughout the Cenozoic and a physical mechanistic basis. In order to provide this, we tackle the following four questions in the remainder of this contribution: 1) Do we know the Cenozoic evolution of deep ocean temperature sufficiently well for it to find utility as a proxy for GMST? (Sec. 3.1), 2) What is the empirical slope of the relationship between deep ocean temperature and GMST? (Sec. 3.2), 3) Are fully coupled climate models characterised by a similar relationship, and what can model data tell us about the mechanistic basis for the deep ocean-GMST relationship? (Sec. 3.3), and 4) Should the approach of H13 be revised in light of this analysis, and if so, how? (Sec. 4). In each case, we introduce the relevant methodology and datasets at the beginning of the section.

3. Results and discussion

3.1 How well do we know the Cenozoic evolution of deep ocean temperature?

Three key proxy methodologies exist for the temperature of the deep ocean (Evans, 2021), all of which are based on the geochemistry of the shells of calcifying benthic foraminifera: i) the magnesium to calcium ratio (Mg/Ca), ii) the oxygen isotopic composition, and clumped isotopic composition of the shell. Each of these has benefits/limitations summarised briefly in the supplementary materials (Text S2). It is important to note that all three proxies have nonthermal controls or present analytical challenges, particularly the extent to which secular changes in the

seawater Mg/Ca ratio impact the Mg/Ca proxy (e.g. Evans and Müller, 2012; Lear et al., 2015), the portioning of measured $\delta^{18}O$ changes between temperature, ice volume, and possible seawater pH (e.g. Meckler et al., 2022; Raymo et al., 2018; Rohling et al., 2021), and the relatively large analytical uncertainty and sample size requirements of Δ_{47} (e.g. Affek, 2012; Meckler et al., 2022).

A compilation of deep ocean temperature data (T_d) derived from these three independent techniques is shown in Fig. 1A. Overall, the datasets are characterised by a remarkable degree of consonance, with all three delineating a long-term cooling trend through the Cenozoic of ~15°C since the EECO and being characterised by several structural similarities, such as the relatively rapid cooling across the Eocene-Oligocene Transition (EOT) and after the Miocene Climactic Optimum (MCO) visible in all or most of the records. On the other hand, major discrepancies exist, most notably that the clumped isotope-derived temperature record is substantially warmer than either of the other proxies throughout much of the Cenozoic (by up to 7°C) and has structure in the early Eocene that is virtually entirely absent in the other proxy data (Meckler et al., 2022). While the $\delta^{18}O$-derived deep ocean temperature record is based on multiple assumptions (Eq. 1.1-1.3), the presence of a previously unidentified deep ocean cooling of ~10°C in the earliest Eocene would have major implications for our interpretation of traditional $\delta^{18}O$ datasets and our understanding of the Cenozoic evolution of $\delta^{18}O_{sw}$ if it is globally representative. Meckler et al. (2022) argue that $\delta^{18}O_b$ may be driven by coincident temperature variation and density-driven changes in $\delta^{18}O_{sw}$, while the overall much warmer Δ_{47} temperatures may imply previously unidentified long-term $\delta^{18}O_{sw}$ shifts related to (e.g.) climactically-driven changes in groundwater storage such that the canonical assumption of bulk ocean $\delta^{18}O_{sw}$ = -1‰ in an ice-free world (Zachos et al., 1994) may often not apply. In contrast, the Mg/Ca-derived record is not systematically offset from either of the other proxies across the Cenozoic, showing good agreement with the clumped isotope data in the Oligocene and Neogene (Fig. 1A) but suggesting substantially cooler temperatures than the clumped isotope record during the mid-late Eocene. Mg/Ca-derived temperature is additionally substantially warmer than that based on $\delta^{18}O$ in the Paleocene, which could, for example, be driven by the lack of an accurate seawater Mg/Ca record for this interval. It is also important to note that the Mg/Ca and $\delta^{18}O$ datasets are characterised by very different resolutions, such that the apparent agreement between the two proxies during the PETM is an artefact of the way the Mg/Ca record has been smoothed.

While it is beyond the scope of this contribution to reconcile all the aforementioned proxy-proxy offsets, reconstructing GMST from deep ocean proxy data with a useful accuracy is contingent on understanding these discrepancies, such that this issue deserves urgent attention. In Sec. 4 we show, as a starting point, how the majority of the Cenozoic Δ_{47} and $\delta^{18}O$ data may be reconciled with each other, therefore also bringing $\delta^{18}O$ and Mg/Ca into agreement during the Neogene. We also note that, notwithstanding the importance of the Mg/Ca data compilation and analysis by Cramer et al. (2011), several aspects of that data analysis require revision, particularly in light of new information regarding the Cenozoic evolution of seawater Mg/Ca and updated benthic foraminiferal Mg/Ca temperature calibrations (Evans et al., 2018; Lear et al., 2015); revisiting the Paleogene portion of the Mg/Ca dataset with these advances in mind may help to resolve the

Eocene Mg/Ca-Δ_{47} offset. Irrespective, the central hypothesis to be tested here is that GMST and T_d are characterised by a linear 1:1 relationship. Fortunately, doing so is reasonably insensitive to the deep ocean temperature proxy discrepancies (Fig. 1) because the method of H13 anchors this 1:1 relationship to the base of the Pliocene (Eq. 3). Given that the clumped isotope record is warmer than the $\delta^{18}O$ transformation (Eq. 1.1-1.3) throughout most of the Cenozoic, the outcome of assessing this relationship empirically via the combination of deep ocean and surface temperature datasets (Sec. 3.2) does not greatly depend on the choice of deep ocean proxy data.

3.2 Empirical evidence for the relationship between deep ocean and global mean surface temperature

In order to assess whether quantitative, independent proxy data support the notion of a linear, 1:1 relationship between T_d and GMST prior to the Pliocene, we combine the deep ocean temperature records described in Sec. 3.1 (Fig. 1A) with curated data compilations from five well-studied intervals: the last glacial maximum (LGM), Pliocene (mid-Piacenzian) Warm Period (PWP), and three early Cenozoic warm intervals described in Sec. 2 and Fig. 2 (the LP, PETM, and EECO).

Global mean surface temperature for each interval is based on the following datasets: the LGM GMST used here is that of the data-model assimilation exercise of Osman et al. (2021), who estimated a LGM-pre industrial (PI) ΔT of 7.0±1.0°C, i.e. LGM GMST = 6.9°C based on a PI GMST = 13.9°C. The GMSST of the PWP (mid-Piacenzian; 3.264-3.025 Ma) was estimated using the surface area-weighted mean of the PRISM3/4 SST dataset (Dowsett et al., 2016, 2013), which is 18.7°C. Alternatively using the alkenone-only GMSST reconstruction of McClymont et al. (2020) yields 17.2°C for a narrower interglacial (KM5c, 3.2 Ma) and would therefore result in pre-Pliocene reconstructions 1.5°C lower when using this as an anchor. The early Cenozoic GMS(S)T estimates were taken from the DeepMIP data compilation and associated GMST analysis (Hollis et al., 2019; Inglis et al., 2020), with GMST/GMSST estimated in five different ways using surface ocean and terrestrial temperature estimates. In addition, we provide a new estimate based on a combination of SST data from the mid/low latitudes and deep ocean temperature as a proxy for high latitude SST avoiding potential seasonal bias (Evans et al., 2018; Hollis et al., 2012; Inglis et al., 2020), with GMST equal to the surface area-weighted mean in each latitudinal band (0-30, 30-65, >65°). In this latter case, the estimates of deep ocean and GMSST are not fully independent of each other, although we note that they agree well with four other GMS(S)T calculation techniques that are independent of the deep ocean temperature data (Inglis et al., 2020). These GMS(S)T estimates are compared to $\delta^{18}O$ and Mg/Ca-derived T_d in Fig. 3, calculated using the mean of all data within the sampled interval in all cases, using the H13 transformation equations in the case of $\delta^{18}O$ (Fig. 1B) and Eq. 7a of Cramer et al. (2011) in the case of Mg/Ca. The only exception to this is the LGM, for which we use the deep ocean temperature estimate of Adkins et al. (2002). An estimate of the relationship between T_d and GMST based on clumped isotope deep ocean temperatures is not given here as a result of the sparsity of data in certain key intervals, but note that a comparison between Δ_{47}-derived T_d and GMST is given by Goudsmit-Harzevoort et al. (2023).

The data compilation shown in Fig. 3 unavoidably combines GMST (PI, LGM, some early Cenozoic estimates) and GMSST reconstructions (PWP, some early Cenozoic estimates), such that caution is required in extrapolating between them. For this reason, the data compilations described above are compared to the T_d-GMST relationship of H13 (Eqs 1.1-3) as well as a data-derived T_d-GMSST relationship (Gaskell et al., 2022). Indeed, coupled climate models consistently predict that GMST and GMSST diverge at GMST <~20-25°C (Haywood et al., 2020; Lunt et al., 2021; Valdes et al., 2021) but are broadly similar above this, discussed in more detail in Sec. 3.3. This is in agreement with a comparison of the proxy-based GMSST analysis of Gaskell et al. (2022) with H13, which suggests coincident GMST/GMSST at very high GMST and a divergence of the two below ~20°C (Fig. 1A; 3A). If correct, this suggests that the Pliocene-Eocene portion of the analysis shown in Fig. 3 should be limited to the relationship between T_d and GMSST, given that the PWP dataset contains only SST estimates, or that the 1:1 line anchored to the Pliocene should be translated down the y axis by several °C. In the absence, to our knowledge, of a true PWP GMST estimate based on a comprehensive terrestrial and marine proxy data compilation, and to avoid complications and uncertainties associated with correcting GMSST to GMST (see Sec. 4) we initially approach the proxy data analysis with both possibilities in mind.

Oxygen isotope-based deep ocean temperature (Fig. 3A): Anchoring a 1:1 GMST-T_d relationship to the mid-Piacenzian $\delta^{18}O$-derived deep ocean and PRISM SST datasets defines a 17.2°C offset between the two and thus a LP, PETM, and EECO GMST of 28.4, 30.8, and 35.4°C respectively (given by the y axis location of the black dashed line at the respective T_d for these intervals). These estimates are substantially warmer than the majority of the independent GMST estimates for the EECO and PETM with the exception of the surface-area weighted SST-derived estimate of this study (solid green EECO datapoint in Fig. 3A), and the majority of the LP estimates, which fall within 2°C of the 1:1 line. In contrast, the transformation equations of H13 result in early Paleogene GMST estimates ~3°C cooler than the 1:1 line anchored to the PWP, and overall excellent agreement between the DeepMIP database GMST and T_d-derived estimates (red line in Fig. 3A). As discussed above, a likely reason for this is that the Pliocene anchor represents PWP GMSST rather than GMST, such that earlier Cenozoic GMST derived from this may be overestimates (black dashed line in Fig. 3A). The agreement between the T_d-GMSST relationship of Gaskell et al. (2022) and the PWP data point (blue line and red data point in Fig. 3A) adds support to this caveat. Either way, minor discrepancies exist, for example the EECO GMSST estimate of this study (solid green symbol; Fig. 3) is ~3°C warmer than the T_d-GMSST relationship of Gaskell et al. (2022). Understanding whether or not these offsets imply (e.g.) a state-dependent GMST-T_d relationship remains challenging given the certainty with which deep-time GMST can be independently estimated from surface proxy datasets (Anagnostou et al., 2020; Inglis et al., 2020) and should be the subject of future research and data compilation efforts. Nonetheless, the proxy data analysis shown in Fig. 3A is consistent with both the GMSST-T_d and GMST-T_d relationships (Gaskell et al. 2022; H13), especially in the latter case if the PWP data point is considered to represent GMSST and thus overestimates GMST.

As an aside, we note that the LGM GMST estimate of Osman et al. (2021) is cooler than that predicted by the $\delta^{18}O$ transformation of H13, requiring a steeper Pleistocene GMST-T_d slope (Eq.

2.1), which directly follows from the greater Holocene-LGM ΔT than that utilised by H13 (7 cf. 4.5°C).

Mg/Ca-based deep ocean temperature (Fig. 3B): Compared to using the $\delta^{18}O$ transformation equations of H13, the Mg/Ca-derived T_d-GMST relationship differs principally in that the Mg/Ca PWP T_d is ~3°C warmer, whereas the early Paleogene deep ocean temperatures are broadly not (Fig. 1A). This has the effect of shifting the 1:1 GMST-T_d relationship to higher T_d for a given GMST, bringing the EECO and PETM GMST estimates, especially those of this study, into excellent agreement with the deep ocean-based estimate (Fig. 3B). In contrast, the LP is offset from the 1:1 line by >5°C, which is driven by the very high Mg/Ca deep ocean temperatures in the late Paleocene compared to those derived from $\delta^{18}O$, in contrast to the EECO, where the two proxies are in good overall agreement. The unexpected nature of the structure of the Mg/Ca temperatures in the earliest Cenozoic suggests that this is probably an artefact of the Mg/Ca data or transformation, given that in this analysis the LP and PETM have similar T_d but were clearly characterised by very different climate states (Dunkley Jones et al., 2013; Penman et al., 2014). The reason for this is likely routed in either the seawater Mg/Ca correction (very little data exist for the Paleocene, none of which was available at the time these Mg/Ca deep ocean temperatures were calculated (see Cramer et al., 2011; Evans et al., 2018b; Gothmann et al., 2015)), or suggests a diagenetic issue with the Paleocene deep ocean Mg/Ca data. Assuming the PWP GMST is an overestimate (see discussion above) would alternatively suggest that T_d underestimates PETM (±EECO) GMST as the red data point and black dashed line anchored to it in Fig. 3B would shift down the y axis, possibly arguing for a relationship between GMST and T_d with a slope >1. Alternatively viewing the PWP datapoint as GMSST and comparing to the early Paleogene GMSST estimates (this study; solid symbols in Fig. 3B) would constrain a pre-Pliocene T_d-GMSST slope of ~1, substantially steeper than that derived by Gaskell et al. (2022). That is, the Mg/Ca-derived T_d analysis cannot be fully reconciled with both the transformation equations of H13 and Gaskell et al. (2022), although the $\delta^{18}O$-derived relationship of that latter study crosses the 1:1 line at a GMSST approximately equidistant between the PWP and EECO, such that the datasets may nonetheless fall within uncertainty of each other.

In conclusion, irrespective of which deep ocean temperature dataset is used, and whether the surface temperature estimates based on the data compilations utilised here are considered to represent GMST or GMSST (or both, in warmer climate states), the above data analysis is consistent with the notion of an approximate 1:1 relationship between T_d and GMST, as proposed by Hansen et al. (2013). In addition, reframing parts of this analysis in terms of GMSST provides support for a T_d-GMSST slope of 0.73 (Gaskell et al., 2022; see the coloured solid data points in Fig. 3A).

