**RESEARCH ARTICLE**

**Special Collection:**
Illuminating a Warmer World: Insights from the Paleogene

**Key Points:**

- Deep ocean temperature changes are used to constrain global mean surface temperature yet the underlying assumptions lack detailed scrutiny
- Both curated data compilations and climate model simulations demonstrate that deep ocean-derived global mean surface temperature (GMST) estimates are robust
- We update the transformation equations and provide a revised estimate of GMST through the Cenozoic

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

D. Evans,
d.evans@soton.ac.uk

**Author Contributions:**

**Conceptualization:** David Evans
**Formal analysis:** David Evans, Julia Brugger, Gordon N. Inglis, Paul Valdes
**Methodology:** David Evans, Julia Brugger
**Resources:** David Evans
**Software:** David Evans, Julia Brugger
**Visualization:** David Evans
**Writing – original draft:** David Evans

# The Temperature of the Deep Ocean Is a Robust Proxy for Global Mean Surface Temperature During the Cenozoic

David Evans[1,2] ⓘ, Julia Brugger[3], Gordon N. Inglis[1] ⓘ, and Paul Valdes[4]

[1]School of Ocean and Earth Science, University of Southampton, Southampton, UK, [2]Formerly at Institute of Geosciences, Goethe University Frankfurt, Frankfurt am Main, Germany, [3]Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main, Germany, [4]School of Geographical Sciences, University of Bristol, Bristol, UK

**Abstract** Reconstructing global mean surface temperature (GMST) is one of the key contributions that paleoclimate science can make in addressing societally relevant questions and is required to determine equilibrium climate sensitivity (ECS). GMST has been derived from the temperature of the deep ocean ($T_d$), with previous work suggesting a simple $T_d$-GMST scaling factor of 1 prior to the Pliocene. However, this factor lacks a robust mechanistic basis, and indeed, 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 and two sets of paleoclimate model simulations. We show that models and data are in full agreement that a 1:1 relationship is a good approximation. Taken together, the two sets of climate models suggest that (a) a lower sensitivity of SST in the season of deep water formation than high latitude mean annual SST in response to climate forcing, and moreover (b) a greater degree 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 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.

## 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 requires a reconstruction of radiative forcing ($CO_2$) and global temperature at the very least (as $CO_2$ forcing is not the only potential factor that affects global mean surface temperature (GMST)), 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 paleoclimate 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 realms.

The use of deep-time paleoclimate 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 & 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 & Rae, 2016; Hönisch et al., 2012; Pagani, 2002; Pearson & 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 & 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 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).

Empirical data sets designed to reconstruct past changes in GMST can be broadly placed into two categories: (a) the synthesis of large amounts of spatially-distributed data sufficient to constrain the global climate at the time, or (b) the parameterization 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 paleotemperature 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 1,500 "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 3.1–4.5°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 data sets 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 & Schrag, 2003; Cramer et al., 2011; Lear et al., 2000; Lisiecki & 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 data sets 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 (e.g., 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 characterized by polar amplification (Burls et al., 2021; Cramwinckel et al., 2018; Evans et al., 2018b; Gaskell et al., 2022; Lunt et al., 2012) and overall increased ocean stratification (Green & 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 data sets (Section 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 (Sections 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 (a) we do not limit our analysis to climate model output, (b) we explicitly set out to test what relationship between GMST and deep ocean temperature is expected, for the entirety of the Cenozoic, and (c) present several advances in the way in which climate model data are evaluated for these purposes, detailed below (Section. 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 foraminifera (Evans, 2021), namely, the stable isotope (the oxygen isotopic ($\delta^{18}O_b$) or clumped isotope composition ($\Delta_{47}$) 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 Text S2 in Supporting Information S1. 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 utilized, 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 (a) 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, and (b) 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:

$$T_d = -4\delta^{18}O + 12 \tag{1}$$

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

$$T_d = 5–8\left(\delta^{18}O–1.75\right)/3 \tag{2}$$

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

$$T_d = 1–4.4\left(\delta^{18}O–3.25\right)/3 \tag{3}$$

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 Equations 1–3 to the most recent version of the benthic foraminiferal $\delta^{18}O$ stack (Westerhold et al., 2020) is shown in Figure 1a, alongside independent temperature proxies based on the Mg/Ca and clumped isotopic composition of benthic foraminifera.

Following the derivation of a deep ocean temperature record, H13 separate the Cenozoic into the Pleistocene, Pliocene, and pre-Pliocene. For the Pleistocene, H13 relate surface temperature (GMST) to the deep ocean by the following relationship:
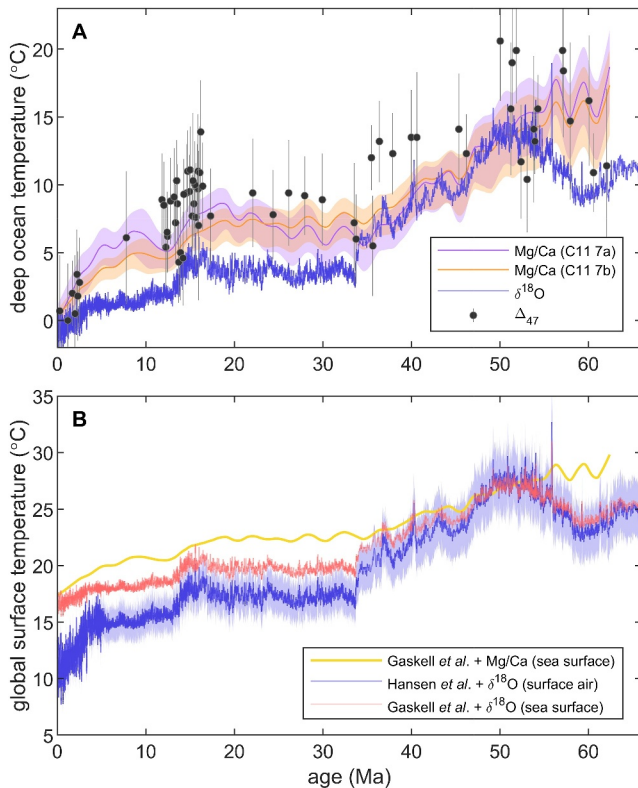
**Figure 1.** (a) Deep ocean temperature based on the oxygen isotope composition of foraminifera (Westerhold et al., 2020) converted to temperature following the approach of Hansen et al. (2013), the Mg to Ca ratio of foraminifera (Cramer et al., 2011), and the clumped isotopic composition of foraminifera ($\Delta_{47}$; 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 parameterized 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.

$$GMST = 2T_d + 12.25 \tag{4}$$

This is Equation 4.1 of H13, which was rooted in a Holocene-LGM $\Delta T$ 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:

$$GMST = 2.5T_d + 12.15 \tag{5}$$

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 GMST to $T_d$ in a 1:1 relationship, that is,

$$GMST^{t=t} - GMST^{t=5.3} = T_d^{t=t} - T_d^{t=5.3} \tag{6}$$

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 Equations 4–6 to both the $\delta^{18}O$ and Mg/Ca-derived deep ocean temperature records (Figure 1a) is displayed in Figure 1b. The focus of the analysis presented here is on the warm intervals of the early Cenozoic, and therefore it is Equation 6 which we most closely scrutinize.

A simple illustration of whether we might expect Equation 6 to be accurate is shown in Figure 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 Text S1 in Supporting Information S1 for a comprehensive description). The contours in Figure 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 and 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 more than the tropics are characterized by $\Delta GMST/\Delta T_{deep}$ relationships <1.

In order to determine the extent to which GMST could be overestimated by Equation 6 within the context of this simple calculation, surface ocean proxy data from three key intervals within the early Palaeogene are overlain on Figure 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 Equations 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
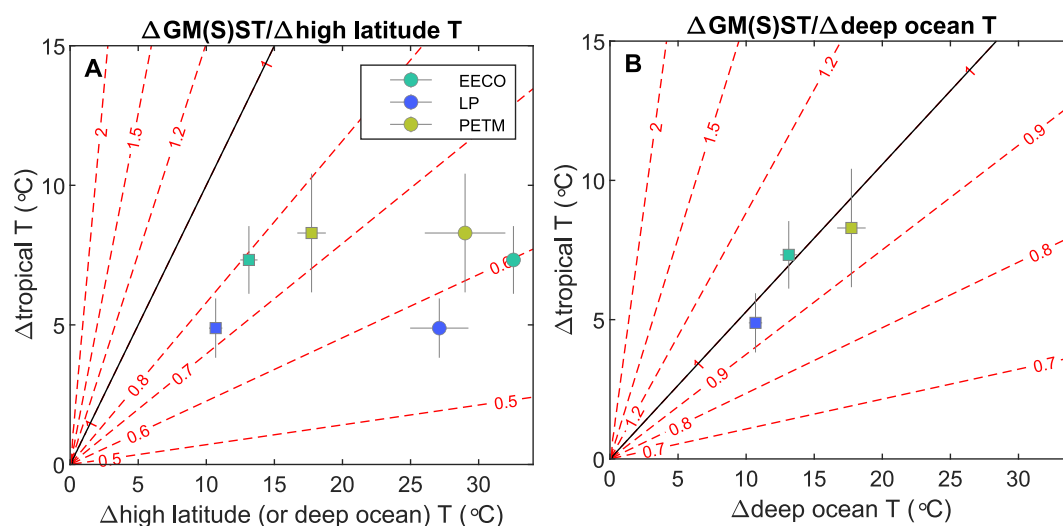
**Figure 2.** (a) The idealized relationship between GM(S)ST and high latitude/deep ocean temperature for a given degree of high latitude/deep ocean and tropical warming (contour lines, see text and Text S1 in Supporting Information S1 for details) for a zonally homogeneous ocean planet. Assuming deep ocean and high latitude SST are exactly coupled and in the absence of other processes, only in a world characterized 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 Figure 1 and the text for methodological details). (b) The same analysis assuming that deep water formation becomes increasingly biased toward winter as high latitude SST increases (or that SST in the months of deep water formation is less sensitive than the annual average to climate forcing; 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 toward higher deep ocean temperatures. We stress 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.