3.3 Constraints from fully coupled climate models

In section 3.2 we show that there is good empirical evidence that the central assumption of the pre-Pliocene GMST estimate of H13 is a reasonable approximation. To mechanistically understand

why this is the case, we interrogate the output of two sets of Paleogene coupled climate model datasets: 1) the DeepMIP set of model simulations (Lunt et al., 2021, 2017), which incorporates eight climate models run under different pCO_2 but otherwise similar (early Paleogene) boundary conditions between models, and identical boundary conditions within a set of model simulations, plus associated PI controls, and 2) the Cenozoic portion of the Phanerozoic HadCM3 simulations (Valdes et al., 2021), which includes 12 simulations (1 per Stage) with varying palaeogeography and other boundary conditions, run under two (broadly similar) pCO_2 within each time slice. These latter simulations were not part of DeepMIP. Both sets of model output are described in detail elsewhere, including: the experimental design and rationale (Lunt et al., 2021; Valdes et al., 2021, and references therein), the degree to which the deep ocean has reached equilibrium (Zhang et al., 2022), the location(s) of deep water formation (Zhang et al., 2022), and the spatial heterogeneity in modelled deep ocean temperature (Goudsmit-Harzevoort et al., 2023; typically <1°C). Those analyses are not repeated here, with the exception of the key regions of deep-water formation, which we interrogate when comparing high latitude SST to T_d, and the relationship between GMST and T_d, which was explored in detail by Goudsmit-Harzevoort et al. (2023), but is expanded upon here and compared in detail to the simulations of Valdes et al. (2021). The majority of the simulations in both sets are considered to have a reached a reasonable degree of equilibrium with respect to the deep ocean (<1°C drift in the volume-integrated mean ocean temperature; Valdes et al. (2021)), with the exception of the 9× CO_2 CESM simulation (~1.5°C in the final kyr; Fig. S2 of Zhang et al. (2022)). This latter data point is clearly an outlier (see below), and while we include it in all relevant figures, we exclude it from any regression analysis on the basis that deep ocean temperature is likely underestimated in this simulation. In all cases in the discussion below we define T_d as equal to the mean of all temperature data below 3000 m.

The two sets of model simulations have their own advantages and limitations. The DeepMIP output allows the role of CO_2 to be more readily separated from other factors, as all other boundary conditions were held constant with the exception of the PI controls. Conversely, key boundary conditions such as paleogeography were modified for each time slice in the HadCM3L simulations of Valdes et al. (2021), making it more challenging to pinpoint the factors driving a given output, but arguably makes them a better test of the relationship between T_d and GMST, given that (e.g.) the Cenozoic paleogeographic changes are incorporated.

3.3.1 Model deep ocean temperature versus GMS(S)T

As shown by Goudsmit-Harzevoort et al. (2023), the DeepMIP simulations are characterised by a T_d-GMST slope close to 1 based on the entire ensemble (excluding the 9× CO_2 CESM simulation and PI controls), with m = 1.026 (Fig. 4A). Anchoring a 1:1 line to the mean of the lowest CO_2 simulation for each model with Eocene paleogeography (in order to broadly follow the assumption of Hansen et al., 2013) demonstrates that all simulations fall within ±2°C, with the majority falling within ±1°C (mean average error = 0.82°C). Thus, a 1:1 relationship between T_d and GMST appears to be a robust assumption based on both proxy data (Sec. 3.2) and climate models. The HadCM3L simulations with variable boundary conditions (hereafter $HadCM3L_{V21}$ to distinguish them from

those performed as part of DeepMIP) paint a similar picture, with most falling with ±2°C of a 1:1 line anchored to the mean of the two simulations at 3 Ma (Fig. 4C). The exception to this are the Miocene simulations, which are characterised by GMST 0-3°C warmer than PI but mean deep ocean temperatures up to 2°C cooler, irrespective of which CO_2 scenario is used. This yields an overall T_d-GMST slope much lower than the DeepMIP simulations (m = 0.715), although excluding these from the analysis results in a slope much closer to unity (m = 0.858). The HadCM3L$_{V21}$ Miocene anomaly appears to be driven by a shift in the dominant region of deep-water formation from the North Atlantic to the Southern Ocean compared to the PI control, resulting in cooler deep ocean temperatures. This is likely caused by salinity-driven changes in density driving a slowdown in N. Atlantic overturning such that deep water formation shifts to the fresher but colder Southern Ocean, yet CO_2 is insufficiently high to drive increases in GMST. Conversely, T_d is higher in the HadCM3L$_{V21}$ Pliocene simulations because the N. Atlantic remains the dominant region of deep water formation. Of the 86 simulations that we consider in our analysis, it is only the six HadCM3$_{V21}$ from the Miocene that show a substantial deviation from a 1:1 T_d-GMST relationship, such that climate models with both constant and temporally varying boundary conditions run at different pCO_2 overall provide strong evidence in favour of the 1:1 T_d-GMST hypothesis. Nonetheless, given that the HadCM3L$_{V21}$ simulations are arguably a better test of whether T_d and GMST are directly related throughout the Cenozoic (as opposed to when pCO_2 changes within a given interval), as paleogeography related phenomena are capable of driving large changes in GMST independent of pCO_2 (Caballero and Huber, 2013), the possible breakdown of this relationship in the Miocene is a key target for future research. That there is limited proxy evidence for a similar Miocene and PI deep ocean temperature (Fig. 1A), possibly suggests an issue with the Miocene model data, and highlights the need for further Miocene pCO_2 estimates (Valdes et al., 2021).

Both sets of model simulations are characterised by a GMSST-T_d and GMSST-GMST relationship with a slope substantially lower than 1, with m = 0.870 and 0.456 in the DeepMIP and HadCM3L$_{V21}$ simulations respectively (GMSST-T_d; Fig. 4) and m increasing to 0.587 in the latter case if the Miocene simulations are excluded. These slopes bracket the empirical ($\delta^{18}O$-derived) relationship of Gaskell et al. (2022), which has m = 0.73, with the DeepMIP suite of models conspicuous in being consistently characterised by a steeper relationship and GMSST consistently (~2-3°C) warmer than the data suggest for a given T_d at pCO_2 >2× PI (Fig. 4B). While the HadCM3L$_{V21}$ dataset is characterised by a lower slope than the proxy data-based estimate, almost all simulations remain similarly offset to higher GMSST. In general, the model-data GMSST-T_d disagreement is likely driven by the model difficulty in capturing the magnitude of polar amplification implied by the proxy data (e.g. Evans et al., 2018b; Kiehl and Shields, 2013; Lunt et al., 2012; Sagoo et al., 2013; Hollis et al., 2019; Lunt et al., 2021), which results in a lower degree of high latitude and therefore deep ocean warming for a given global GMSST increase, rather than an issue with the analysis of Gaskell et al. (2022). However, we note that this cannot explain the good agreement between the early Cenozoic HadCM3L$_{V21}$ simulations with Gaskell et al. (2022) (Fig. 4D).

3.3.2 Mechanistic basis for a ~1:1 deep ocean-GMST relationship

Both model and proxy data are in overall agreement that GMST and the temperature of the deep ocean are linearly related with a slope close to 1, supporting one of the central assumptions of Hansen et al. (2013). We next revisit the two key reasons that this relationship might emerge, outlined in Sec. 2, despite polar amplification and the stratification of the ocean in warm climate states, using both the DeepMIP and HadCM3L$_{V21}$ simulations. That is: 1) whether deep water formation becomes increasingly seasonally biased in warm climate states, and 2) if a faster rate of land versus ocean warming counteracts the effect of polar amplification to result in a 1:1 T_d-GMST relationship.

Is deep water formation seasonally biased in warm climate states? In order to address this question, we examine the relationship between T_d and high latitude SST (SST_{HL}) in the regions of deep-water formation. The sensitivity of this analysis to three different methodologies was explored, in which SST_{HL} was variously calculated as:

1. Mean annual SST in all regions of the ocean likely to be important for deep-water formation, that is, the N. Atlantic, N. Pacific, and Atlantic, Pacific, and Indian sectors of the Southern Ocean. See Zhang et al. (2022) and Fig. 5 for the definition of these boxes.
2. Winter SST in only the boxes relevant for deep-water formation, on a simulation-specific basis. We define winter as June, July, and August in all southern hemisphere boxes, January and February in the N. Pacific, and March, April, and May in the N. Atlantic to reflect that maximum overturning occurs shortly after winter in that region (Wang et al., 2021). The choice of relevant box(es) for a given simulation was based on the region of maximum mixed layer depth, detailed in Tab. S1. In cases where this is ambiguous, such as the 9× CO_2 CESM simulation (Fig. 5), we use the boxes from the closest CO_2 simulation which clearly shows deep water formation as evidenced by the mixed layer depth (mld).
3. The SST in the grid cells that have a simulation-specific mld at least 90% as deep as the global seasonal maximum, only considering grid cells >50°N/S (Valdes et al., 2021). T_d was compared to both mean annual and winter SST in the mean of all grid cells meeting these criteria.

In all cases, SST is based on the mean of all relevant grid cells averaged over 0-100 m. A sensitivity analysis was performed to understand the impact of in/excluding the Arctic, which may at times have been disproportionately fresher and warmer than other high latitude regions (Brinkhuis et al., 2006). Doing so has no significant impact on any aspect of the data analysis presented below.

The results of the first two of these analyses is shown in Figs 6 and 7 for the DeepMIP and HadCM3L$_{V21}$ simulations respectively, in all cases anchored to i) the simulation with the lowest pCO_2 but Eocene paleogeography in the case of the DeepMIP simulations, as the closest possible representation of Pliocene-like conditions, and ii) to the Pliocene in the case of HadCM3$_{V21}$, again, because the principal aim of this study is to test the pre-Pliocene T_d-GMST assumption of Hansen et al. (2013). Note that while a similar interrogation of the relationship between deep ocean

temperature and GMST was performed by Goudsmit-Harzevoort et al. (2023), the analysis presented here differs in that we interrogate all simulations by determining the relative change from the Pliocene or that with the lowest CO_2 and paleo boundary conditions, although we replot the data in absolute terms for comparison (e.g. Fig. 6C,F). This is an important distinction, given that it is the *relative* change in T_d and GMST prior to the base of the Pliocene that we are principally interested in, and doing so avoids potential bias derived from model-specific skill in capturing past climate states.

In the case of the DeepMIP models, mean annual SST (MASST) in all high latitude boxes rises at a faster rate than GMSST in response to CO_2 (ensemble m = 0.88; Fig. 6A), as expected given enhanced poleward heat transport in warm climate states (Kelemen et al., 2023). Limiting the analysis to include only winter SST in the relevant deep water boxes (Fig. 5; Tab. S1) is a simple method of removing the summer bias in SST_{HL} in the comparison. Indeed, doing so results in an increase in the whole ensemble GMSST-SST_{HL} slope (m = 0.96; Fig. 6B), albeit with a greater degree of variance around the least squares regression (the uncertainty in the slope increases from 0.018 to 0.038). A similar picture emerges when assessing the change in T_d as a function of SST_{HL} (Fig. 6D,E), which is overall characterised by a slope close to 1 (m = 1.05) such that $\Delta SST_{HL} \approx \Delta T_d$, implying that changes in deep ocean temperature are directly coupled to the high latitude surface ocean in all simulations. As before, an increased gradient (to m = 1.13) and variance is observed when winter SST in the filtered high latitude boxes is used (0.034 versus 0.074), implying slightly worse agreement between SST_{HL} and T_d in absolute terms across the ensemble (Fig. 6F). A faster rate of temperature increase in the deep ocean than high latitude SST (Fig. 6E) is physically implausible and implies a limitation of this simple analysis (e.g. the inclusion of grid cells in the high latitude boxes (Fig. 5) that do not contribute to overturning), or that a portion of deep water is formed outside of the months and/or boxes considered here, explored in more detail below. Overall, however, this analysis must mean that seasonally biased deep water formation is important in offsetting T_d from mean annual SST_{HL}, strongly arguing for this as a mechanistic cause of a 1:1 T_d-GMST relationship.

In the case of the $HadCM3L_{V21}$ simulations, the results of the analysis is broadly similar in that accounting for the seasonality and relevant location of deep water formation results in a ΔGMST-ΔSST_{HL} slope marginally close to 1 compared to a similar analysis based on MASST in all high latitude boxes (Fig. 7A,B). While this adds further model-based support for a 1:1 relationship between GMST and SST_{HL}, these simulations differ from the DeepMIP set in that the unfiltered analysis (Fig. 7A) is characterised by a slope marginally >1, i.e. GMSST broadly increases at a faster rate than SST_{HL}. This is not the case for the HadCM3 DeepMIP simulations, highlighting the potential impact of palaeogeography over CO_2 alone, although we also note that the version of HadCM3L utilised to produce the DeepMIP simulations is one of several models less able to capture the proxy data-derived degree of Eocene polar amplification in several deep-time warm intervals (Burls et al., 2021; Evans et al., 2018; Lunt et al., 2021). The $HadCM3L_{V21}$ simulations are also characterised by a T_d-SST_{HL} slope >1, irrespective of whether the relevant boxes and season of deep water formation are taken into account (Fig. 7D,E), and absolute deep ocean temperature

substantially offset from winter SST in the boxes relevant for deep water formation (Fig. 7F). This is also not the case for the DeepMIP HadCM3 simulations and is mostly easily explicable a limitation of our analysis, likely suggesting the inclusion of cooler high latitude grid cells that are not relevant for deep water formation and/or that there is an important contribution to deep water formation beyond the core winter months.

Both the DeepMIP and HadCM3L$_{V21}$ simulations overall provide strong evidence that deep ocean temperature, GMSST, and high latitude SST are tightly, linearly linked to each other, especially when a winter season bias in deep water formation is accounted for. While the degree of divergence from 1:1 high latitude seasonal SST-T_d and GMSST-T_d relationship is small in both sets of simulations (within ±5°C in almost all cases, e.g. Fig. 6B,E; Fig. 7B, E), we explore how much of the remaining variance in the data analysis is a result of the approach of averaging data across large high latitude ocean boxes (Fig. 5), which is the simplifying approach of both our analysis thus far and previous studies (Goudsmit-Harzevoort et al., 2023; Valdes et al., 2021; Zhang et al., 2022). To assess this, Fig. 8 shows the same analysis of the DeepMIP simulations as in Fig. 6, except that only grid cells with a mixed layer depth (mld) within 90% of the global seasonal maxima were considered (approach 3 above), thus avoiding possible bias from the inclusion of grid cells within the broad boxes shown in Fig. 5 that are not relevant for deep water formation. For example, in the 3×CO_2 CESM simulation, this approach excludes coastal and more northerly grid cells in the Weddell Sea, as well as the eastern portion of the Indian Ocean sector of the Southern Ocean box (Fig. 5).