ocean may more closely represent mean annual high latitude SST than the direct surface ocean proxy data (Evans et al., 2018b). The results of this simple analysis confirm that, whichever data set is used to constrain high latitude SST, all three intervals are characterized 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 Figure 2a (e.g., ignoring for the moment that this exercise is based on a zonally homogeneous sphere with perfectly smoothly-varying latitudinal temperature), the data suggest that, in the absence of other processes driving the system in the opposite direction, the temperature of the global surface ocean would be expected to rise at a rate of ~0.6–0.8 times that of the deep ocean for this degree of polar amplification. If this simple analysis is applicable to Earth, then the corollary of this is that H13 likely over-estimate Palaeogene GMST on the order of 10s of percent of the deep ocean temperature change.

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 (Equation 6) 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 data sets (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 Equation 6 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; Lunt et al., 2012).

For H13 to accurately reconstruct GMST requires process(es) that either (a) drive GMST increase to a greater degree than the average global ocean, and/or (b) 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 characterized by the assumption of Equation 6, the magnitude of their combined effects must be exactly equal to the surface area-weighted difference between the change in global temperature relative to that of the high latitude oceans. 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. A greater degree of land surface versus sea surface warming/cooling in response to a $CO_2$ forcing could offset a polar amplification signal in high latitude/deep ocean SST. 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 & 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. Figure 1b; Gaskell et al. (2022)).

2. A bias in the season of deep water formation toward 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 characterized 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 toward 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 Figure 2b. Here, the same calculation of the relationship between polar amplification and the ratio of GMST/deep ocean temperature as in Figure 2a 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 (or similarly, a lower sensitivity of winter compared to mean annual temperature to climate forcing) 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 characterized 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 (assessed in detail in Section 3.3).

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: (a) Do we know the Cenozoic evolution of deep ocean temperature sufficiently well for it to find utility as a proxy for GMST? (Section 3.1), (b) What is the empirical slope of the relationship between deep ocean temperature and GMST? (Section 3.2), (c) Are fully coupled climate models characterized by a similar relationship, and what can model data tell us about the mechanistic basis for the deep ocean-GMST relationship? (Section 3.3), and (d) Should the approach of H13 be revised in light of this analysis, and if so, how? (Section 4). In each case, we introduce the relevant methodology and data sets 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: (a) the magnesium to calcium ratio (Mg/Ca), (b) the oxygen isotopic composition, and (c) clumped isotopic composition of the shell. Each of these has benefits/limitations summarized briefly in Text S2 in Supporting Information S1. 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 & Müller, 2012; Lear et al., 2015), the portioning of measured $\delta^{18}O$ changes between temperature, ice volume, and possibly 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 $\Delta_{47}$ (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 Figure 1a. Overall, the data sets are characterized 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 characterized 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 (Equations 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 $\delta^{18}O$ data sets 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 $\Delta_{47}$ temperatures may imply previously unidentified long-term $\delta^{18}O_{sw}$ shifts related to (e.g.,) climatically-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 (Figure 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 $\delta^{18}O$ data sets are characterized by very different resolutions, such that the apparent agreement between the two proxies during the PETM is an artifact 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 Section 4, we show, as a starting point, how the majority of the Cenozoic $\Delta_{47}$ 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., 2018b; Lear et al., 2015); revisiting the Paleogene portion of the Mg/Ca data set with these advances in mind may help to resolve the Eocene Mg/Ca-$\Delta_{47}$ offset. Irrespective, the central hypothesis to be tested here is that GMST and $T_d$ are characterized by a linear 1:1 relationship. Fortunately, doing so is reasonably insensitive to the deep ocean temperature proxy discrepancies (Figure 1) because the method of H13 anchors this 1:1 relationship to the base of the Pliocene (Equation 6). Given that the clumped isotope record is warmer than the $\delta^{18}O$ transformation (Equations 1–3) throughout most of the Cenozoic, the outcome of assessing this relationship empirically via the combination of deep ocean and surface temperature data sets (Section 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
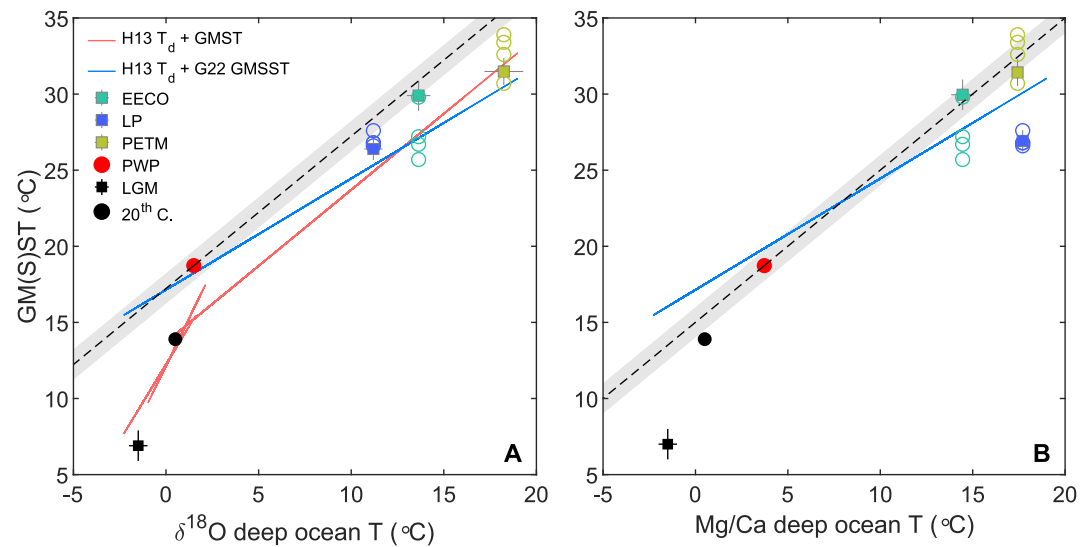
**Figure 3.** Empirical estimates of GMS(S)T as a function of deep ocean temperature for key Cenozoic intervals for which curated data compilation efforts exist (Dowsett et al., 2016; Hollis et al., 2019; Osman et al., 2021). Estimates of GMST for three DeepMIP target intervals (Inglis et al., 2020) 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 $\delta^{18}O$/Mg/Ca, see text for details. The gray shaded region depicts a 1:1 increase in deep ocean and GMST anchored to the Pliocene Warm Period (red circles), with an arbitrary $\pm 1°C$ 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.

Section 3.1 (Figure 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 Section 2 and Figure 2 (the LP, PETM, and EECO). We initially do so without revising either the $T_d$-GMST relationships of H13 or substantially revising any GMST estimate in order to test whether parameterizing GMST as a function of $T_d$ is empirically likely to be a fruitful methodology. Later (Section 4) we revisit and revise both aspects in detail, following this preliminary analysis and drawing on information from climate model simulations.

Global mean surface temperature for each interval is based on the following data sets: the LGM GMST used here is that of the data-model assimilation output of Osman et al. (2021), who estimated a LGM-pre industrial (PI) $\Delta T$ of $7.0 \pm 1.0°C$, that is, LGM GMST = 6.9°C based on a PI GMST of 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 data set (Dowsett et al., 2013, 2016), 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., 2018b; Hollis et al., 2012; Inglis et al., 2020), with GMSST 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 reasonably 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 Figure 3, calculated using the mean of all data within the sampled interval in all cases, using the H13 transformation equations in the case of $\delta^{18}O$ (Figure 1b) and Equation 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 $\Delta_{47}$-derived $T_d$ and GMST is given by Goudsmit-Harzevoort et al. (2023).

The data compilation shown in Figure 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 (Equations 1–6) 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 $<\sim 20$–$25°C$ (Haywood et al., 2020; Lunt et al., 2021; Valdes et al., 2021) but are broadly similar above this, discussed in more detail in Section. 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 $\sim 20°C$ (Figures 1a and 3a). If correct, this suggests that the Pliocene-Eocene portion of the analysis shown in Figure 3 should be limited to the relationship between $T_d$ and GMSST, given that the PWP data set 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 Section 4) we initially approach the proxy data analysis with both possibilities in mind.