This more nuanced analysis highlights a stronger degree of polar amplification in both mean annual and seasonal SST in grid squares with the deepest seasonal mld relative to the all grid cells within the high latitude boxes discussed previously. For example, there is a shallower slope between ΔSST_{HL} and ΔGMST in the model ensemble relationship between GMSST and mean annual SST_{HL}, which is characterised by a slope that is reduced to 0.81±0.05 in the mld analysis compared to m = 0.88±0.02 when using all grid cells in the high latitude boxes (cf. Fig. 8A, Fig. 6A). A similar reduction in slope is observed when comparing the seasonal SST in the grid cells characterised by the deepest mld and limiting the analysis to include only grid cells in the hemisphere in which deep water formation dominantly occurred, with a reduction in slope from 0.96±0.04 to 0.89±0.04 (Fig. 8B, Fig. 6B; with CESM being a notable exception). Therefore, winter SST in these grid cells warms faster than global average MASST, which means that overall, while an increasing seasonal bias in deep water formation is a key process that results in an approximate 1:1 T_d-GMSST relationship (all DeepMIP simulations fall within ±5% of the 1:1 line; Fig. 8B), a polar amplification signal is present in subducting water, such that a seasonal bias in deep water formation cannot be the only process resulting in a 1:1 GMST-T_d relationship (Fig. 4A). Unsurprisingly, the mld analysis fully resolves the discrepancy between SST_{HL} and T_d observed previously (Fig. 6E). Considering only the relevant grid squares results in a slope of 0.99±0.07 (Fig. 8D), i.e., the two parameters have the same value.

The above analysis demonstrates that an increasing seasonal bias in deep water formation offsets polar amplification in the regions of deep water formation to a substantial degree, but is, alone, insufficient to mechanistically explain a 1:1 T_d-GMST relationship. The other key factor, as noted by Hansen et al. (2013) and Goudsmit-Harzevoort et al. (2023) is that land surface air temperature are more sensitive to CO_2 than SST. This is the case to a greater degree in the HadCM3L$_{V21}$ simulations than the DeepMIP ensemble (GMSST versus GMST m = 0.84 and 0.72 respectively, Fig. 4E,F), implying that non-CO_2 boundary conditions can be important in modulating this slope given that isolating the DeepMIP HadCM3 and HadCM3L simulations demonstrates that these are characterised by a steeper slope than HadCM3L$_{V21}$ (m = 0.795 versus 0.716). Nonetheless, in the case of the DeepMIP ensemble, the slopes between ΔGMSST and ΔSST$_{HL}$ (accounting for seasonality in the relevant high latitude boxes) and GMSST-GMST are almost identical (m = 0.89 and 0.84, respectively; Fig. 8B, Fig. 4E), which is the explanation for a 1:1 T_d-GMST relationship as hypothesised by H13 and the reason that this is an emergent model property (Fig. 4A).

We stress that while previous analyses have reached the same or similar conclusions (Goudsmit-Harzevoort et al., 2023), the key assumption of H13 was that this is the case in the earlier Cenozoic when GMST and T_d are *anchored to the base of the Pliocene*, i.e. prior to the strong modulation of this relationship by ice sheet growth and ice-sheet climate feedbacks. Thus, while it is strongly encouraging that this conclusion has now independently been reached several times, it is only by performing the analysis in the way presented here, and assessing whether it is the case in model simulations with both variable palaeogeography and constant palaeogeography but variable pCO_2 that we can mechanistically understand whether or not this relationship is likely to have been the case for the entirety of Cenozoic prior to ~5 Ma.

4 Reformulation of the deep ocean-GMST relationship

Climate model simulations and proxy data are in remarkable full agreement (within uncertainty) that the temperature of the deep ocean and GMST are characterised by a 1:1 relationship prior to the Pliocene. Crucially, given that this is empirically the case, this lends strong support to the notion that a reconstruction of deep ocean temperature is a reliable proxy of GMST irrespective of whether the model-derived mechanistic basis for this relationship discussed in Sec. 3.3.2 is correct. Nonetheless, the GMST approach of H13 requires revision, particularly in light of our greatly improved understanding of i) the Cenozoic evolution of T_d, including a more thorough grasp of the nonthermal controls on some key temperature proxies, and ii) the Cenozoic evolution of continental ice volume, i.e., that the assumption of ice-free conditions before the base of the Pliocene is no longer tenable (Lear et al., 2015; Leutert et al., 2021; Rohling et al., 2022). Focusing on the benthic oxygen isotope stack ($\delta^{18}O_b$) because the temporal resolution of the dataset is unparalleled by the other proxy datasets (Fig. 1), we explore whether $\delta^{18}O_b$-derived temperatures can be reconciled with constraints from clumped isotope analysis of benthic foraminifera, and how our improved understanding of sea level variation impacts GMST estimates based on these data. In order to do so, we i) revisit the ice volume/sea level component of $\delta^{18}O_b$ using the analysis of

Rohling et al. (2022), ii) explore the impact of a pH/[CO_3^{2-}] correction on $\delta^{18}O_b$, and iii) rescale the resulting deep sea temperature record using three intervals with reasonable constraints on both deep ocean and GMST (the LGM, present-day, and PWP).

Sea level: An extremely comprehensive analysis of the sea level/ice volume contribution to $\delta^{18}O_b$ is available (Rohling et al., 2022), which uses a process-based model to determine the nonlinear relationship between $\delta^{18}O_b$ and sea level. This nonlinearity largely results from the relationship between GMST and ice volume (e.g., the absence of ice above a certain GMST) and the change in mean ice sheet $\delta^{18}O$ as a function of total ice volume (Rohling et al., 2022, 2021; Spratt and Lisiecki, 2016). Here, we use the median of the boot-strapped Monte Carlo results of the preferred process-based model of Rohling et al. (2022), from which we derive a sea level-free $\delta^{18}O_b$ record by converting the deep sea temperature of that study back to $\delta^{18}O_b$ simply by dividing by -0.25‰°C^{-1}. As Rohling et al. (2022) studied the interval 0-41 Ma, we extend the record to the entirety of the Cenozoic by appending the remainder of the $\delta^{18}O_b$ record of Westerhold et al. (2020), assuming no ice volume contribution to $\delta^{18}O_b$ before 41 Ma (e.g. Scotese et al., 2021).

pH effect on $\delta^{18}O_b$: A seawater carbonate chemistry effect on $\delta^{18}O_b$ has been found for both species of foraminifera (*Orbulina universa* and *Globigerina bulloides*) for which sufficient data are available to make an assessment (Bijma et al., 1999; Spero et al., 1997), as well as coccolithophores and calcareous dinoflagellates (Ziveri et al., 2012), and inorganic calcite (McCrea, 1950). While there is no direct evidence for a similar impact on the oxygen isotopic composition of the shells of benthic foraminifera, we advocate for a correction because it is a ubiquitous feature of all calcitic plankton studied so far, and has a strong basis in theory, being rooted in the pH-dependent speciation of dissolved inorganic carbon (Zeebe, 1999). We nonetheless stress that correcting $\delta^{18}O_b$ data for past changes in seawater carbonate chemistry remains fraught with uncertainty because the slope of the relationship strongly differs between both foraminifera species studied thus far (by a factor of ~2), which has a large impact on the resulting correction when considering large whole-ocean changes in pH (Evans et al., 2016). Here, we explore a correction using the theoretical slope between pH and $\delta^{18}O$ across the pH range 7-9 (Eq. 2 of Zeebe (1999)), which covers the possible range of past ocean pH variation, across which a linear approximation suffices. Doing so yields a pH-$\delta^{18}O$ slope of -1.50‰ per pH unit (see supplementary Text S3), which is intermediate between the two planktonic foraminifera species for which data are available (-0.89 and -2.51‰ per pH unit). While this control on $\delta^{18}O$ is often thought of as a carbonate ion effect (e.g. Gaskell et al., 2022; Spero et al., 1997), it is more appropriately mechanistically ascribed to pH because this is, in effect, the dominant control on seawater [HCO_3^-]/[CO_3^{2-}], and the effect of seawater carbonate chemistry on carbonate $\delta^{18}O$ occurs via the differential fractionation factor between water and these DIC species (Zeebe, 1999). Parameterising the seawater carbonate chemistry effect on $\delta^{18}O$ as a function of pH additionally has the advantage that a direct proxy for seawater pH is available from measurements of the boron isotopic composition of foraminifera (Anagnostou et al., 2020; Foster and Rae, 2016; Hönisch et al., 2012; Penman et al., 2014). In order to apply a pH correction, we fit a smoothing spline to the benthic foraminifera $\delta^{11}B$-derived pH record (Greenop et al., 2014; Meckler et al., 2022). The available benthic foraminifera-derived pH dataset is low-resolution (31

datapoints spanning the last ~60 Ma) and contains no data between the mid-Eocene and Miocene. As such, our T_d and derived GMST record will require revision as more data become available, and it is possible or likely that the details of the records presented here contain artefacts related to the long-term smooth applied to the pH data.

The resulting Cenozoic T_d reconstruction is shown in Fig. 9A, with that of Hansen et al. (2013) and independent estimates of deep ocean temperature for comparison (Fig. 1A). Notwithstanding the uncertainties in the approach driven by the sparsity of deep ocean pH data, pH correcting $\delta^{18}O_b$ following the ice volume deconvolution of Rohling et al. (2022) results in a Cenozoic $\delta^{18}O_b$-derived T_d record which agrees well with the majority of the clumped isotope data, such that this revised analysis of the $\delta^{18}O$ data resolves much of the pre-existing discrepancy between the $\delta^{18}O$ and Δ_{47} records (cf. Meckler et al. (2022); Westerhold et al. (2020)), especially when the $\Delta_{47}/\delta^{18}O$ reanalysis of Daëron and Gray (2023) is considered (Fig. 9A). The magnitude of the Cenozoic T_d decrease between the EECO and late Pleistocene (~17°C) is indistinguishable between the proxies, with major discrepancies remaining only in the early Eocene, with a transient cooling event constrained by Δ_{47} but not $\delta^{18}O$, and in the Miocene, wherein the majority of the Δ_{47} data are ~5°C warmer than the $\delta^{18}O$-derived record presented here. This latter discrepancy either implies that the clumped data record regionally warmer-than-global temperature at ODP Site 761 (where the majority of the Δ_{47} data come from in this interval, NW Australian margin, see the discussion in Evans (2021) and Modestou et al. (2020)), or that mid-Miocene deep ocean pH and/or sea level are substantially overestimated. Further work is of course required to understand whether the Δ_{47} data from this interval is a truly global signal and to determine the cause of the remaining discrepancies within the Eocene portion of the datasets .

We convert the $\delta^{18}O$-derived T_d record into GMST using a similar approach to H13, by splitting the Cenozoic into three intervals characterised by overall different relationships between T_d and GMST, namely, the Plio-Pleistocene intervals with T_d cooler than modern, the Plio-Pleistocene with T_d warmer than modern, and the remainder of the Cenozoic before, in this case, the PWP. Doing so requires three tie points at which GMST is well characterised and the assumption that the relationship between T_d and GMST remains constant within each of these portions of the dataset. As tie points, we use the 20th Century GMST (13.9°C, e.g. Trenberth and Fasullo (2013)), the LGM GMST analysis of Osman et al. (2021), and the mid-Piacenzian PRISM4 GMSST (Dowsett et al., 2016, 2013), coupled with the minimum T_d of the last 25 ka and mean T_d of the interval 3.00-3.05 Ma. Here, the switch between LGM and Pliocene scaling occurs at a $\Delta T_d = 0$ based on the 1 ka T_d datapoint. The accuracy of this analysis obviously depends on the quality of the GMS(S)T data, all of which are based on large, independent proxy data compilations or observations. In the case of the mid-Piacenzian, we unavoidably use a GMSST rather than a GMST estimate to anchor the Cenozoic GMST reconstruction, because there is, to our knowledge, currently no curated data compilation on which such an estimate could be based. However, we note that both the HadCM3L$_{V21}$ and DeepMIP suite of climate models (Fig. 4E,F) fall onto a single emergent GMST-GMSST relationship, which we use to convert Pliocene GMSST to GMST. Performing this calculation using HadCM3L$_{V21}$ yields a mid-Piacenzian GMST = 15.0°C, preferred here because these

simulations have variable palaeogeography. However, the exercise is overall insensitive to this choice; alternatively using the DeepMIP set of simulations would result in pre-Pliocene GMST 0.45°C cooler. Together, this exercise defines $\Delta GMST/\Delta T_d$ scaling factors of 2.2, 1.3, and 1.0 for the Plio-Pleistocene cooler and warmer than present, and pre-3.025 Ma, respectively.

The resulting GMST estimates, both with and without a pH correction on $\delta^{18}O$, are shown in Fig. 9B. This updated Cenozoic GMST reconstruction constrains the magnitude of cooling from the EECO (53-50 Ma) to the 21st Century to 17.3°C, of which 45% occurs during the Eocene, 20% across the Eocene-Oligocene Transition, 20% during the Miocene, and 15% during the Plio-Pleistocene. The pH correction on $\delta^{18}O_b$ exerts a strong control on Palaeogene-Miocene GMST reconstructed in this way, e.g. elevating EECO GMST by ~5°C. Understanding whether or not this correction should be applied is clearly an urgent priority, and if so, as is the production of a high-resolution deep ocean pH record. More broadly, placing an uncertainty estimate on T_d-derived GMST is challenging because the pre-Pliocene scaling factor is an assumption that has not been derived from any specific dataset. While we show that it appears to be a good assumption, the independent proxy data and associated transformations are currently insufficiently error-free to place further constraint on the scaling factor (Fig. 3) while the climate model simulations interrogated here suggest that it lies between 1.03 (DeepMIP) and 0.86 ($HadCM3L_{V21}$), see Fig. 4. Notwithstanding the potential pitfalls in using the model derived slopes to constrain uncertainty in the approach overall, we apply a ±2°C uncertainty to the T_d and GMST reconstructions shown in Fig. 9, which is the approximate difference that would result in the Eocene between a scaling factor of 0.86 and 1.03, but again highlight that further systematic bias is possible, particularly related to the pH correction outlined above.

The main differences between our GMST reconstruction (Fig. 9) and that of H13 derive from i) the revision of the LGM-modern ΔGMST from 4.5 to 7°C by Osman et al. (2021), resulting in a substantially greater glacial-interglacial GMST change in the late Pleistocene, and ii) from the pH correction on $\delta^{18}O$, resulting in substantially warmer Palaeogene GMST as a result (average EECO GMST of ~27°C in H13 compared to ~31.°C in this study (Fig. 9B)). The revised Cenozoic GMST reconstruction agrees well with several independent lines of evidence from both models and proxy data (Fig. 10). For example, the Pliocene Model Intercomparison Project range (ΔGMST = 1.8-5.2°C relative to the pre-industrial era; Haywood et al. (2020)) covers the deep ocean-derived maxima for this interval (2.5°C). Our results are also in reasonable agreement with the model-informed Oligocene GMST estimates of O'Brien et al. (2020), with the latter offset to values ~0-3°C higher (Fig. 10). In addition, we observe excellent agreement between the data compilation-derived GMSST estimates of Ring et al. (2022) coupled with our assessment of deep ocean temperature and the T_d-GMSST relationship of Gaskell et al. (2022), see Fig. 10. However, our GMST estimates are broadly substantially cooler for much of the Neogene, and warmer during the early Paleogene than those of Ring et al. (2022). We note that the GMST and GMSST estimates of that study are broadly similar, which is at odds with modern observations and climate model simulations which require substantially warmer GMSST than GMST in cooler climate states (Fig. 4E,F), with the two converging only under early Paleogene-like global warmth. This discrepancy potentially points to a

815 systematic bias in the terrestrial proxy records (also discussed in Ring et al. (2022)), which warrants
816 further investigation, rather than a failure of climate models to capture this feature of Earth's
817 climate.

818 In the early Paleogene, the PETM and pre-PETM GMST estimates of Tierney et al. (2022) of 34.1°C
819 (33.1-35.5) and 28.5°C (27.5-30.1°C) are within uncertainty of this study (35.2 and 28.1°C,
820 respectively; note that the resolution of the core-PETM $\delta^{18}O$ data in the stack utilised here is
821 insufficient to place a precise estimate on the PETM using this approach). Our analysis constrains
822 EECO GMST to 31.3±1.3°C, slightly higher than the upper range of the estimates provided in Inglis
823 et al. (2020) but within uncertainty of both the data analysis and 6× CO_2 CESM simulation of Zhu et
824 al. (2019). While there are important aspects of the deep ocean-derived estimates that require
825 further research (see above), if correct, this would also constrain 'bulk' equilibrium climate
826 sensitivity at the upper end of the range reported in Inglis et al. (2020), i.e. ~5°C, in agreement with
827 the LP-PETM-derived ECS estimate of Tierney et al. (2022). The offset of this study as well as that of
828 Tierney et al. (2022) and Zhu et al. (2019) compared to Inglis et al. (2020) is likely driven in large
829 part by the inclusion of a substantial amount of terrestrial temperature data which may be cool
830 biased in several of the approaches included in that latter study (also compare our early Eocene
831 GMST estimate to that of Ring et al. (2022)). Support for this is provided by the fact that our T_d-
832 derived GMST is in excellent agreement with the DeepMIP database-derived estimate of coeval
833 GMSST of this study and the T_d-GMSST analysis of Gaskell et al. (2022), see Fig. 10, potentially
834 implying an issue with some of the terrestrial but not the ocean datasets in that database. This is
835 encouraging because both all model simulations considered here and our T_d-GMST analysis
836 coupled with the T_d-GMSST analysis of Gaskell et al. (2022) suggest a convergence, within ~1-3°C,
837 of GMST and GMSST under early Palaeogene-like climate states (Fig. 4E,F; 10). As such, cooler
838 EECO GMST estimates (Fig. 10) are difficult to reconcile with our data-derived GMSST, and/or
839 require this aspect of the climate model output to be inaccurate. We alternatively argue that the
840 consistency between this study, Gaskell et al. (2022), and this emergent property of climate models
841 provides strong support for the very warm EECO GMST presented here.

842

843 ## 5 Conclusions

844 Here, we interrogate the use of a deep ocean temperature (T_d) record to infer global mean surface
845 temperature (GMST) in detail, using both curated data compilations and two sets of climate model
846 simulations. In particular, we address the question of whether T_d and GMST are linearly related with
847 a slope of 1, as previously hypothesised (Hansen et al., 2013). We show that no such relationship
848 would be expected in a world characterised by polar amplification, because i) the high latitude
849 regions warm faster than the global mean, and ii) deep water is thought to have formed at high
850 latitudes throughout most, if not all, of the Cenozoic. However, proxy data compilations of the
851 Pliocene and early Palaeogene fall within uncertainty of a 1:1 T_d-GMST relationship, suggesting that
852 (some) other process(es) act to balance polar amplification. Using both the DeepMIP set of
853 simulations (Lunt et al., 2021) with varying CO_2 and fixed palaeogeography and a set of Cenozoic

HadCM3L simulations with covarying palaeogeography and CO_2 (Valdes et al., 2021), we show that these processes are: 1) an increasing seasonal bias in deep water formation as GMST increases, and 2) the fact that the land surface warms at a faster rate than the ocean surface (see also Goudsmit-Harzevoort et al. (2023)). While this provides a mechanistic basis for a 1:1 T_d-GMST relationship prior to the (mid)Pliocene, we note that some HadCM3 simulations do not adhere to this, with GMST underestimated by up to 3°C during the Miocene (Fig. 4C). This occurs when a relatively small CO_2 change is sufficient to shift the principal locus of deep-water formation without a large associated change in GMST, resulting, in these simulations, in a cooler deep ocean as North Atlantic overturning ceases to be an important source of deep water. Although there is, to our knowledge, no direct evidence for this scenario in the Cenozoic, it highlights that there is at least the potential for substantial deviations from a 1:1 T_d-GMST relationship, particularly in deeper time. More broadly, we stress that our key finding – that T_d-GMST is characterised by a 1:1 relationship prior to the Pliocene within the certainty of the proxy data records – is robust irrespective of the mechanism, and further work will of course be required to empirically determine whether the above causal processes inferred from climate model simulations were indeed responsible.

Our contribution substantially strengthens the notion that GMST may be simply calculated from that of the deep ocean with a useful degree of precision. However, we highlight that recent advances in proxy methodologies for deep ocean temperature have arguably increased the uncertainty in our knowledge of T_d itself, particularly in the early Cenozoic. Specifically, clumped isotope-derived palaeotemperatures are substantially warmer and more variable than our canonical understanding of benthic foraminiferal $\delta^{18}O$ and/or the Cenozoic evolution of $\delta^{18}O_{sw}$. Solving this issue is clearly an urgent priority, although we show that the majority of the discrepancy can be explained by a seawater carbonate chemistry (pH) effect on $\delta^{18}O$ (Fig. 9A), as also suggested by Meckler et al. (2022).

Using the pH-corrected $\delta^{18}O$ data, we construct a revised Cenozoic GMST record, broadly following the methodology of Hansen et al. (2013), but incorporating advances in our understanding of LGM and Pliocene GMST as well as the evolution of sea level throughout the past 40 Ma (Dowsett et al., 2016; Osman et al., 2021; Rohling et al., 2022). Our GMST record is warmer throughout much of the Cenozoic, with large (up to ~5°C) differences present prior to the Miocene. Overall, these estimates are in excellent agreement with several independent early Cenozoic GMST reconstructions, adding confidence to all of these various lines of evidence. Finally, we constrain the magnitude of the Cenozoic GMST decrease to 17.3°C (EECO to 20th Century), and EECO GMST to 31.3±1.3°C, slightly above the upper end of previous reconstructions. If correct, this would support the notion that 'bulk' equilibrium climate sensitivity of was higher than modern in this past warm climate state (Tierney et al., 2022).

Acknowledgements
DE and JB acknowledge support via the VeWA consortium funded by the Hessen State Ministry for Higher Education, Research, and the Arts through the LOEWE program. DE also acknowledges support from the Royal Society (award reference URF\R1\221735). 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). GNI also acknowledges support from NERC (NE/V018388/1). We thank Felix Strnad for adjusting the geoutils python package for the purposes of analysing the climate model data considered here.

Data availability statement

No new datasets are presented here. The climate model output interrogated here is available from https://www.deepmip.org/data/ and https://www.paleo.bristol.ac.uk/ummodel/scripts/papers/Valdes_et_al_2021.html (Lunt et al., 2021; Valdes et al., 2021). They key datasets used in the data analysis performed here are the WOA2018 ocean temperature dataset (https://www.ncei.noaa.gov/products/world-ocean-atlas (Locarnini et al., 2018)), the DeepMIP proxy database (Hollis et al., 2019), Cenozoic benthic foraminifera $\delta^{18}O$, Mg/Ca, Δ_{47}, $\delta^{11}B$, and sea level/temperature deconvolution (Cramer et al., 2011; Leutert et al., 2021; Meckler et al., 2022; Modestou et al., 2020; Rohling et al., 2022; Westerhold et al., 2020), and mid-Piacenzian SST synthesis (Dowsett et al., 2016). The python and Matlab scripts used to perform the data analysis presented here, and to produce all figures, can be found at https://zenodo.org/xxxxxxxxx [provided as a supplementary zip during review].

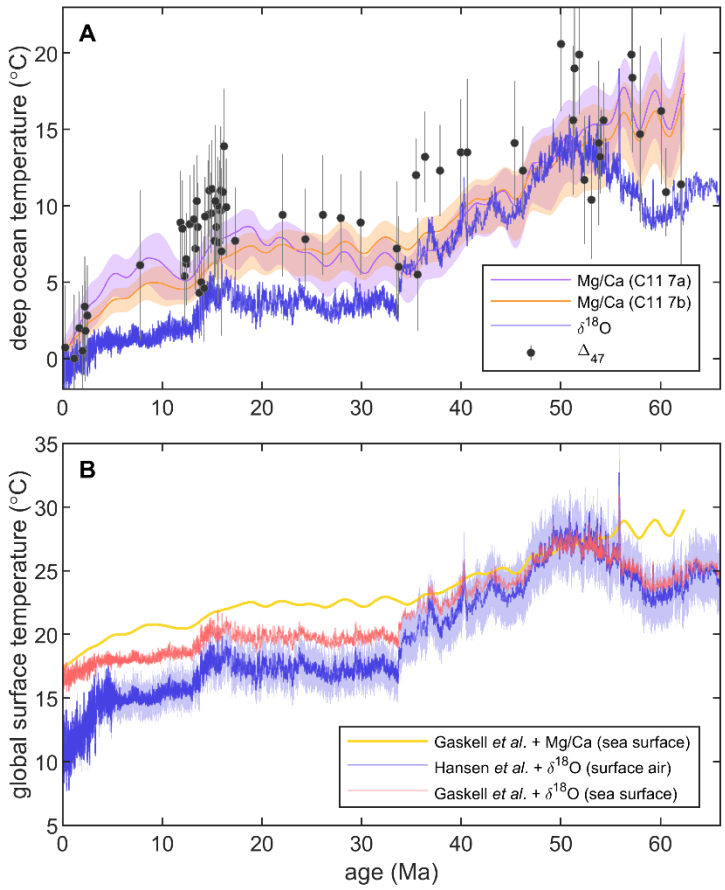

Fig. 1. (A) Deep ocean temperature based on the oxygen isotope composition of foraminifera (Westerhold et al., 2020) converted to temperature following the approach of Hansen et al. (2013), the Mg to Ca ratio of foraminifera (Cramer et al., 2011), and the clumped isotopic composition of foraminifera (Δ_{47}; Leutert et al., 2021; Meckler et al., 2022; Modestou et al., 2020). In the case of Mg/Ca, the two transformation equations refer to the two Mg/Ca-temperature calibrations explored by (Cramer et al., 2011). (B) Global mean surface air temperature (Hansen et al., 2013) and global mean sea surface temperature (Gaskell et al., 2022) calculated according to the methodologies outlined in the original studies except using the revised deep ocean benthic foraminiferal oxygen isotope stack (Westerhold et al., 2020). An arbitrary ±10% uncertainty has been added to the global mean surface air temperature estimate. Note that the deep ocean to surface temperature transformation of (Hansen et al., 2013) is parameterised according to some specific features of the benthic oxygen isotope stack, such that calculating GMST from benthic foraminifera Mg/Ca in a directly comparable way is not possible and was not attempted.

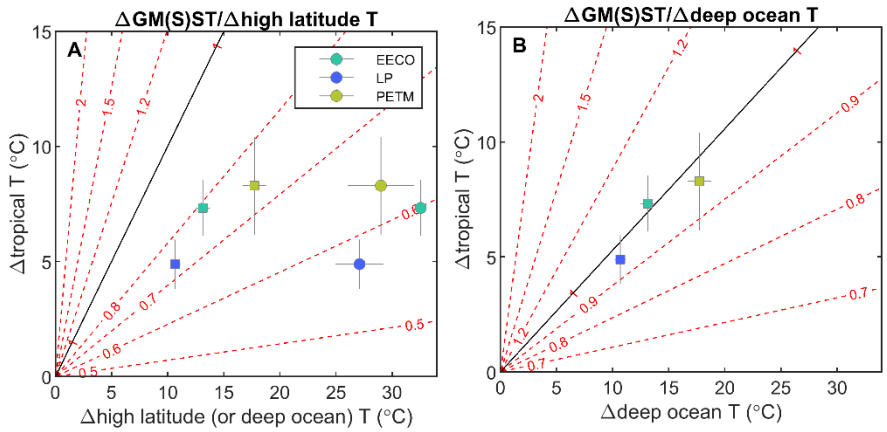

Fig. 2. (A) The idealised relationship between GM(S)ST and high latitude/deep ocean temperature for a given degree of high latitude/deep ocean and tropical warming (contour lines, see text and supplementary Text S1 for details) for a zonally homogeneous waterworld. Assuming deep ocean and high latitude SST are exactly coupled and in the absence of other processes, only in a world characterised by exactly zero polar amplification is a 1:1 relationship between GM(S)ST and high latitude SST expected (black line). Estimates of high latitude and tropical SST change for three DeepMIP target intervals (Inglis et al., 2020) are shown (circles). Given a likely seasonal bias in high latitude proxy SST, high latitude temperature is alternatively calculated for the same intervals using the benthic foraminiferal oxygen isotope stack (squares; see Fig. 1 and the text for methodological details). (B) The same analysis assuming that deep water formation becomes increasingly biased towards winter as high latitude SST increases (specifically, for every 5°C high latitude SST increase, deep ocean temperature is biased by 1°C below the mean annual high latitude average). This has the effect of stretching the contour lines shown in panel A towards higher deep ocean temperatures. Note that the prescribed change in the seasonal bias in deep water formation is entirely without basis; the exercise is intended as a simple illustration of a process that could result in a 1:1 relationship between deep ocean temperature and GM(S)ST.