*Oxygen isotope-based deep ocean temperature* (Figure 3a): Anchoring a 1:1 GMST-$T_d$ relationship to the mid-Piacenzian $\delta^{18}O$-derived deep ocean and PRISM SST data sets 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 Figure 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 $\sim 3°C$ 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 Figure 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 Figure 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 Figure 3a) adds support to this caveat. Either way, minor discrepancies exist, for example, the EECO GMSST estimate of this study (solid green symbol; Figure 3) is $\sim 3°C$ 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 data sets (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 Figure 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.

Lastly, 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 (Equation 4), which directly follows from the greater Holocene-LGM $\Delta T$ than that utilized by H13 (7 vs. 4.5°C).

*Mg/Ca-based deep ocean temperature* (Figure 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 $\sim 3°C$ warmer, whereas the early Paleogene deep ocean temperatures are broadly not (except in the Paleocene; Figure 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 (Figure 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 artifact of the Mg/Ca data or transformation, given that in this analysis the LP and PETM have similar $T_d$ but were clearly characterized by very different climate states (Dunkley Jones et al., 2013; Penman et al., 2014). The reason for this is likely rooted 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 (we note that the $\delta^{18}O_{bf}$ data for this interval are reproducible between sites, arguing against a diagenetic effect in that case (Evans et al., 2018a; Westerhold et al., 2020), whereas relatively little Mg/Ca data exist). Assuming the PWP GMSST reconstruction overestimates GMST (see discussion above) would alternatively suggest that $T_d$ underestimates PETM (and possibly EECO) GMST as the red data point and black dashed line anchored to it in Figure 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 Figure 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 data sets may nonetheless fall within uncertainty of each other.

In conclusion, irrespective of which deep ocean temperature data set is used, and whether the surface temperature estimates based on the data compilations utilized 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 colored solid data points in Figure 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 data sets: (a) the DeepMIP set of model simulations (Lunt et al., 2017, 2021), which incorporates eight climate models run under different $p$CO$_2$ but otherwise similar (early Paleogene) boundary conditions between models, and identical boundary conditions within a set of model simulations, plus associated PI controls, and (b) the Cenozoic portion of the Phanerozoic HadCM3 simulations of Valdes et al. (2021), which includes 12 simulations (1 per Stage) with varying paleo-geography and other boundary conditions, run under two (broadly similar) $p$CO$_2$ within each time slice, referred to as "F17" and "smoothed" CO$_2$ respectively, where CO$_2$ is derived from the same underlying data set but within a given interval taken either from either the LOWESS fit of Foster et al. (2017) or the much smoother fit of Valdes et al. (2021). These latter simulations were not part of DeepMIP and are hereafter referred to as HadCM3L$_{V21}$ to distinguish them from the DeepMIP HadCM3(L) simulations. 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 modeled 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 per 1,000 years; Valdes et al. (2021)), with the exception of the 9 × CO$_2$ CESM simulation (~1.5°C in the final kyr; Figure 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 3,000 m, as Goudsmit-Harzevoort et al. (2023) show that global mean $T_d$ varies by ~<1°C in response to chosen cutoff depth below this point.

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 characterized by a $T_d$-GMST slope close to 1 based on the entire ensemble (excluding the $9 \times CO_2$ CESM simulation and PI controls), with $m = 1.026$ (Figure 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 $\pm 2°C$, with the majority falling within $\pm 1°C$ (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 (Section. 3.2) and climate models. The HadCM3L$_{V21}$ simulations with variable boundary conditions paint a similar picture, with most falling with $\pm 2°C$ of a 1:1 line anchored to the mean of the two simulations at 3 Ma (Figure 4c). The exception to this are the Miocene simulations, which are characterized 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 closer to, albeit still lower than, unity ($m = 0.824$). 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 total, it is only the six (out of 24) 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 $p$CO$_2$ overall provide strong evidence in favor 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 $p$CO$_2$ changes within a given interval), as paleogeography related phenomena are capable of driving large changes in GMST independent of $p$CO$_2$ (Caballero & 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 (Figure 1a), possibly suggests an issue with the Miocene model data, and highlights the need for further Miocene $p$CO$_2$ estimates (Valdes et al., 2021).

Both sets of model simulations are characterized 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-$T_d$; Figure 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 characterized by a steeper relationship and GMSST consistently ($\sim$2–3°C) warmer than the data suggest for a given $T_d$ at $p$CO$_2 > 2 \times$ PI (Figure 4b). While the HadCM3L$_{V21}$ data set is characterized 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; Hollis et al., 2019; Kiehl & Shields, 2013; Lunt et al., 2012; Lunt et al., 2021; Sagoo et al., 2013), 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) (upper-right data points in Figure 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 Section 2, despite polar amplification and the stratification of the ocean in warm climate states, using both the DeepMIP and HadCM3L$_{V21}$ simulations. That is: (a) whether a differential sensitivity of SST in the season of deep water formation compared to the annual mean,
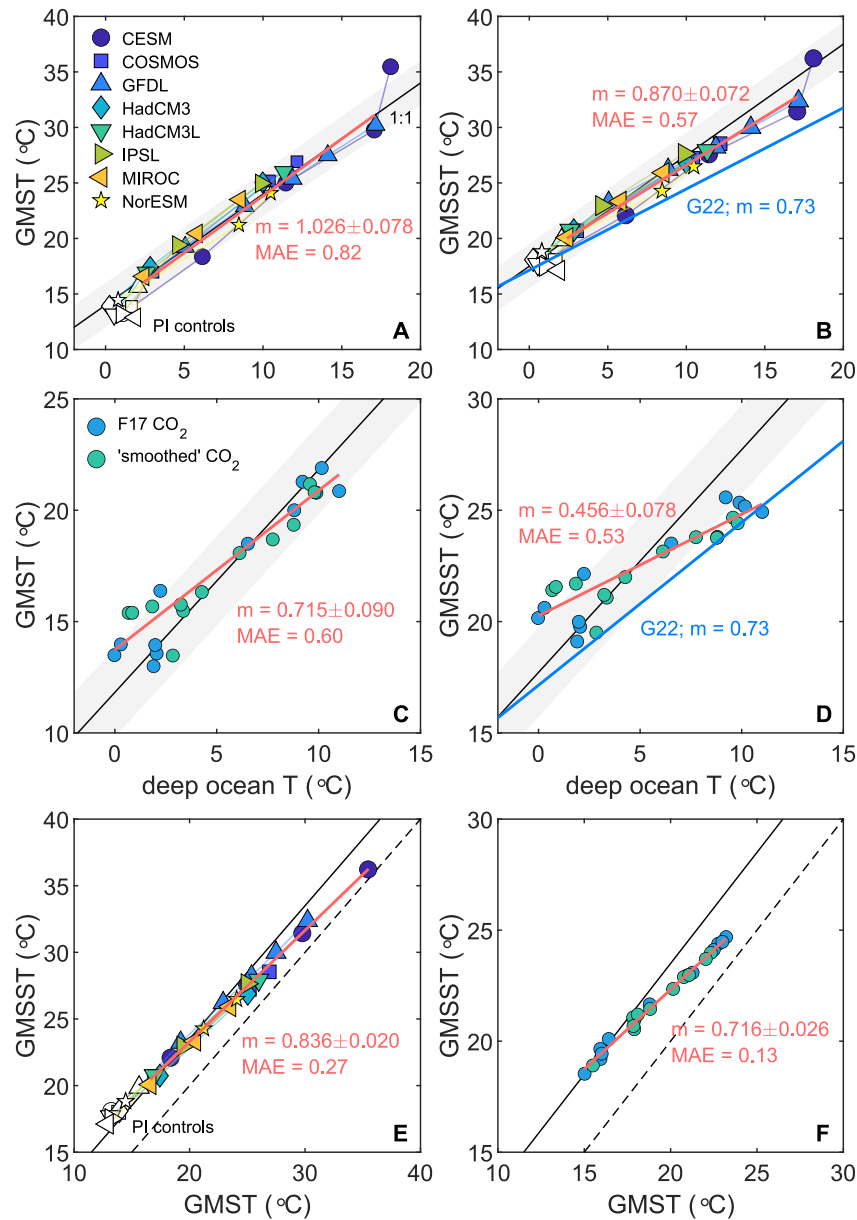
**Figure 4.** (a) Global mean surface temperature and (b) Global mean sea surface temperature as a function of deep ocean temperature (>3,000 m) in the DeepMIP set of model simulations, see panels C and D for $x$ axis labels. One-to-one lines are anchored to the mean of the simulations conducted at $1 \times CO_2$ and Eocene paleogeography plus the IPSL simulation at $1.5 \times CO_2$; shaded regions depict $\pm 1$ and $2°C$ from this line. The least squares linear regressions (red lines) include all model simulations with Eocene paleogeography, except for the $9 \times CO_2$ 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 $1 \times 1°$ grid before further calculations were performed.

and (b) if a greater sensitivity of land versus ocean surface temperature to climate forcing counteracts the effect of polar amplification to result in a 1:1 $T_d$-GMST relationship.