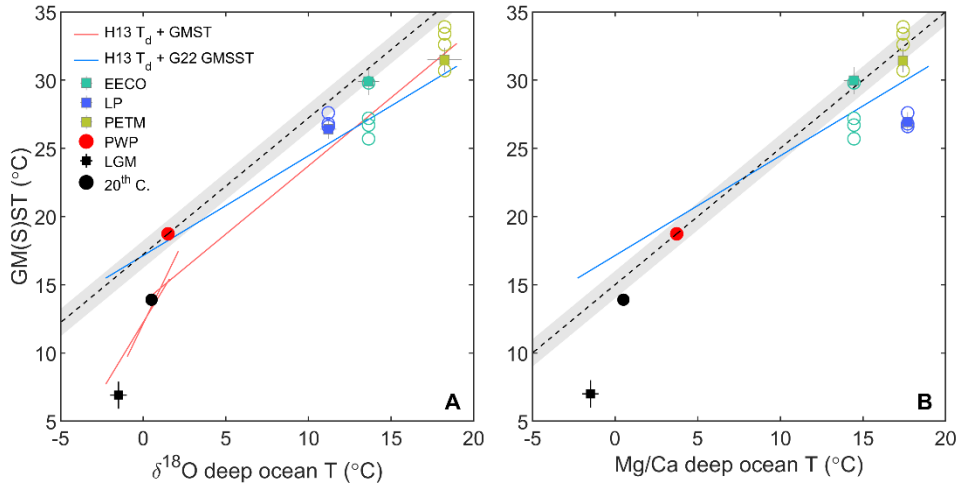

Fig. 3. Empirical estimates of GMS(S)T as a function of deep ocean temperature for key Cenozoic intervals for which curated data compilation efforts exist (Dowsett et al., 2016; Hollis et al., 2019; Osman et al., 2021). Estimates of GMST for three DeepMIP target intervals (Inglis et al., 2020) are shown (open circles, see that study for uncertainties) as well as GMSST (filled squares; this study) (A) Deep ocean temperature calculated from the benthic foraminifera oxygen isotope stack following Hansen et al. (2013). (B) Deep ocean temperature calculated from benthic foraminiferal Mg/Ca following Cramer et al. (2011). LGM and pre-industrial deep ocean temperature is not based on foraminiferal δ^{18}O/Mg/Ca, see text for details. The grey shaded region depicts a 1:1 increase in deep ocean and GMST anchored to the Pliocene Warm Period (red circles), with an arbitrary ±1°C uncertainty. Previously suggested relationships between deep ocean temperature and GMST (Hansen et al., 2013) and GMSST (Gaskell et al., 2022) are shown with red and blue lines respectively. Note that: 1) the Paleogene GMST estimates of this study are not fully independent from estimated deep ocean temperature, because deep ocean temperature was used to estimate high latitude SST in order to avoid a seasonal bias in the surface proxy data, and 2) the PWP data point is a GMSST estimate that may overestimate GMST. This may explain the discrepancy between the Palaeogene estimates from the 1:1 line anchored to the PWP, see text for details.

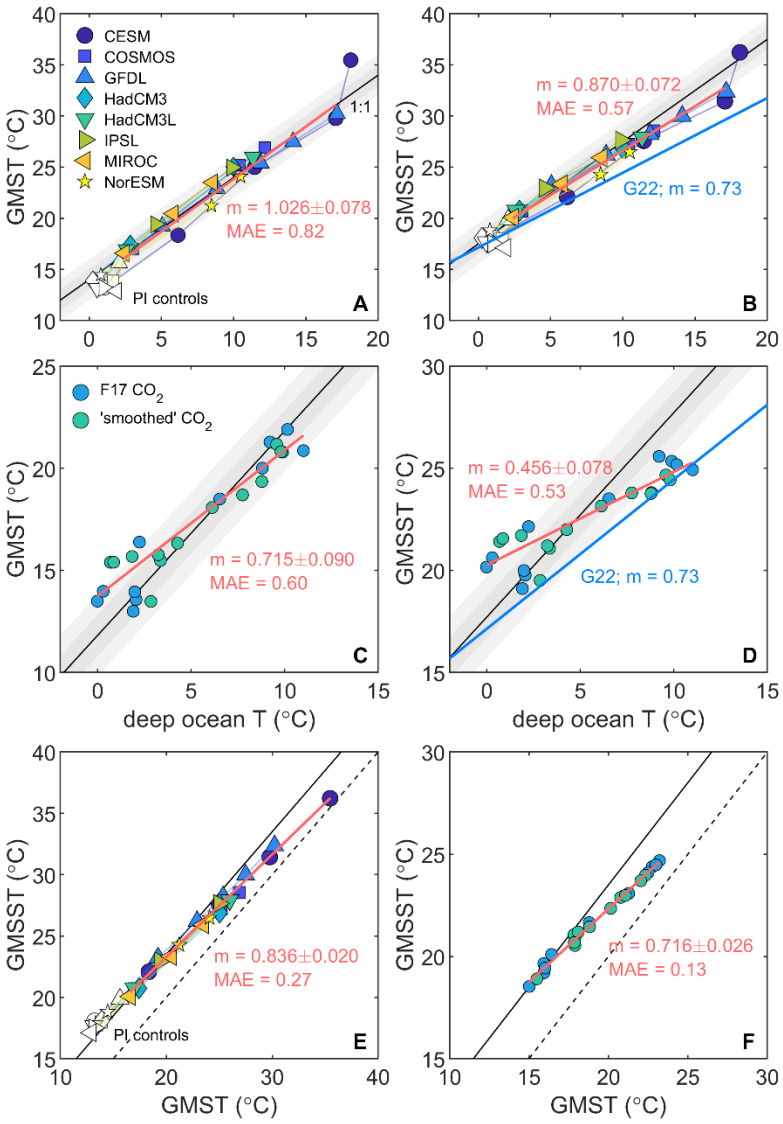

Fig. 4. (A) Global mean surface temperature and (B) Global mean sea surface temperature as a function of deep ocean temperature (>3000 m) in the DeepMIP set of model simulations, see panels C and D for x axis labels. One-to-one lines are anchored to the mean of the simulations conducted at 1×CO_2 and Eocene palaeogeography plus the IPSL simulation at 1.5×CO_2; shaded regions depict ±1 & 2°C from this line. The least squares linear regressions (red lines) include all model simulations with Eocene palaeogeography, except for the 9×CO_2 CESM simulation (see text). (C,D) A similar analysis performed for the Cenozoic HadCM3 simulations of Valdes et al. (2021) with 1:1 lines anchored to the mean of the two 3 Ma simulations. A foraminiferal $\delta^{18}O$-derived relationship between GMSST and deep ocean temperature is shown in panels B and D, calculated following Gaskell et al. (2022). (E,F) The relationship between global mean surface temperature and global mean sea surface temperature in the DeepMIP and HadCM3 simulations of Valdes et al. (2021), respectively. Lines with a slope of 1 are shown anchored as described for the

other panels (solid) as well as 1:1 lines (dashed). All model data were interpolated to a 1×1° grid before further calculations were performed.

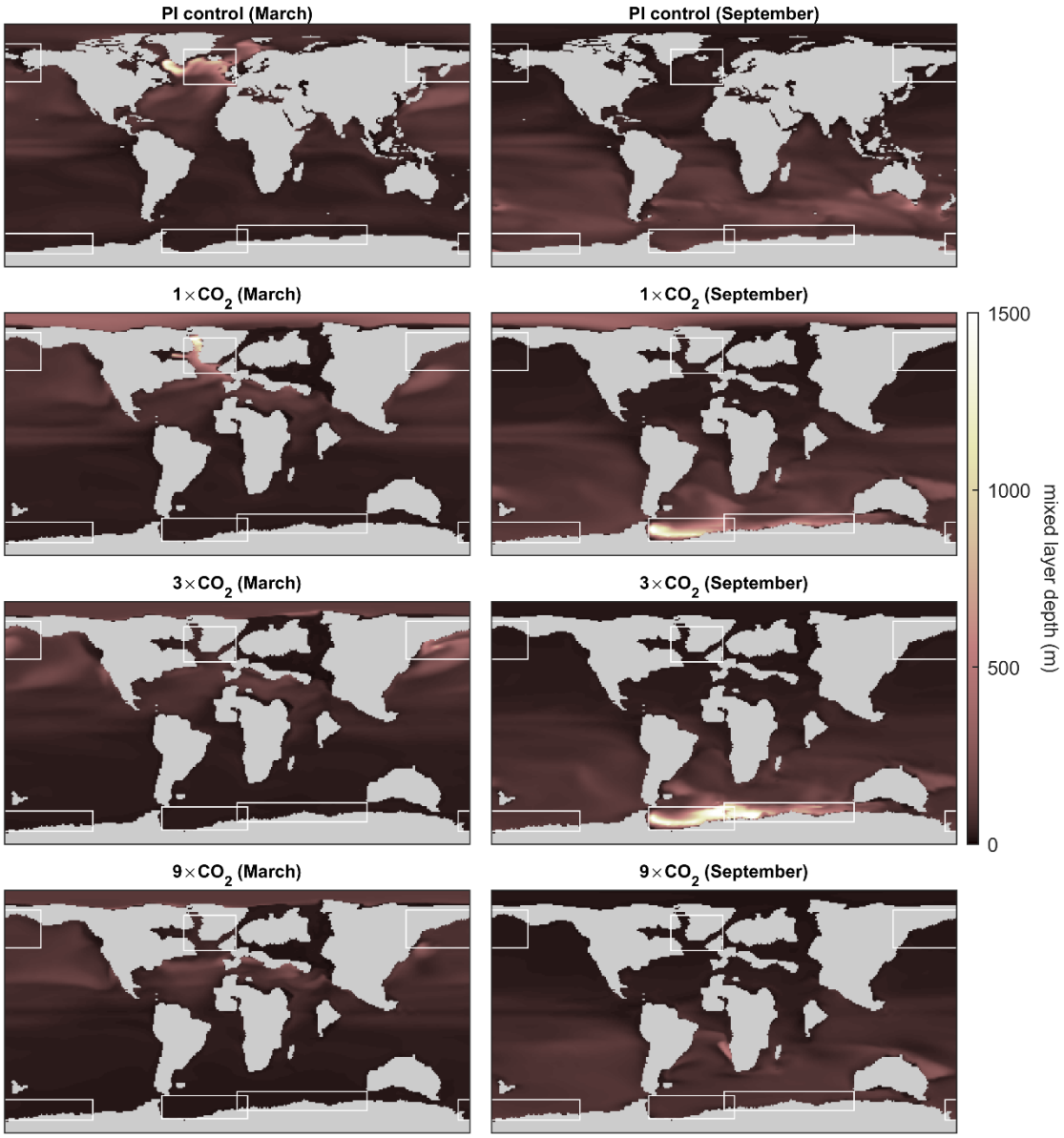

Fig. 5. March and September mixed layer depth in the four DeepMIP CESM simulations (PI control, as well as 1,3, and 9× CO_2). Boxes showing the key regions of deep water formation in the full suite of models contributing to the DeepMIP set of simulations are overlain (Zhang et al., 2022).

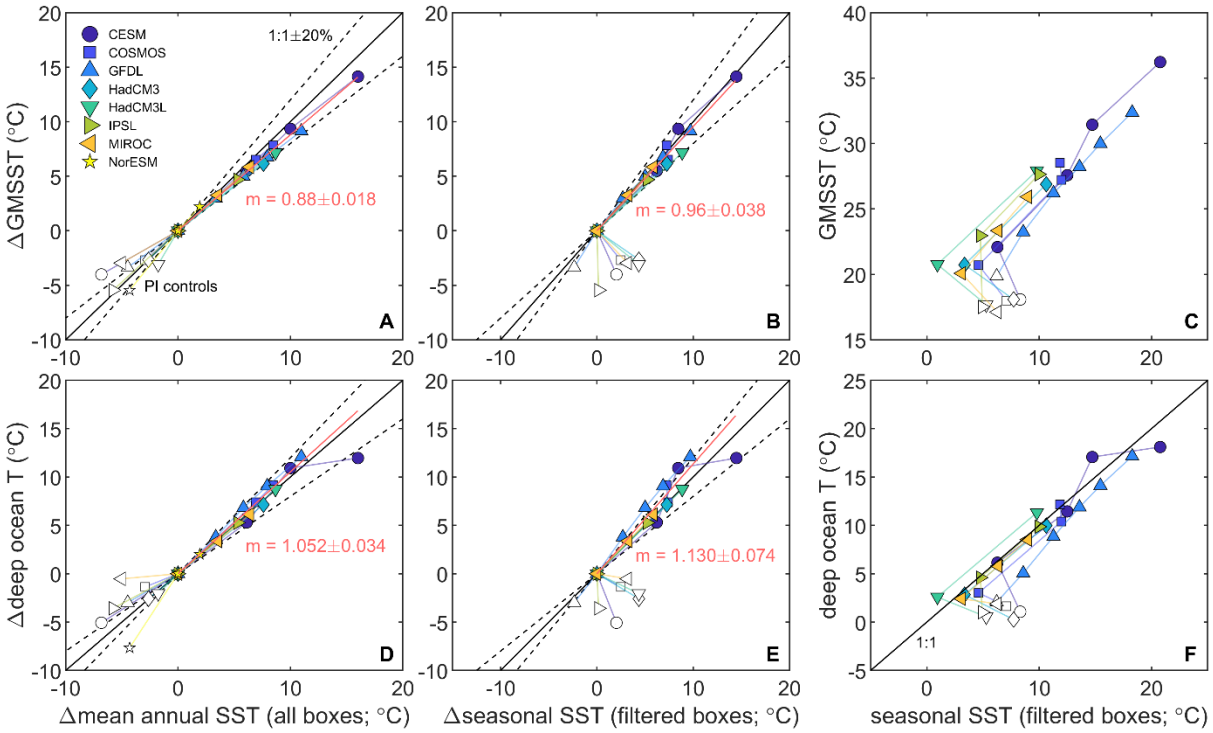

Fig. 6. The relationship between sea surface temperature (0-100 m) in the broad regions of deep-water formation and global mean sea surface temperature (GMSST) or deep ocean temperature in the DeepMIP set of model simulations. (A,D) Mean annual SST in all boxes shown in Fig. 5 plotted relative to the 1×CO_2 simulation with Eocene palaeogeography (except IPSL; 1.5×CO_2). (B,E) Winter SST in the model-specific box(es) relevant for deep water formation. Note that seasonal SST data for NorESM was not available. (C,F) As panel B/E, except in absolute temperature space. In all cases, the least squares linear regressions are forced through the origin and fit to the ensemble, excluding the pre-industrial controls. The 9×CO_2 CESM simulation was excluded from the fit.