*Is winter SST in the regions of deep water formation less sensitive to climate forcing that mean annual high latitude SST?* 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:
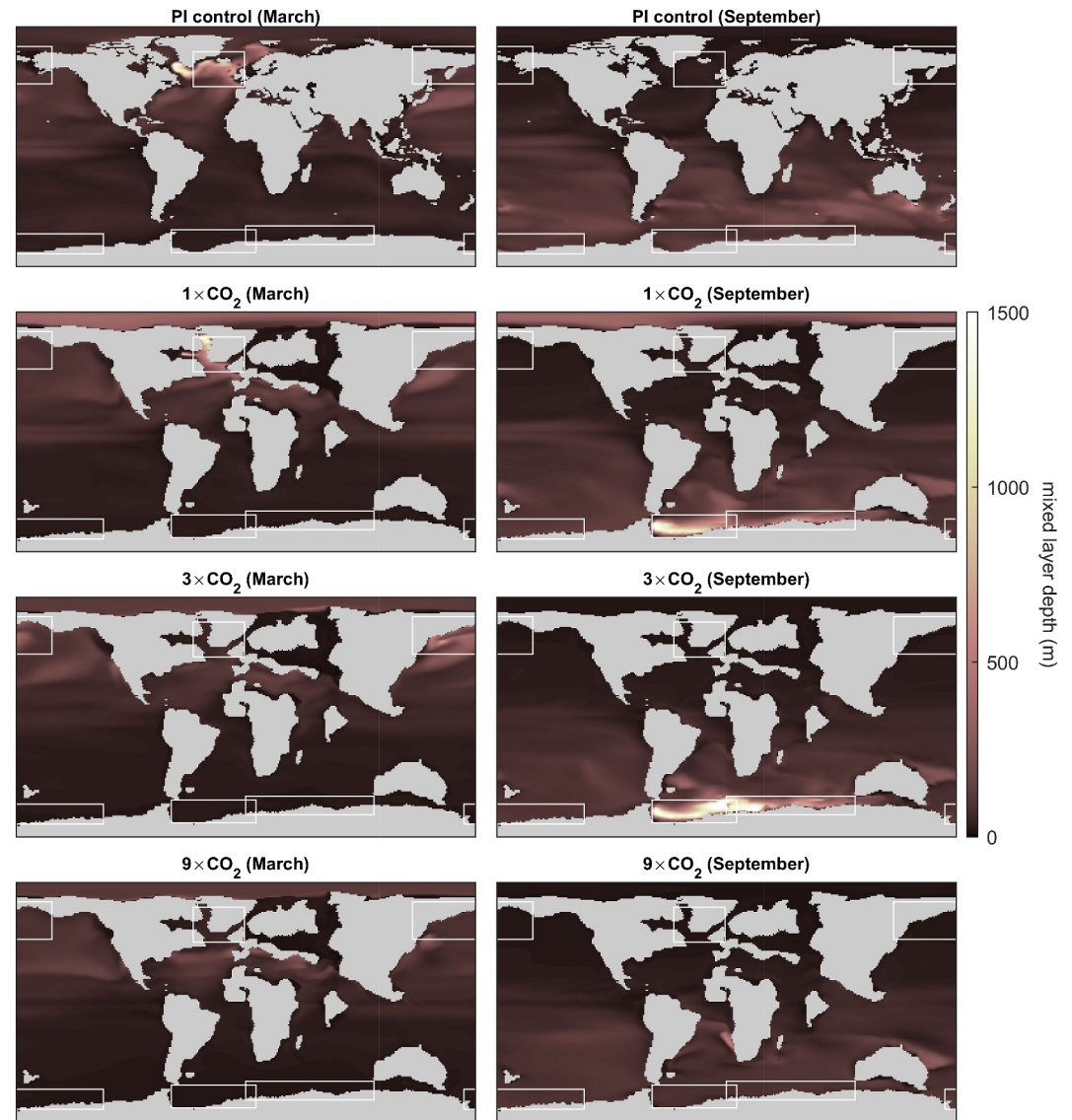
**Figure 5.** March and September mixed layer depth in the four DeepMIP CESM simulations (PI control, as well as 1, 3, and $9 \times CO_2$). Boxes showing the key regions of deep water formation in the full suite of models contributing to the DeepMIP set of simulations are overlain (Zhang et al., 2022).

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 Figure 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 initially 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 Table S1 in Supporting Information S1. In cases where this is ambiguous, such as the $9 \times CO_2$ CESM simulation (Figure 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). In our later more in-depth analysis, we define the month of deep-water formation more flexibly on a simulation-specific basis (see below).
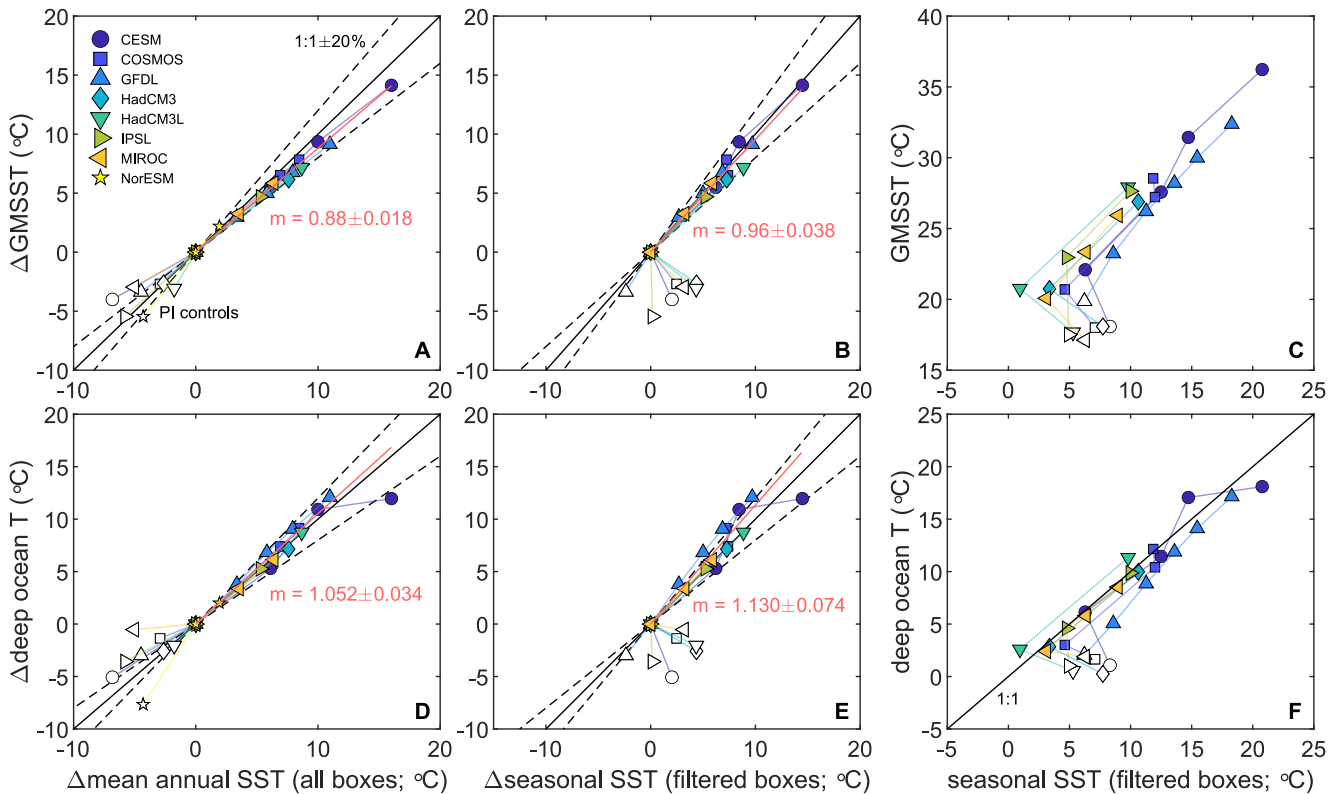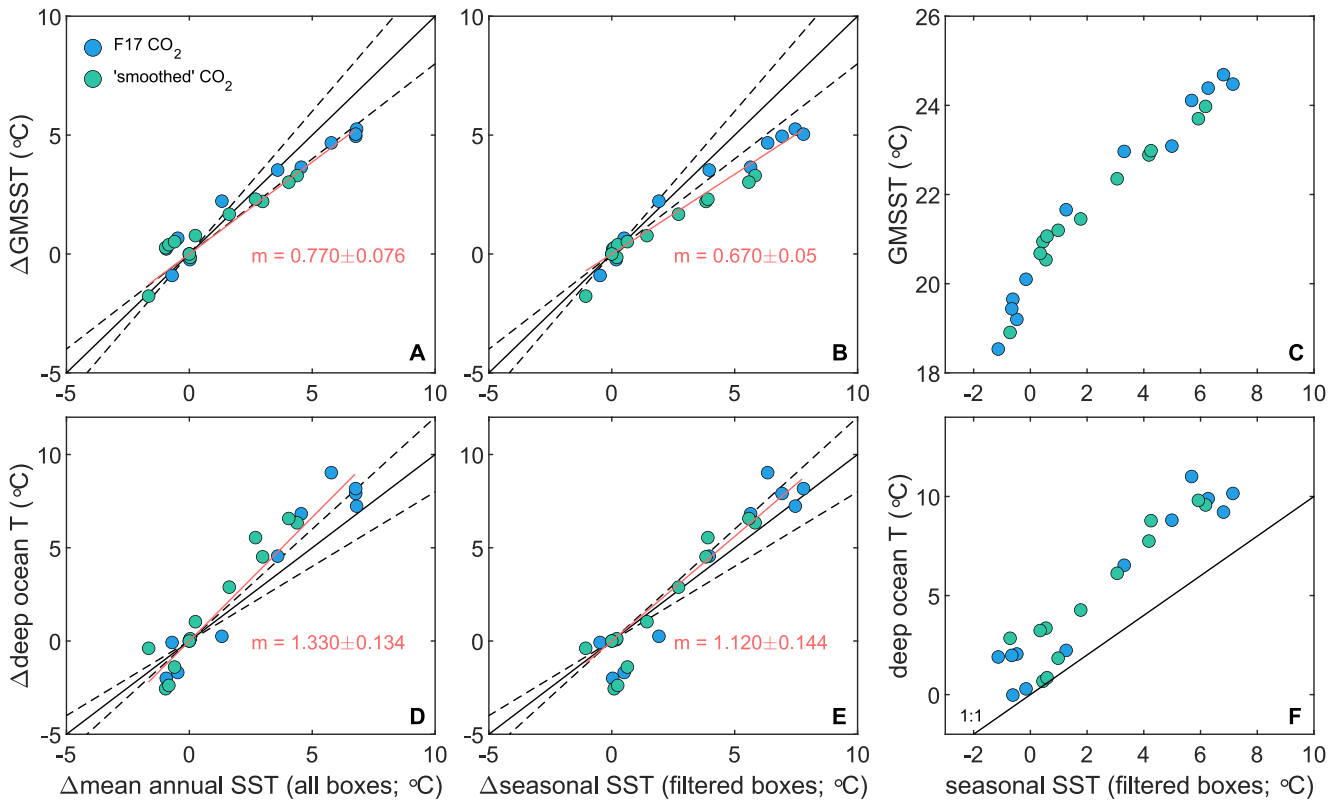
**Figure 6.** The relationship between sea surface temperature (0–100 m) in the broad regions of deep-water formation and global mean sea surface temperature (GMSST) or deep ocean temperature in the DeepMIP set of model simulations. (a, d) Mean annual SST in all boxes shown in Figure 5 plotted relative to the $1 \times CO_2$ simulation with Eocene paleogeography (except IPSL; $1.5 \times CO_2$). (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 $9 \times CO_2$ CESM simulation was excluded from the fit.