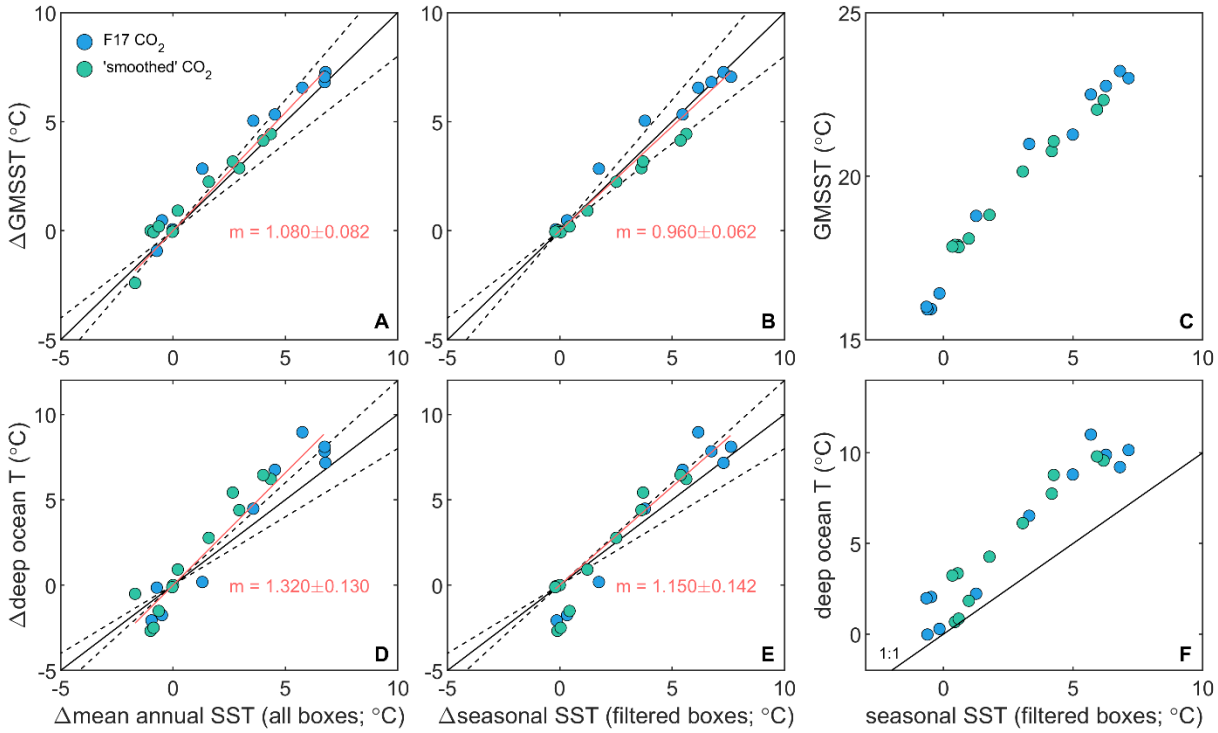

Fig. 7. The relationship between sea surface temperature (0-100 m) in the broad regions of deep-water formation and global mean sea surface temperature (GMSST) or deep ocean temperature in the HadCM3L simulations of Valdes et al. (2021). Two simulations were performed for each time slice, at two different CO_2. (A,D) Mean annual SST in all boxes shown in Fig. 5 plotted relative to the 3 Ma simulation. (B,E) Winter SST in the model-specific box(es) relevant for deep water formation. Data from the 0 Ma simulation are not shown as these fall off of the trend as a result of being characterised by N. Atlantic deep water formation. (C,F) As panel B/E, except in absolute temperature space. In all cases, the least squares linear regressions are forced through the origin and fit to all simulations.

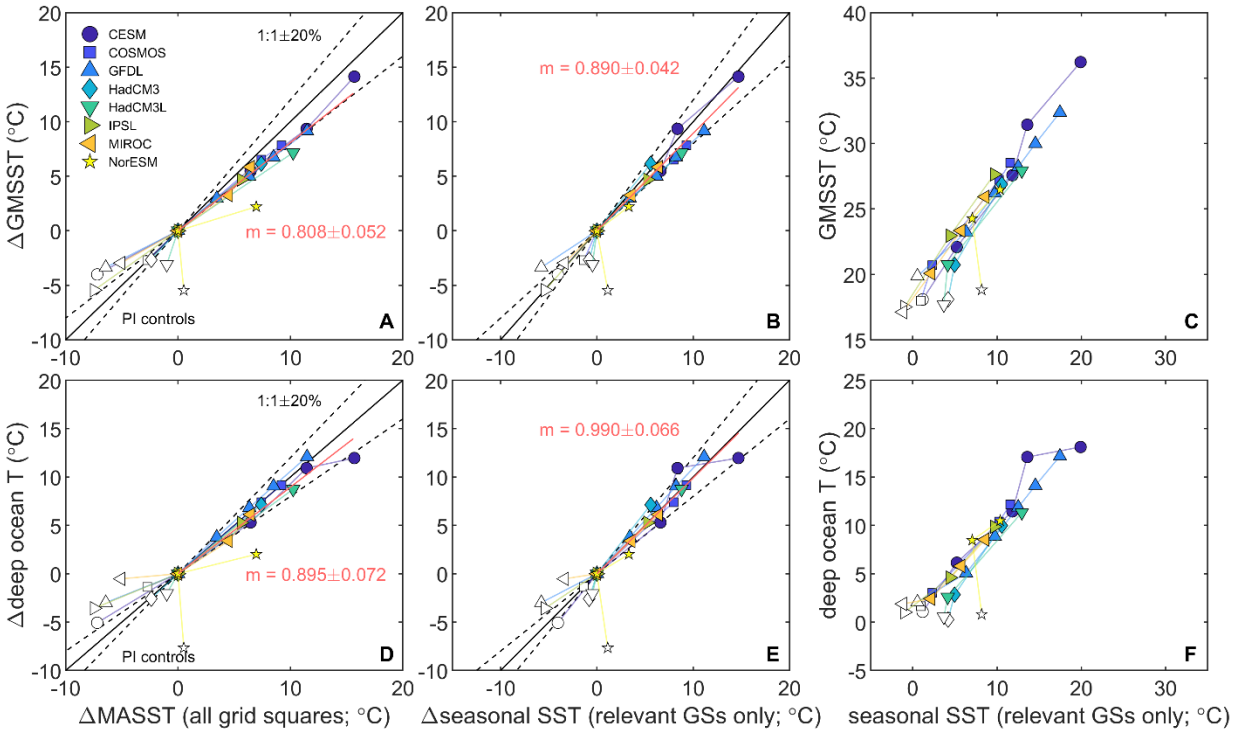

Fig. 8. The relationship between SST in the model grid cells with a mixed layer depth (mld) at least 90% of the global seasonal maximum and global mean sea surface temperature (GMSST) or deep ocean temperature in the DeepMIP set of model simulations. (A,D) Mean annual SST in all grid cells meeting the mld criteria plotted relative to the 1×CO_2 simulation with Eocene palaeogeography (except IPSL; 1.5×CO_2). (B,E) SST in all grid cells (GSs) meeting the mld criteria during the season of maximum mixed layer depth, and limited to the model-specific hemisphere(s) relevant for deep water formation. Note that seasonal SST data for NorESM was not available. (C,F) As panel B/E, except in absolute temperature space. In all cases, the least squares linear regressions are fit to the ensemble, excluding the pre-industrial controls. The 6× and 9×CO_2 CESM simulations were excluded from the fit.

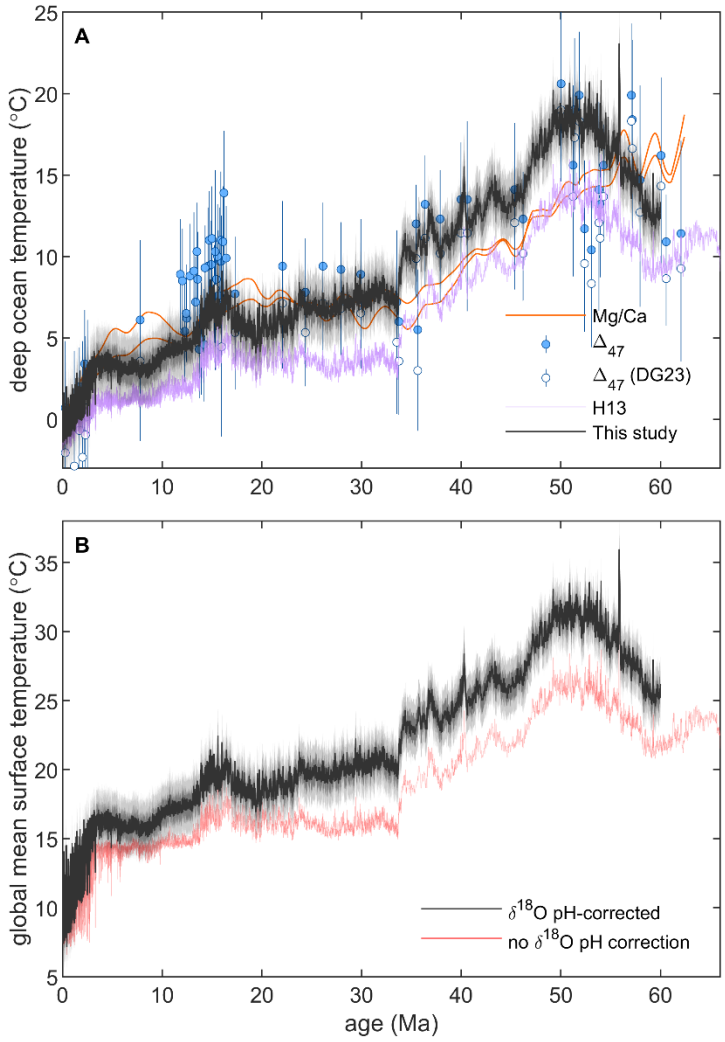

Figure 9. (A) A revised estimate of the Cenozoic evolution of deep ocean temperature based on the sea-level and pH-corrected benthic foraminifera oxygen isotope stack (black line with an arbitrary ±2°C uncertainty, see text for details) in the context of other proxy estimates (see Fig. 1), including the Δ_{47} reanalysis of Daëron and Gray (2023). (B) GMST based on the deep ocean temperature record from this study (see text) and the sensitivity of this reconstruction to whether or not the benthic foraminiferal $\delta^{18}O$ data are pH corrected.

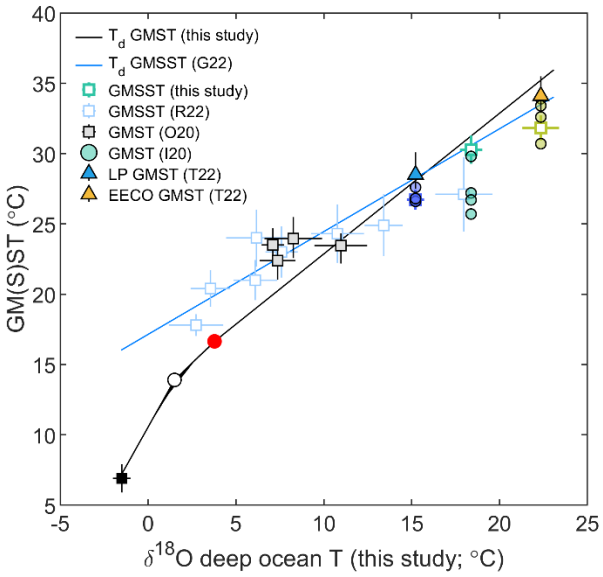

Figure 10. The revised relationship between deep ocean temperature (T_d) and GMST (black line, this study), showing the three anchor points used here (LGM, 20th Century, and PWP, black square, white circle, and red circle respectively; see text for details). Following our analysis and H13, the relationship between T_d and GMST has a slope of 1 for all climate states warmer than the PWP. Independent estimates of GMST (Inglis et al., 2020; O'Brien et al., 2020; Tierney et al., 2022 (I20, O20, T22, respectively) and GMSST (open squares; this study, based on the DeepMIP database (Hollis et al. 2019) and the revised assessment of T_d (Fig. 9), as well as those of Ring et al. (2022); R22) are shown. Note that the blue line is the T_d-GMSST relationship of Gaskell et al. (2022) and not the best fit regression of the estimates from Ring et al. (2022).

References

Adkins, J.F., McIntyre, K., Schrag, D.P., 2002. The Salinity, Temperature, and δ18O of the Glacial Deep Ocean. Science 298, 1769–1773. https://doi.org/10.1126/science.1076252

Affek, H.P., 2012. Clumped Isotope Paleothermometry: Principles, Applications, and Challenges. The Paleontological Society Papers 18, 101–114. https://doi.org/10.1017/S1089332600002576

Anagnostou, E., John, E.H., Babila, T.L., Sexton, P.F., Ridgwell, A., Lunt, D.J., Pearson, P.N., Chalk, T.B., Pancost, R.D., Foster, G.L., 2020. Proxy evidence for state-dependence of climate sensitivity in the Eocene greenhouse. Nat Commun 11, 4436. https://doi.org/10.1038/s41467-020-17887-x

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

Bijma, J., Spero, H.J., Lea, D.W., 1999. Reassessing Foraminiferal Stable Isotope Geochemistry: Impact of the Oceanic Carbonate System (Experimental Results), in: Fischer, G., Wefer, G. (Eds.), Use of Proxies in Paleoceanography: Examples from the South Atlantic. Springer, Berlin, Heidelberg, pp. 489–512. https://doi.org/10.1007/978-3-642-58646-0_20

Billups, K., Schrag, D.P., 2003. Application of benthic foraminiferal Mg/Ca ratios to questions of Cenozoic climate change. Earth and Planetary Science Letters 209, 181–195. https://doi.org/10.1016/S0012-821X(03)00067-0

Brinkhuis, H., Schouten, S., Collinson, M.E., Sluijs, A., Damsté, J.S.S., Dickens, G.R., Huber, M., Cronin, T.M., Onodera, J., Takahashi, K., Bujak, J.P., Stein, R., van der Burgh, J., Eldrett, J.S., Harding, I.C., Lotter, A.F., Sangiorgi, F., Cittert, H. van K., de Leeuw, J.W., Matthiessen, J., Backman, J., Moran, K., 2006. Episodic fresh surface waters in the Eocene Arctic Ocean. Nature 441, 606–609. https://doi.org/10.1038/nature04692