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 Figures 6 and 7 for the DeepMIP and HadCM3L$_{V21}$ simulations respectively, in all cases anchored to (a) the simulation with the lowest $p$CO$_2$ but Eocene paleogeography in the case of the DeepMIP simulations, as the closest possible representation of Pliocene-like conditions, and (b) 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., Figures 6c and 6f). 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 is more sensitive than GMSST to CO$_2$ (ensemble $m = 0.88$; Figure 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 (Figure 5; Table S1 in Supporting Information S1) is a simple method of removing the summer bias in

**Figure 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 Figure 5 plotted relative to the 3 Ma simulation. (b, e) Winter SST in the simulation-specific box(es) relevant for 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.

$MASST_{HL}$ in the comparison. Indeed, doing so results in an increase in the whole ensemble $GMSST-SST_{HL}$ slope ($m = 0.96$; Figure 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 $SST_{HL}$ (Figures 6d and 6e), which is overall characterized 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 vs. 0.074), implying slightly worse agreement between $SST_{HL}$ and $T_d$ in absolute terms across the ensemble (Figure 6f). An increased degree of temperature increase in the deep ocean than high latitude SST (Figure 6e) is physically implausible and implies a limitation of this simple analysis (e.g., the inclusion of grid cells in the high latitude boxes (Figure 5) that do not contribute to overturning, see below), that a portion of deep water is formed outside of the months and/or boxes considered here, or that this is an inherent limitation of using mean monthly output in addressing an instantaneous process in the model (convection) that may be better-resolved using time-step resolution output (unavailable for these simulations). Overall, however, this analysis must mean that seasonally biased deep water formation is important in offsetting $T_d$ from $MASST_{HL}$, 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 differ in that a greater polar amplification signal is present in the regions of deep water formation, that is, the $\Delta GMST-\Delta SST_{HL}$ slope is shallower than that of the DeepMIP ensemble (0.77 vs. 0.88; Figure 7a; Figure 6a), and accounting for the seasonality and relevant location of deep water formation results in a further reduced $\Delta GMSST-\Delta SST_{HL}$ slope (Figure 7, compare panels A and B). This is the case to a lesser degree for the HadCM3(L) DeepMIP simulations, highlighting the potential impact of paleogeography over $CO_2$ alone, although we also note that the version of HadCM3L utilized 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., 2018b; Lunt

et al., 2021). The HadCM3L$_{V21}$ simulations are also characterized by a $T_d$-SST$_{HL}$ slope >1, irrespective of whether the relevant boxes and season of deep water formation are taken into account (Figures 7d and 7e), and absolute deep ocean temperature substantially offset from winter SST in the boxes relevant for deep water formation (Figure 7f). This is also not the case for the DeepMIP HadCM3 simulations and is mostly easily explicable as a limitation of this 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 (excluding sea ice-covered grid cells has no appreciable impact on the analysis). With these caveats in mind we note that, unlike the DeepMIP ensemble, there is no evidence for a lower sensitivity of SST in the months of deep water formation compared to MASST contributing to a 1:1 $T_d$-GMST relationship in the analysis shown in Figure 7. Rather, the lack of this phenomenon in HadCM3L$_{V21}$ is offset by the lower GMST-GMSST slope in these compared to the DeepMIP simulations (Figures 4e and 4f), indicating that the two sets fall onto a 1:1 $T_d$-GMST relationship for different reasons. We also note that the HadCM3L$_{V21}$ simulations exhibit a switch in the main location of deep water formation around a threshold of ∼300 ppm $CO_2$, such that anchoring the above analysis to the Pliocene incorporates an effect that is not present in the majority of the DeepMIP analysis, in that the location of deep water formation is broadly consistent between the simulations with Eocene paleogeography. Alternatively anchoring our HadCM3L$_{V21}$ analysis to the 11 Ma Miocene simulations in order to understand the extent to which the results shown in Figure 7 are driven by a switch in the location of deep water formation does change the details of the analysis (i.e., the regression slopes; Figure S2 in Supporting Information S1) but not the overall patterns described above or below.

Both the DeepMIP and HadCM3L$_{V21}$ simulations overall provide strong evidence that deep ocean temperature, GMSST, and high latitude SST are 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. Figures 6b and 6e; Figures 7b and 7e), 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 (Figure 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, Figure 8 shows the same analysis of the DeepMIP simulations as in Figure 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 Figure 5 that are not relevant for deep water formation. For example, in the $3 \times CO_2$ 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 (Figure 5). In addition, we use MLD to define the month or season of deep water formation, as opposed to considering only winter months in our analysis (point 2 above) but note that imposing this constraint or otherwise has no substantial impact on the results.

This more nuanced analysis highlights a stronger degree of polar amplification in both mean annual and seasonal SST in grid cells 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 ΔGMSST in the model ensemble, which is characterized by a slope that is reduced to 0.81 ± 0.05 in the MLD analysis compared to $m = 0.88 \pm 0.02$ when using all grid cells in the high latitude boxes (cf. Figures 6a and 8a). A similar reduction in slope is observed when comparing the seasonal SST in the grid cells characterized 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 (Figures 6b and 8b; with CESM being a notable exception when comparing the results of the analysis within individual models). Therefore, winter SST in these grid cells warms more than global average MASST in the high latitudes, which means that overall, while a lower sensitivity of SST$_{HL}$ in the months of deep water formation than the annual average 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; Figure 8b), a polar amplification signal is present in subducting water (as is also the case for HadCM3L$_{V21}$), such that a seasonal bias in deep water formation cannot be the only process resulting in a 1:1 GMST-$T_d$ relationship (Figure 4a). Unsurprisingly, the MLD analysis fully resolves the discrepancy between SST$_{HL}$ and $T_d$ observed previously (Figure 6e). Considering only the relevant grid cells results in a slope of 0.99 ± 0.07 (Figure 8d), that is, the two parameters have the same value.

Performing the same analysis for the HadCM3L$_{V21}$ simulations (Figure 9) results in an overall very similar outcome, albeit substantially increasing the variance in the data. Specifically, the slope of the relationship
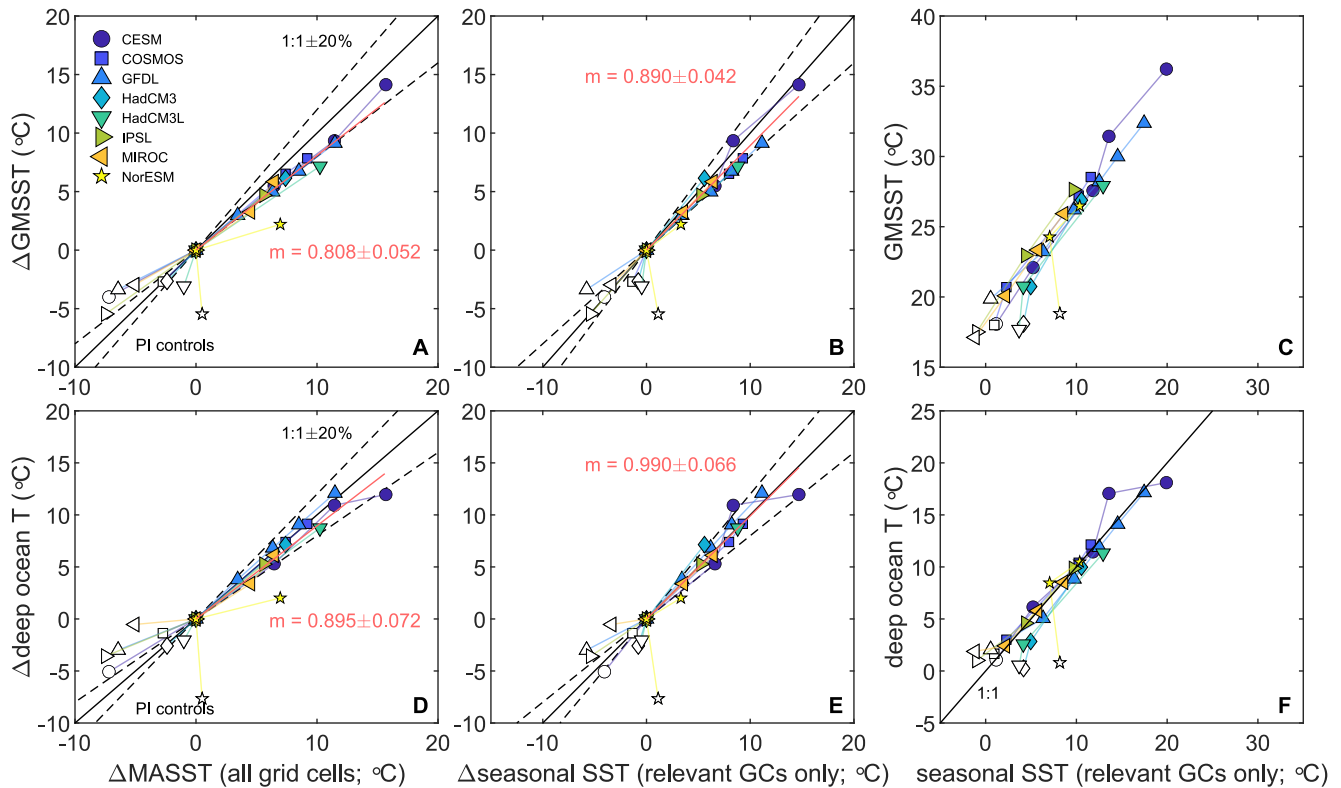
**Figure 8.** The relationship between SST in the model grid cells with a mixed layer depth (MLD) at least 90% of the global seasonal maximum and global mean sea surface temperature (GMSST) or deep ocean temperature in the DeepMIP set of model simulations. (a, d) Mean annual SST in all grid cells meeting the MLD criteria plotted relative to the $1 \times CO_2$ simulation with Eocene paleogeography (except IPSL; $1.5 \times CO_2$). (b, e) As in panels A/D except mean SST was calculated from all grid cells (GCs) 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 $6\times$ and $9 \times CO_2$ CESM simulations were excluded from the fit.

between $SST_{HL}$ and GMSST decreases from 0.77 to 0.46 and 0.67 to 0.49 when using mean annual SST or winter SST in the relevant high latitude grid cells respectively (compare Figures 7a, 7b and 9a, 9b). That is, an even more pronounced polar amplification signal is present in the grid cells with a deep mixed layer depth. As is also the case for the DeepMIP simulations, considering only the grid cells with a deep mixed layer depth results in a $SST_{HL}$-$T_d$ relationship closer to 1, as expected (compare Figures 7e and 9e), and brings the two onto a 1:1 relationship (Figure 9f), resolving this contradiction in the simpler analysis (Figure 7f). The increased variance introduced in this analysis is a result of averaging over fewer grid cells (and is largely driven by a small number of Miocene simulations), coupled with the specific characteristics of deep water formation in these simulations. For example, the 36 Ma "smoothed $CO_2$" simulation is characterized by a substantially (7.4°C) warmer than modern winter SST in the grid cells with a deep MLD (in this case in the Pacific sector of the Southern Ocean) compared to the 3 Ma simulation, while GMSST increases by just 2.3°C (Figure 9e).

The above analysis demonstrates that winter $SST_{HL}$ is less sensitive to climate forcing than the annual average, offsetting deep ocean temperature from the effects of polar amplification in the regions of deep water formation to a substantial degree in the DeepMIP simulations, but is, alone, insufficient to mechanistically explain a 1:1 $T_d$-GMST relationship and is potentially not the main mechanism responsible in the case of $HadCM3L_{V21}$ based on the MLD analysis. The other key factor, as noted by Hansen et al. (2013) and Goudsmit-Harzevoort et al. (2023) is that land surface air temperature is more sensitive to $CO_2$ forcing than SST. This is the case to a greater degree in the $HadCM3L_{V21}$ simulations than the DeepMIP ensemble (GMSST vs. GMST $m = 0.84$ and 0.72 respectively, Figures 4e and 4f), 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 characterized by a steeper slope than $HadCM3L_{V21}$ ($m = 0.795$ vs. 0.716). Nonetheless, in the case of the DeepMIP ensemble, the slopes between $\Delta$GMSST and $\Delta SST_{HL}$ (accounting for seasonality in the relevant high latitude boxes) and
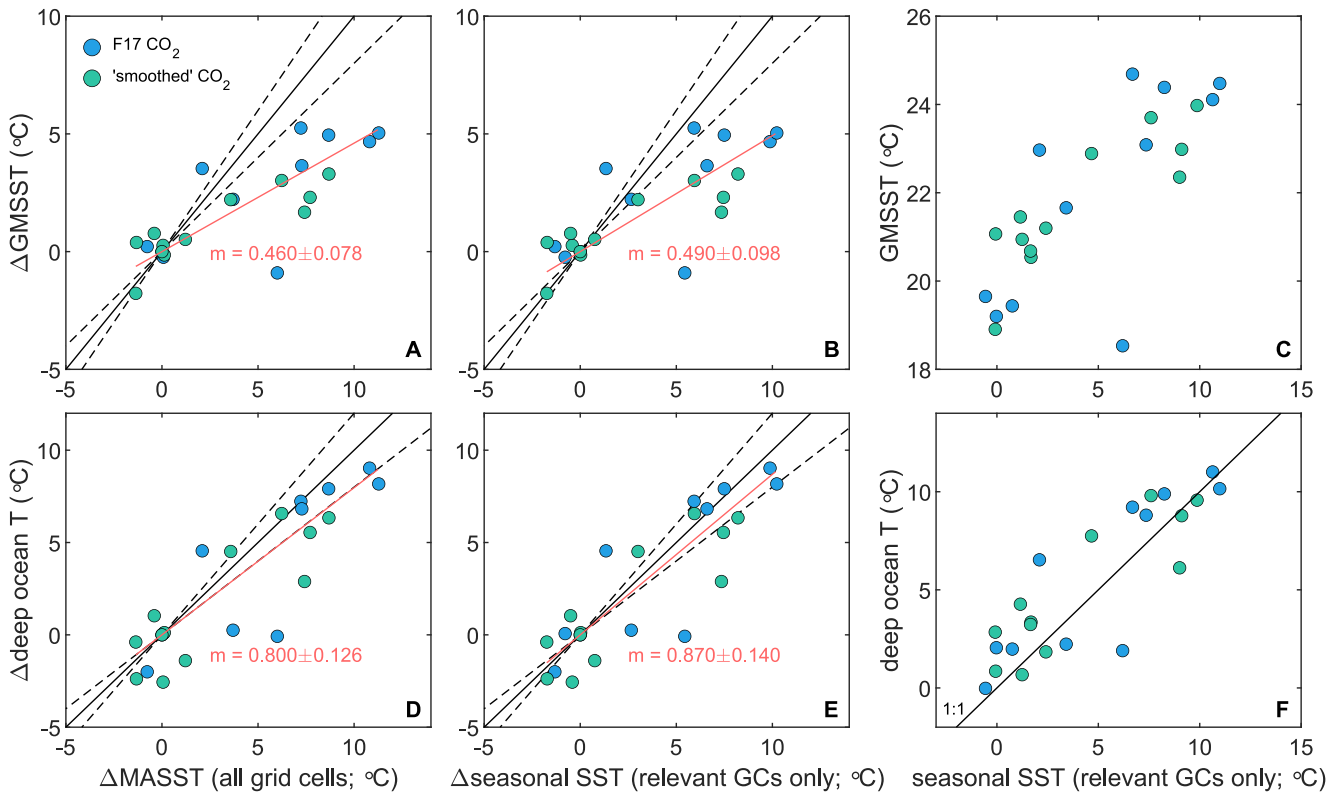
**Figure 9.** The relationship between SST in the model grid cells with a mixed layer depth (MLD) at least 90% of the global seasonal maximum and global mean sea surface temperature (GMSST) or deep ocean temperature in the HadCM3L$_{V21}$ set of model simulations. (a, d) Mean annual SST in all grid cells meeting the MLD criteria plotted relative to the 3 Ma simulation. (b, e) As in panels A/D except mean SST was calculated from all grid cells (GCs) meeting the MLD criteria during the season of maximum mixed layer depth, and limited to the southern hemisphere (as this is the dominant locus of deep water formation in all simulations >0 Ma). Data from the 0 Ma simulation are not included in the regression slops as these fall off of the trend largely as a result of being characterized 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 (note that panel F shows a 1:1 line rather than a regression).