Burls, N.J., Bradshaw, C.D., De Boer, A.M., Herold, N., Huber, M., Pound, M., Donnadieu, Y., Farnsworth, A., Frigola, A., Gasson, E., von der Heydt, A.S., Hutchinson, D.K., Knorr, G., Lawrence, K.T., Lear, C.H., Li, X., Lohmann, G., Lunt, D.J., Marzocchi, A., Prange, M., Riihimaki, C.A., Sarr, A.-C., Siler, N., Zhang, Z., 2021. Simulating Miocene Warmth: Insights From an Opportunistic Multi-Model Ensemble (MioMIP1). Paleoceanography and Paleoclimatology 36, e2020PA004054. https://doi.org/10.1029/2020PA004054

Caballero, R., Huber, M., 2013. 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. https://doi.org/10.1073/pnas.1303365110

Covey, C., Sloan, L.C., Hoffert, M.I., 1996. Paleoclimate data constraints on climate sensitivity: The paleocalibration method. Climatic Change 32, 165–184. https://doi.org/10.1007/BF00143708

Cramer, B.S., Miller, K.G., Barrett, P.J., Wright, J.D., 2011. Late Cretaceous–Neogene trends in deep ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal geochemistry (δ18O and Mg/Ca) with sea level history. Journal of Geophysical Research: Oceans 116. https://doi.org/10.1029/2011JC007255

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., Röhl, U., Schouten, S., Sluijs, A., 2018. Synchronous tropical and polar temperature evolution in the Eocene. Nature 559, 382–386. https://doi.org/10.1038/s41586-018-0272-2

Daëron, M., Gray, W.R., 2023. Revisiting Oxygen-18 and Clumped Isotopes in Planktic and Benthic Foraminifera. Paleoceanography and Paleoclimatology 38, e2023PA004660. https://doi.org/10.1029/2023PA004660

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A.M., Mitrovica, J.X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K., Haywood, A., 2016. The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. Climate of the Past 12, 1519–1538. https://doi.org/10.5194/cp-12-1519-2016

Dowsett, H.J., Foley, K.M., Stoll, D.K., Chandler, M.A., Sohl, L.E., Bentsen, M., Otto-Bliesner, B.L., Bragg, F.J., Chan, W.-L., Contoux, C., Dolan, A.M., Haywood, A.M., Jonas, J.A., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Nisancioglu, K.H., Abe-Ouchi, A., Ramstein, G., Riesselman, C.R., Robinson, M.M., Rosenbloom, N.A., Salzmann, U., Stepanek, C., Strother, S.L., Ueda, H., Yan, Q., Zhang, Z., 2013. Sea Surface Temperature of the mid-Piacenzian Ocean: A Data-Model Comparison. Sci Rep 3, 2013. https://doi.org/10.1038/srep02013

Dunkley Jones, T., Lunt, D.J., Schmidt, D.N., Ridgwell, A., Sluijs, A., Valdes, P.J., Maslin, M., 2013. Climate model and proxy data constraints on ocean warming across the Paleocene–Eocene Thermal Maximum. Earth-Science Reviews 125, 123–145. https://doi.org/10.1016/j.earscirev.2013.07.004

Evans, D., 2021. Deep Heat: Proxies, Miocene Ice, and an End in Sight for Paleoclimate Paradoxes? Paleoceanography and Paleoclimatology 36, e2020PA004174. https://doi.org/10.1029/2020PA004174

Evans, D., Müller, W., 2012. Deep time foraminifera Mg/Ca paleothermometry: Nonlinear correction for secular change in seawater Mg/Ca. Paleoceanography 27. https://doi.org/10.1029/2012PA002315

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., Affek, H.P., 2018. Eocene greenhouse climate revealed by coupled clumped isotope-Mg/Ca thermometry. PNAS 115, 1174–1179. https://doi.org/10.1073/pnas.1714744115

Evans, D., Wade, B.S., Henehan, M., Erez, J., Müller, W., 2016. Revisiting carbonate chemistry controls on planktic foraminifera Mg / Ca: implications for sea surface temperature and hydrology shifts over the Paleocene–Eocene Thermal Maximum and Eocene–Oligocene transition. Climate of the Past 12, 819–835. https://doi.org/10.5194/cp-12-819-2016

Farnsworth, A., Lunt, D.J., O'Brien, C.L., Foster, G.L., Inglis, G.N., Markwick, P., Pancost, R.D., Robinson, S.A., 2019. Climate Sensitivity on Geological Timescales Controlled by Nonlinear Feedbacks and Ocean Circulation. Geophysical Research Letters 46, 9880–9889. https://doi.org/10.1029/2019GL083574

Ferreira, D., Cessi, P., Coxall, H.K., de Boer, A., Dijkstra, H.A., Drijfhout, S.S., Eldevik, T., Harnik, N., McManus, J.F., Marshall, D.P., Nilsson, J., Roquet, F., Schneider, T., Wills, R.C., 2018. Atlantic-Pacific Asymmetry in Deep Water Formation. Annual Review of Earth and Planetary Sciences 46, 327–352. https://doi.org/10.1146/annurev-earth-082517-010045

Ford, H.L., Burls, N.J., Jacobs, P., Jahn, A., Caballero-Gill, R.P., Hodell, D.A., Fedorov, A.V., 2022. Sustained mid-Pliocene warmth led to deep water formation in the North Pacific. Nat. Geosci. 15, 658–663. https://doi.org/10.1038/s41561-022-00978-3

Foster, G.L., Rae, J.W.B., 2016. Reconstructing Ocean pH with Boron Isotopes in Foraminifera. Annual Review of Earth and Planetary Sciences 44, 207–237. https://doi.org/10.1146/annurev-earth-060115-012226

Foster, G.L., Rohling, E.J., 2013. Relationship between sea level and climate forcing by CO2 on geological timescales. PNAS 110, 1209–1214. https://doi.org/10.1073/pnas.1216073110

Gaskell, D., Huber, M., O'Brien, C.L., Inglis, G.N., Acosta, P., Poulsen, C.J., Hull, P.M., 2022. The latitudinal temperature gradient and its climate dependence as inferred from foraminiferal $\delta^{18}O$ over the past 95 million years. PNAS 119, e2111332119.

Gothmann, A.M., Stolarski, J., Adkins, J.F., Schoene, B., Dennis, K.J., Schrag, D.P., Mazur, M., Bender, M.L., 2015. Fossil corals as an archive of secular variations in seawater chemistry since the Mesozoic. Geochimica et Cosmochimica Acta 160, 188–208. https://doi.org/10.1016/j.gca.2015.03.018

Goudsmit-Harzevoort, B., Lansu, A., Baatsen, M.L.J., von der Heydt, A.S., de Winter, N.J., Zhang, Y., Abe-Ouchi, A., de Boer, A., Chan, W.-L., Donnadieu, Y., Hutchinson, D.K., Knorr, G., Ladant, J.-B., Morozova, P., Niezgodzki, I., Steinig, S., Tripati, A., Zhang, Z., Zhu, J., Ziegler, M., 2023. The Relationship Between the Global Mean Deep-Sea and Surface Temperature During the Early Eocene.

Paleoceanography and Paleoclimatology 38, e2022PA004532. https://doi.org/10.1029/2022PA004532
Green, J. a. M., Huber, M., 2013. Tidal dissipation in the early Eocene and implications for ocean mixing. Geophysical Research Letters 40, 2707–2713. https://doi.org/10.1002/grl.50510
Greenop, R., Foster, G.L., Wilson, P.A., Lear, C.H., 2014. Middle Miocene climate instability associated with high-amplitude CO2 variability. Paleoceanography 29, 845–853. https://doi.org/10.1002/2014PA002653
Gulev, S.K., Thorne, P.W., Ahn, J., Dentener, F.J., Domingues, C.M., Gerland, S., Gong, D., Kaufman, D.S., Nnamchi, H.C., Quaas, J., Rivera, J.A., Sathyendranath, S., Smith, S.L., Trewin, B., von Shuckmann, K., Vose, R.S., 2021. Changing state of the climate system, in: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, Ö., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
Hansen, J., Ruedy, R., Sato, M., Lo, K., 2010. Global Surface Temperature Change. Reviews of Geophysics 48. https://doi.org/10.1029/2010RG000345
Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., Pagani, M., Raymo, M., Royer, D.L., Zachos, J.C., 2008. Target Atmospheric CO: Where Should Humanity Aim? The Open Atmospheric Science Journal 2. https://doi.org/10.2174/1874282300802010217
Hansen, J., Sato, M., Russell, G., Kharecha, P., 2013. Climate sensitivity, sea level and atmospheric carbon dioxide. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371, 20120294. https://doi.org/10.1098/rsta.2012.0294
Haywood, A.M., Dolan, A.M., Pickering, S.J., Dowsett, H.J., McClymont, E.L., Prescott, C.L., Salzmann, U., Hill, D.J., Hunter, S.J., Lunt, D.J., Pope, J.O., Valdes, P.J., 2013. On the identification of a Pliocene time slice for data–model comparison. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371, 20120515. https://doi.org/10.1098/rsta.2012.0515
Haywood, A.M., Tindall, J.C., Dowsett, H.J., Dolan, A.M., Foley, K.M., Hunter, S.J., Hill, D.J., Chan, W.-L., Abe-Ouchi, A., Stepanek, C., Lohmann, G., Chandan, D., Peltier, W.R., Tan, N., Contoux, C., Ramstein, G., Li, X., Zhang, Z., Guo, C., Nisancioglu, K.H., Zhang, Q., Li, Q., Kamae, Y., Chandler, M.A., Sohl, L.E., Otto-Bliesner, B.L., Feng, R., Brady, E.C., von der Heydt, A.S., Baatsen, M.L.J., Lunt, D.J., 2020. The Pliocene Model Intercomparison Project Phase 2: large-scale climate features and climate sensitivity. Climate of the Past 16, 2095–2123. https://doi.org/10.5194/cp-16-2095-2020
Henry, M., Vallis, G.K., 2022. Variations on a Pathway to an Early Eocene Climate. Paleoceanography and Paleoclimatology 37, e2021PA004375. https://doi.org/10.1029/2021PA004375
Hoffert, M.I., Covey, C., 1992. Deriving global climate sensitivity from palaeoclimate reconstructions. Nature 360, 573–576. https://doi.org/10.1038/360573a0
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., Lunt, D.J., 2019. 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. https://doi.org/10.5194/gmd-12-3149-2019
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., Zachos, J.C., 2012. Early Paleogene

temperature history of the Southwest Pacific Ocean: Reconciling proxies and models. Earth and Planetary Science Letters 349–350, 53–66. https://doi.org/10.1016/j.epsl.2012.06.024

Hönisch, B., Ridgwell, A., Schmidt, D.N., Thomas, E., Gibbs, S.J., Sluijs, A., Zeebe, R., Kump, L., Martindale, R.C., Greene, S.E., Kiessling, W., Ries, J., Zachos, J.C., Royer, D.L., Barker, S., Marchitto, T.M., Moyer, R., Pelejero, C., Ziveri, P., Foster, G.L., Williams, B., 2012. The Geological Record of Ocean Acidification. Science 335, 1058–1063. https://doi.org/10.1126/science.1208277

Inglis, G.N., Bragg, F., Burls, N.J., Cramwinckel, M.J., Evans, D., Foster, G.L., Huber, M., Lunt, D.J., Siler, N., Steinig, S., Tierney, J.E., Wilkinson, R., Anagnostou, E., de Boer, A.M., Dunkley Jones, T., Edgar, K.M., Hollis, C.J., Hutchinson, D.K., Pancost, R.D., 2020. 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. https://doi.org/10.5194/cp-16-1953-2020

Kelemen, F.D., Steinig, S., de Boer, A., Zhu, J., Chan, W.-L., Niezgodzki, I., Hutchinson, D.K., Knorr, G., Abe-Ouchi, A., Ahrens, B., 2023. Meridional Heat Transport in the DeepMIP Eocene Ensemble: Non-CO2 and CO2 Effects. Paleoceanography and Paleoclimatology 38, e2022PA004607. https://doi.org/10.1029/2022PA004607

Lear, C.H., Coxall, H.K., Foster, G.L., Lunt, D.J., Mawbey, E.M., Rosenthal, Y., Sosdian, S.M., Thomas, E., Wilson, P.A., 2015. Neogene ice volume and ocean temperatures: Insights from infaunal foraminiferal Mg/Ca paleothermometry. Paleoceanography 30, 1437–1454. https://doi.org/10.1002/2015PA002833

Lear, C.H., Elderfield, H., Wilson, P.A., 2000. Cenozoic Deep-Sea Temperatures and Global Ice Volumes from Mg/Ca in Benthic Foraminiferal Calcite. Science 287, 269–272. https://doi.org/10.1126/science.287.5451.269

Leutert, T.J., Modestou, S., Bernasconi, S.M., Meckler, A.N., 2021. Southern Ocean bottom-water cooling and ice sheet expansion during the middle Miocene climate transition. Climate of the Past 17, 2255–2271. https://doi.org/10.5194/cp-17-2255-2021

Leutert, T.J., Sexton, P.F., Tripati, A., Piasecki, A., Ho, S.L., Meckler, A.N., 2019. Sensitivity of clumped isotope temperatures in fossil benthic and planktic foraminifera to diagenetic alteration. Geochimica et Cosmochimica Acta 257, 354–372. https://doi.org/10.1016/j.gca.2019.05.005

Lisiecki, L.E., Raymo, M.E., 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography 20. https://doi.org/10.1029/2004PA001071

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., Reagan, J., Seidov, D., Weathers, K., Paver, C., Smolyar, I., 2018. World Ocean Atlas 2018, Volume 1: Temperature.