GMSST-GMST are almost identical ($m = 0.89$ and 0.84, respectively; Figures 4e and 8b), which is also the case for the simpler HadCM3L$_{V21}$ high latitude data analysis (i.e., assuming that the more nuanced MLD analysis introduces too much variance in the results to identify this phenomenon), with $m = 0.67$ and 0.72, respectively (Figures 4f and 7b). Overall, this is the explanation for a 1:1 $T_d$-GMST relationship as hypothesized by H13 and the reason that this is an emergent model property (Figure 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*, that is, 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 paleogeography and constant paleogeography 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 characterized by a 1:1 relationship prior to the Pliocene. Crucially, given that this is empirically the case for the early Palaeogene, 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 Section 3.3.2 is correct. Nonetheless, the GMST approach of

[H13] requires revision, particularly in light of our greatly improved understanding of (a) the Cenozoic evolution of $T_d$, including a more thorough grasp of the nonthermal controls on some key temperature proxies, and (b) the Cenozoic evolution of continental ice volume, that is, 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 data set is unparalleled by the other proxy data sets (Figure [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 (a) revisit the ice volume/sea level component of $\delta^{18}O_b$ using the analysis of Rohling et al. ([2022]), (b) explore the impact of a pH/[$CO_3^{2-}$] correction on $\delta^{18}O_b$, and (c) 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., [2021], [2022]; Spratt & 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 $-0.25\%\circ^\circ C^{-1}$. 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 (Equation 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 $-1.50\%\circ$ per pH unit (see Text S3 in Supporting Information [S1]), which is intermediate between the two planktonic foraminifera species for which data are available ($-0.89$ and $-2.51\%\circ$ 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]). Parameterizing 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 & 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 data set is low-resolution (31 datapoints spanning the last ∼60 Ma) and contains no data between the mid-Eocene and Miocene. In addition, secular changes in the boron isotopic composition of seawater are not well-constrained for parts of the Cenozoic (Rae et al., [2021]) but is required information in deriving pH from $\delta^{11}B$. For these reasons, 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 artifacts related to the long-term smooth applied to the pH data.

The resulting Cenozoic $T_d$ reconstruction is shown in Figure [10a], with that of Hansen et al. ([2013]) and independent estimates of deep ocean temperature for comparison (Figure [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
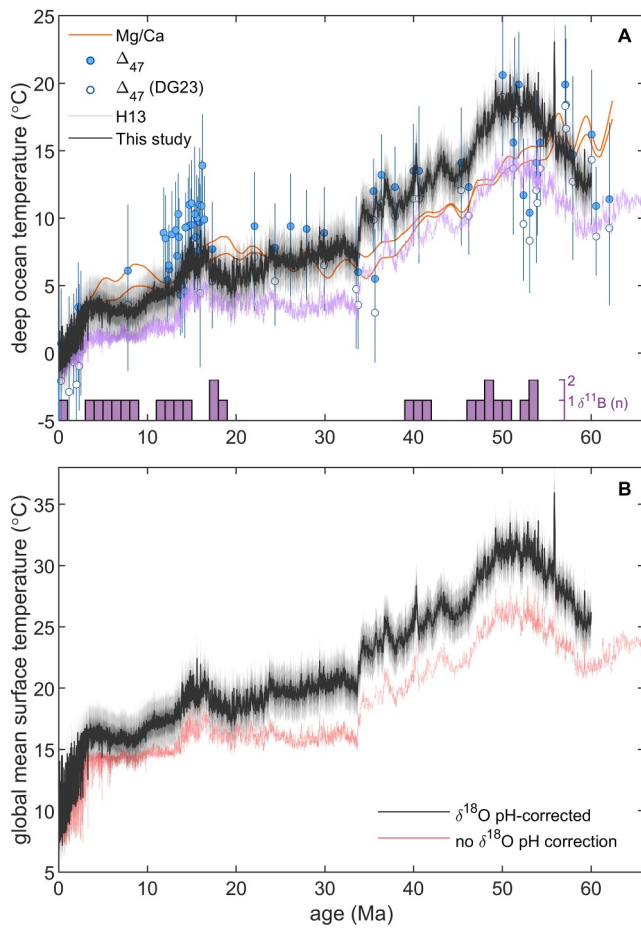
**Figure 10.** (a) A revised estimate of the Cenozoic evolution of deep ocean temperature based on the sea-level and pH-corrected benthic foraminifera oxygen isotope stack (black line with an arbitrary ±2°C uncertainty, see text for details) in the context of other proxy estimates (see Figure 1), including the $\Delta_{47}$ reanalysis of Daëron and Gray (2023). The histogram shows the density of the benthic foraminifer boron isotope measurements used to pH correct $\delta^{18}O_{bf}$. (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.

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 $\Delta_{47}$ 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 (Figure 10a). 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 $\Delta_{47}$ but not $\delta^{18}O$, and in the Miocene, wherein the majority of the $\Delta_{47}$ 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 $\Delta_{47}$ 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 $\Delta_{47}$ data from this interval is a truly global signal and to determine the cause of the remaining discrepancies within the Eocene portion of the data sets.

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 characterized 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 characterized and the assumption that the relationship between $T_d$ and GMST remains constant within each of these portions of the data set. 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., 2013, 2016), coupled with the minimum reconstructed $T_d$ 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 $\Delta T_d = 0$ relative to the youngest $T_d$ 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 (Figures 4e and 4f) 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 paleogeography. However, the calculations are overall insensitive to this choice; alternatively using the DeepMIP set of simulations would result in pre-Pliocene GMST 0.45°C cooler. Specifically, this results in the following transformation equations:

$$GMST = 2.23\Delta T_d + 13.9 \qquad (7)$$

For all $T_d$ lower than the present average deep ocean (1.5°C),

$$GMST = 1.20\Delta T_d + 13.9 \qquad (8)$$

For all $T_d$ between the present average and that of the interval 3.0–3.05 Ma (2.3°C higher than present based on our conversion of $\delta^{18}O_{bf}$ to temperature described above), and
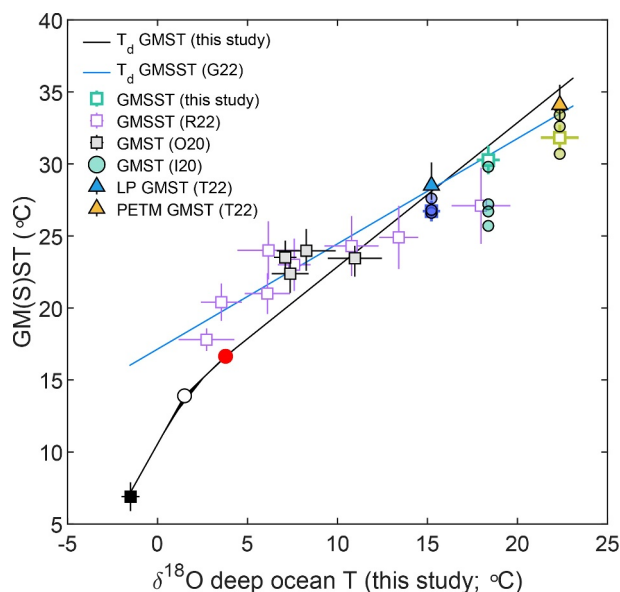
**Figure 11.** The revised relationship between deep ocean temperature ($T_d$) and GMST (black line, this study), showing the three anchor points used here (LGM, 20th Century, and PWP, black square, white circle, and red circle respectively; see text for details). Following our analysis and H13, the 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 $T_d$ (Figure 10)), 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).

$$GMST = 1.20 \times 2.3 + (\Delta T_d - 2.3) + 13.9 \quad (9)$$

For all $T_d$ warmer than the average during the interval 3.0–3.05 Ma. The equivalent scaling factors and Pliocene deep ocean temperature are 2.22, 1.97, and 0.16°C respectively in the case that the oxygen isotope stack is not pH corrected.

The resulting GMST estimates, both with and without a pH correction on $\delta^{18}O$, are shown in Figure 10b. 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, for example, 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 data set. 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 (Figure 3) while the climate model simulations interrogated here suggest that it lies between 1.03 (DeepMIP) and 0.86 (HadCM3L$_{V21}$, or 0.72 if all simulations are included, see above), see Figure 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 Figure 10, 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 (Figure 10) and that of H13 derive from (a) the revision of the LGM-modern $\Delta$GMST 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 (b) 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 ~31.0°C in this study (Figure 10b)). The revised Cenozoic GMST reconstruction agrees well with several independent lines of evidence from both models and proxy data (Figure 11). For example, the Pliocene Model Intercomparison Project range ($\Delta$GMST = 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 (Figure 11). 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 Figure 11. 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 (Figures 4e and 4f), with the two converging only under early Paleogene-like global warmth. This discrepancy potentially points to a systematic bias in the terrestrial proxy records (also discussed in Ring et al. (2022)), which warrants further investigation, rather than a failure of climate models to capture this feature of Earth's climate.