Lunt, D.J., Bragg, F., Chan, W.-L., Hutchinson, D.K., Ladant, J.-B., Morozova, P., Niezgodzki, I., Steinig, S., Zhang, Z., Zhu, J., Abe-Ouchi, A., Anagnostou, E., de Boer, A.M., Coxall, H.K., Donnadieu, Y., Foster, G., Inglis, G.N., Knorr, G., Langebroek, P.M., Lear, C.H., Lohmann, G., Poulsen, C.J., Sepulchre, P., Tierney, J.E., Valdes, P.J., Volodin, E.M., Dunkley Jones, T., Hollis, C.J., Huber, M., Otto-Bliesner, B.L., 2021. 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. https://doi.org/10.5194/cp-17-203-2021

Lunt, D.J., Farnsworth, A., Loptson, C., Foster, G.L., Markwick, P., O'Brien, C.L., Pancost, R.D., Robinson, S.A., Wrobel, N., 2016. Palaeogeographic controls on climate and proxy interpretation. Climate of the Past 12, 1181–1198. https://doi.org/10.5194/cp-12-1181-2016

Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., Dowsett, H.J., Loptson, C.A., 2012. On the causes of mid-Pliocene warmth and polar amplification. Earth and Planetary Science Letters 321–322, 128–138. https://doi.org/10.1016/j.epsl.2011.12.042

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., Zeebe, R.E., 2017. 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. https://doi.org/10.5194/gmd-10-889-2017

Marchitto, T.M., Curry, W.B., Lynch-Stieglitz, J., Bryan, S.P., Cobb, K.M., Lund, D.C., 2014. Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera. Geochimica et Cosmochimica Acta 130, 1–11. https://doi.org/10.1016/j.gca.2013.12.034

Martínez-Botí, M.A., Foster, G.L., Chalk, T.B., Rohling, E.J., Sexton, P.F., Lunt, D.J., Pancost, R.D., Badger, M.P.S., Schmidt, D.N., 2015. Plio-Pleistocene climate sensitivity evaluated using high-resolution CO2 records. Nature 518, 49–54. https://doi.org/10.1038/nature14145

McClymont, E.L., Ford, H.L., Ho, S.L., Tindall, J.C., Haywood, A.M., Alonso-Garcia, M., Bailey, I., Berke, M.A., Littler, K., Patterson, M.O., Petrick, B., Peterse, F., Ravelo, A.C., Risebrobakken, B., De Schepper, S., Swann, G.E.A., Thirumalai, K., Tierney, J.E., van der Weijst, C., White, S., Abe-Ouchi, A., Baatsen, M.L.J., Brady, E.C., Chan, W.-L., Chandan, D., Feng, R., Guo, C., von der Heydt, A.S., Hunter, S., Li, X., Lohmann, G., Nisancioglu, K.H., Otto-Bliesner, B.L., Peltier, W.R., Stepanek, C., Zhang, Z., 2020. Lessons from a high-CO_2 world: an ocean view from ~ 3 million years ago. Climate of the Past 16, 1599–1615. https://doi.org/10.5194/cp-16-1599-2020

McCrea, J.M., 1950. On the Isotopic Chemistry of Carbonates and a Paleotemperature Scale. The Journal of Chemical Physics 18, 849–857. https://doi.org/10.1063/1.1747785

Meckler, A.N., Sexton, P.F., Piasecki, A.M., Leutert, T.J., Marquardt, J., Ziegler, M., Agterhuis, T., Lourens, L.J., Rae, J.W.B., Barnet, J., Tripati, A., Bernasconi, S.M., 2022. Cenozoic evolution of deep ocean temperature from clumped isotope thermometry. Science 377, 86–90. https://doi.org/10.1126/science.abk0604

Mills, B.J.W., Krause, A.J., Scotese, C.R., Hill, D.J., Shields, G.A., Lenton, T.M., 2019. Modelling the long-term carbon cycle, atmospheric CO2, and Earth surface temperature from late Neoproterozoic to present day. Gondwana Research 67, 172–186. https://doi.org/10.1016/j.gr.2018.12.001

Modestou, S.E., Leutert, T.J., Fernandez, A., Lear, C.H., Meckler, A.N., 2020. Warm Middle Miocene Indian Ocean Bottom Water Temperatures: Comparison of Clumped Isotope and Mg/Ca-Based Estimates. Paleoceanography and Paleoclimatology 35, e2020PA003927. https://doi.org/10.1029/2020PA003927

Morice, C.P., Kennedy, J.J., Rayner, N.A., Jones, P.D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. Journal of Geophysical Research: Atmospheres 117. https://doi.org/10.1029/2011JD017187

O'Brien, C.L., Huber, M., Thomas, E., Pagani, M., Super, J., Elder, L.E., Hull, P.M., 2020. The enigma of Oligocene climate and global surface temperature evolution. Proceedings of the National Academy of Sciences 117, 25302–25309.

Osman, M.B., Tierney, J.E., Zhu, J., Tardif, R., Hakim, G.J., King, J., Poulsen, C.J., 2021. Globally resolved surface temperatures since the Last Glacial Maximum. Nature 599, 239–244. https://doi.org/10.1038/s41586-021-03984-4

Pagani, M., 2002. The alkenone-CO2 proxy and ancient atmospheric carbon dioxide. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 360, 609–632. https://doi.org/10.1098/rsta.2001.0959

Pagani, M., Liu, Z., LaRiviere, J., Ravelo, A.C., 2010. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. Nature Geosci 3, 27–30. https://doi.org/10.1038/ngeo724

Pearson, P.N., Palmer, M.R., 1999. Middle Eocene Seawater pH and Atmospheric Carbon Dioxide Concentrations. Science 284, 1824–1826. https://doi.org/10.1126/science.284.5421.1824

Pellichero, V., Sallée, J.-B., Chapman, C.C., Downes, S.M., 2018. The southern ocean meridional overturning in the sea-ice sector is driven by freshwater fluxes. Nat Commun 9, 1789. https://doi.org/10.1038/s41467-018-04101-2

Pellichero, V., Sallée, J.-B., Schmidtko, S., Roquet, F., Charrassin, J.-B., 2017. The ocean mixed layer under Southern Ocean sea-ice: Seasonal cycle and forcing. Journal of Geophysical Research: Oceans 122, 1608–1633. https://doi.org/10.1002/2016JC011970

Penman, D.E., Hönisch, B., Zeebe, R.E., Thomas, E., Zachos, J.C., 2014. Rapid and sustained surface ocean acidification during the Paleocene-Eocene Thermal Maximum. Paleoceanography 29, 357–369. https://doi.org/10.1002/2014PA002621

Rae, J.W.B., Zhang, Y.G., Liu, X., Foster, G.L., Stoll, H.M., Whiteford, R.D.M., 2021. Atmospheric CO_2 over the Past 66 Million Years from Marine Archives. Annual Review of Earth and Planetary Sciences 49, 609–641. https://doi.org/10.1146/annurev-earth-082420-063026

Raymo, M.E., Kozdon, R., Evans, D., Lisiecki, L., Ford, H.L., 2018. The accuracy of mid-Pliocene δ18O-based ice volume and sea level reconstructions. Earth-Science Reviews 177, 291–302. https://doi.org/10.1016/j.earscirev.2017.11.022

Raymo, M.E., Ruddiman, W.F., Backman, J., Clement, B.M., Martinson, D.G., 1989. Late Pliocene variation in northern hemisphere ice sheets and North Atlantic deep water circulation. Paleoceanography 4, 413–446. https://doi.org/10.1029/PA004i004p00413

Ring, S.J., Mutz, S.G., Ehlers, T.A., 2022. Cenozoic Proxy Constraints on Earth System Sensitivity to Greenhouse Gases. Paleoceanography and Paleoclimatology 37, e2021PA004364. https://doi.org/10.1029/2021PA004364

Roderick, M.L., Sun, F., Lim, W.H., Farquhar, G.D., 2014. A general framework for understanding the response of the water cycle to global warming over land and ocean. Hydrology and Earth System Sciences 18, 1575–1589. https://doi.org/10.5194/hess-18-1575-2014

Rohling, E.J., Foster, G.L., Gernon, T.M., Grant, K.M., Heslop, D., Hibbert, F.D., Roberts, A.P., Yu, J., 2022. Comparison and Synthesis of Sea-Level and Deep-Sea Temperature Variations Over the Past 40 Million Years. Reviews of Geophysics 60, e2022RG000775. https://doi.org/10.1029/2022RG000775

Rohling, E.J., Medina-Elizalde, M., Shepherd, J.G., Siddall, M., Stanford, J.D., 2012. Sea Surface and High-Latitude Temperature Sensitivity to Radiative Forcing of Climate over Several Glacial Cycles. Journal of Climate 25, 1635–1656. https://doi.org/10.1175/2011JCLI4078.1

Rohling, E.J., Yu, J., Heslop, D., Foster, G.L., Opdyke, B., Roberts, A.P., 2021. Sea level and deep-sea temperature reconstructions suggest quasi-stable states and critical transitions over the past 40 million years. Science Advances 7, eabf5326. https://doi.org/10.1126/sciadv.abf5326

Rosenthal, Y., Boyle, E.A., Slowey, N., 1997. Temperature control on the incorporation of magnesium, strontium, fluorine, and cadmium into benthic foraminiferal shells from Little Bahama Bank: Prospects for thermocline paleoceanography. Geochimica et Cosmochimica Acta 61, 3633–3643. https://doi.org/10.1016/S0016-7037(97)00181-6

Schmittner, A., Urban, N.M., Shakun, J.D., Mahowald, N.M., Clark, P.U., Bartlein, P.J., Mix, A.C., Rosell-Melé, A., 2011. Climate Sensitivity Estimated from Temperature Reconstructions of the Last Glacial Maximum. Science 334, 1385–1388. https://doi.org/10.1126/science.1203513

Scotese, C.R., Song, H., Mills, B.J.W., van der Meer, D.G., 2021. Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. Earth-Science Reviews 215, 103503. https://doi.org/10.1016/j.earscirev.2021.103503

Sherwood, S.C., Webb, M.J., Annan, J.D., Armour, K.C., Forster, P.M., Hargreaves, J.C., Hegerl, G., Klein, S.A., Marvel, K.D., Rohling, E.J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C.S., Foster, G.L., Hausfather, Z., von der Heydt, A.S., Knutti, R., Mauritsen, T., Norris, J.R., Proistosescu, C., Rugenstein, M., Schmidt, G.A., Tokarska, K.B., Zelinka, M.D., 2020. An Assessment of Earth's Climate Sensitivity

Using Multiple Lines of Evidence. Reviews of Geophysics 58, e2019RG000678. https://doi.org/10.1029/2019RG000678

Spero, H.J., Bijma, J., Lea, D.W., Bemis, B.E., 1997. Effect of seawater carbonate concentration on foraminiferal carbon and oxygen isotopes. Nature 390, 497–500. https://doi.org/10.1038/37333

Spratt, R.M., Lisiecki, L.E., 2016. A Late Pleistocene sea level stack. Climate of the Past 12, 1079–1092. https://doi.org/10.5194/cp-12-1079-2016

Tierney, J.E., Poulsen, C.J., Montañez, I.P., Bhattacharya, T., Feng, R., Ford, H.L., Hönisch, B., Inglis, G.N., Petersen, S.V., Sagoo, N., Tabor, C.R., Thirumalai, K., Zhu, J., Burls, N.J., Foster, G.L., Goddéris, Y., Huber, B.T., Ivany, L.C., Kirtland Turner, S., Lunt, D.J., McElwain, J.C., Mills, B.J.W., Otto-Bliesner, B.L., Ridgwell, A., Zhang, Y.G., 2020. Past climates inform our future. Science 370, eaay3701. https://doi.org/10.1126/science.aay3701

Tierney, J.E., Zhu, J., Li, M., Ridgwell, A., Hakim, G.J., Poulsen, C.J., Whiteford, R.D.M., Rae, J.W.B., Kump, L.R., 2022. Spatial patterns of climate change across the Paleocene–Eocene Thermal Maximum. Proceedings of the National Academy of Sciences 119, e2205326119. https://doi.org/10.1073/pnas.2205326119

Trenberth, K.E., Fasullo, J.T., 2013. An apparent hiatus in global warming? Earth's Future 1, 19–32. https://doi.org/10.1002/2013EF000165

Valdes, P.J., Scotese, C.R., Lunt, D.J., 2021. Deep ocean temperatures through time. Climate of the Past 17, 1483–1506. https://doi.org/10.5194/cp-17-1483-2021

von der Heydt, A.S., Köhler, P., van de Wal, R.S.W., Dijkstra, H.A., 2014. On the state dependency of fast feedback processes in (paleo) climate sensitivity. Geophysical Research Letters 41, 6484–6492. https://doi.org/10.1002/2014GL061121

Wang, H., Zhao, J., Li, F., Lin, X., 2021. Seasonal and Interannual Variability of the Meridional Overturning Circulation in the Subpolar North Atlantic Diagnosed From a High Resolution Reanalysis Data Set. Journal of Geophysical Research: Oceans 126, e2020JC017130. https://doi.org/10.1029/2020JC017130

Westerhold, T., Marwan, N., Drury, A.J., Liebrand, D., Agnini, C., Anagnostou, E., Barnet, J.S.K., Bohaty, S.M., Vleeschouwer, D.D., Florindo, F., Frederichs, T., Hodell, D.A., Holbourn, A.E., Kroon, D., Lauretano, V., Littler, K., Lourens, L.J., Lyle, M., Pälike, H., Röhl, U., Tian, J., Wilkens, R.H., Wilson, P.A., Zachos, J.C., 2020. An astronomically dated record of Earth's climate and its predictability over the last 66 million years. Science.

Winguth, A.M.E., Thomas, E., Winguth, C., 2012. Global decline in ocean ventilation, oxygenation, and productivity during the Paleocene-Eocene Thermal Maximum: Implications for the benthic extinction. Geology 40, 263–266. https://doi.org/10.1130/G32529.1

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. Science 292, 686–693. https://doi.org/10.1126/science.1059412

Zachos, J.C., Dickens, G.R., Zeebe, R.E., 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. Nature 451, 279–283. https://doi.org/10.1038/nature06588

Zachos, J.C., Stott, L.D., Lohmann, K.C., 1994. Evolution of Early Cenozoic marine temperatures. Paleoceanography 9, 353–387. https://doi.org/10.1029/93PA03266

Zeebe, R.E., 1999. An explanation of the effect of seawater carbonate concentration on foraminiferal oxygen isotopes. Geochimica et Cosmochimica Acta 63, 2001–2007. https://doi.org/10.1016/S0016-7037(99)00091-5

Zhang, Y., de Boer, A.M., Lunt, D.J., Hutchinson, D.K., Ross, P., van de Flierdt, T., Sexton, P., Coxall, H.K., Steinig, S., Ladant, J.-B., Zhu, J., Donnadieu, Y., Zhang, Z., Chan, W.-L., Abe-Ouchi, A., Niezgodzki, I., Lohmann, G., Knorr, G., Poulsen, C.J., Huber, M., 2022. Early Eocene Ocean Meridional Overturning Circulation: The Roles of Atmospheric Forcing and Strait Geometry. Paleoceanography and Paleoclimatology 37, e2021PA004329. https://doi.org/10.1029/2021PA004329

1373 Zhu, J., Poulsen, C.J., Tierney, J.E., 2019. Simulation of Eocene extreme warmth and high climate sensitivity
1374 through cloud feedbacks. Science Advances 5, eaax1874. https://doi.org/10.1126/sciadv.aax1874
1375 Ziveri, P., Thoms, S., Probert, I., Geisen, M., Langer, G., 2012. A universal carbonate ion effect on stable oxygen
1376 isotope ratios in unicellular planktonic calcifying organisms. Biogeosciences 9, 1025–1032.
1377 https://doi.org/10.5194/bg-9-1025-2012
1378