In the early Paleogene, the PETM and pre-PETM GMST estimates of Tierney et al. (2022) of 34.1°C (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, respectively; note that the resolution of the core-PETM $\delta^{18}O$ data in the stack utilized here is insufficient to place a precise estimate on the PETM using this approach). Our analysis constrains EECO GMST to 31.3 ± 1.3°C, slightly higher than the upper

range of the estimates provided in Inglis et al. (2020) but within uncertainty of both the data analysis and $6 \times CO_2$ CESM simulation of Zhu et al. (2019). While there are important aspects of the deep ocean-derived estimates that require further research (see above), if correct, this would also constrain "bulk" equilibrium climate sensitivity at the upper end of the range reported in Inglis et al. (2020), that is, ~5°C, in agreement with the LP-PETM-derived ECS estimate of Tierney et al. (2022). The offset of this study as well as that of Tierney et al. (2022) and Zhu et al. (2019) compared to Inglis et al. (2020) is likely driven in large part by the inclusion of a substantial amount of terrestrial temperature data which may be cool biased in several of the approaches included in that latter study (also compare our early Eocene GMST estimate to that of Ring et al. (2022)). Support for this is provided by the fact that our $T_d$-derived GMST is in excellent agreement with the DeepMIP database-derived estimate of coeval GMSST of this study and the $T_d$-GMSST analysis of Gaskell et al. (2022), see Figure 11, potentially implying an issue with some of the terrestrial but not the ocean data sets in that database. This is encouraging because both all model simulations considered here and our $T_d$-GMST analysis coupled with the $T_d$-GMSST analysis of Gaskell et al. (2022) suggest a convergence, within ~1–3°C, of GMST and GMSST under early Palaeogene-like climate states (Figures 4e and 4f; 11). As such, cooler EECO GMST estimates (Figure 11) are difficult to reconcile with our data-derived GMSST, and/or require this aspect of the climate model output to be inaccurate. We alternatively argue that the consistency between this study, Gaskell et al. (2022), and this emergent property of climate models provides strong support for the very warm EECO GMST presented here.

## 5. Conclusions

Here, we interrogate the use of a deep ocean temperature ($T_d$) record to infer global mean surface temperature (GMST) in detail, using both curated data compilations and two sets of climate model simulations. In particular, we address the question of whether $T_d$ and GMST are linearly related with a slope of 1, as previously hypothesized (Hansen et al., 2013). We show that no such relationship would be expected in a world characterized by polar amplification, because (a) the high latitude regions warm to a greater degree than the global mean, and (b) deep water is thought to have formed at high latitudes throughout most, if not all, of the Cenozoic. However, proxy data compilations of the Pliocene and early Palaeogene fall within uncertainty of a 1:1 $T_d$-GMST relationship, suggesting that (some) other process(es) act to balance polar amplification. Using both the DeepMIP set of simulations (Lunt et al., 2021) with varying $CO_2$ and fixed paleogeography and a set of Cenozoic HadCM3L simulations with covarying paleogeography and $CO_2$ (Valdes et al., 2021), we show that these processes are: (a) that SST in the regions and season of deep water formation is less sensitive to climate forcing than high latitude mean annual SST (evident in the DeepMIP ensemble but not clearly the case in the HadCM3L$_{V21}$ simulations), and (b) the fact that the land surface warming is more sensitive than the ocean surface (see also Goudsmit-Harzevoort et al. (2023)) to $CO_2$ and paleogeographic-driven climate change over the Cenozoic. While this provides a mechanistic basis for a 1:1 $T_d$-GMST relationship prior to the (mid)Pliocene, we note that some HadCM3L$_{V21}$ simulations do not adhere to this, with GMST underestimated by up to 3°C during the Miocene (Figure 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, which could be tested, for example, by using geochemical approaches capable of tracking the locus of deep water formation such as neodymium isotope data or carbon isotope gradients. More broadly, we stress that our key finding—that $T_d$-GMST is characterized 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 paleotemperatures 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 $\delta^{18}O$ (Figure 10a), 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 20$^{th}$ 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).

## Data Availability Statement

No new data sets 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/Valdesetal2021.html (Lunt et al., 2021; Valdes et al., 2021). They key data sets used in the data analysis performed here are the WOA2018 ocean temperature data set (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, $\Delta_{47}$, $\delta^{11}B$, 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, are given in Evans et al. (2024).

## References

Adkins, J. F., McIntyre, K., & Schrag, D. P. (2002). The salinity, temperature, and $\delta^{18}O$ of the glacial deep ocean. *Science*, *298*(5599), 1769–1773. https://doi.org/10.1126/science.1076252

Affek, H. P. (2012). Clumped isotope paleothermometry: Principles, applications, and challenges. *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., et al. (2020). Proxy evidence for state-dependence of climate sensitivity in the Eocene greenhouse. *Nature Communications*, *11*(1), 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., et al. (2016). Changing atmospheric $CO_2$ concentration was the primary driver of early Cenozoic climate. *Nature*, *533*(7603), 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 G. Fischer & G. Wefer (Eds.), *Use of proxies in paleoceanography: Examples from the South Atlantic* (pp. 489–512). Springer. 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*(1–2), 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., et al. (2006). Episodic fresh surface waters in the Eocene Arctic Ocean. *Nature*, *441*(7093), 606–609. https://doi.org/10.1038/nature04692

Burls, N. J., Bradshaw, C. D., De Boer, A. M., Herold, N., Huber, M., Pound, M., et al. (2021). Simulating Mocene warmth: Insights from an opportunistic multi-model ensemble (MioMIP1). *Paleoceanography and Paleoclimatology*, *36*(5), 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 of the United States of America*, *110*(35), 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*(2), 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 ($\delta^{18}O$ and Mg/Ca) with sea level history. *Journal of Geophysical Research*, *116*(C12), C12023. https://doi.org/10.1029/2011JC007255

Cramwinckel, M. J., Huber, M., Kocken, I. J., Agnini, C., Bijl, P. K., Bohaty, S. M., et al. (2018). Synchronous tropical and polar temperature evolution in the Eocene. *Nature*, *559*(7714), 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*(10), e2023PA004660. https://doi.org/10.1029/2023PA004660

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., et al. (2016). The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past*, *12*(7), 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., et al. (2013). Sea surface temperature of the mid-Piacenzian Ocean: A data-model comparison. *Scientific Reports*, *3*(1), 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*(3), e2020PA004174. https://doi.org/10.1029/2020PA004174

Evans, D., Badger, M. P. S., Foster, G. L., Henehan, M. J., Lear, C. H., & Zachos, J. C. (2018a). No substantial long-term bias in the Cenozoic benthic foraminifera oxygen-isotope record. *Nature Communications*, *9*(1), 2875. https://doi.org/10.1038/s41467-018-05303-4

Evans, D., Brugger, J., Inglis, G. N., & Valdes, P. (2024). Code and datasets to accompany Evans et al. (2024): The temperature of the deep ocean is a robust proxy for global mean surface temperature during the Cenozoic [Software]. *Paleoceanography and paleoclimatology*. Zenodo. https://doi.org/10.5281/zenodo.11059068

Evans, D., & Müller, W. (2012). Deep time foraminifera Mg/Ca paleothermometry: Nonlinear correction for secular change in seawater Mg/Ca. *Paleoceanography*, *27*(4), PA4205. https://doi.org/10.1029/2012PA002315

Evans, D., Sagoo, N., Renema, W., Cotton, L. J., Müller, W., Todd, J. A., et al. (2018b). Eocene greenhouse climate revealed by coupled clumped isotope-Mg/Ca thermometry. *PNAS*, *115*(6), 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*(4), 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., et al. (2019). Climate sensitivity on geological timescales controlled by nonlinear feedbacks and ocean circulation. *Geophysical Research Letters*, *46*(16), 9880–9889. https://doi.org/10.1029/2019GL083574

Ferreira, D., Cessi, P., Coxall, H. K., de Boer, A., Dijkstra, H. A., Drijfhout, S. S., et al. (2018). Atlantic-Pacific asymmetry in deep water formation. *Annual Review of Earth and Planetary Sciences*, *46*(1), 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. *Nature Geoscience*, *15*(8), 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*(1), 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 $CO_2$ on geological timescales. *PNAS*, *110*(4), 1209–1214. https://doi.org/10.1073/pnas.1216073110

Foster, G. L., Royer, D. L., & Lunt, D. J. (2017). Future climate forcing potentially without precedent in the last 420 million years. *Nature Communications*, *8*(1), 14845. https://doi.org/10.1038/ncomms14845

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*(11), e2111332119. https://doi.org/10.1073/pnas.2111332119

Gothmann, A. M., Stolarski, J., Adkins, J. F., Schoene, B., Dennis, K. J., Schrag, D. P., et al. (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., et al. (2023). The relationship between the global mean deep-sea and surface temperature during the early Eocene. *Paleoceanography and Paleoclimatology*, *38*(3), 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*(11), 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 $CO_2$ variability. *Paleoceanography*, *29*(9), 845–853. https://doi.org/10.1002/2014PA002653

Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., et al. (2021). Changing state of the climate system. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (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*(4), RG4004. https://doi.org/10.1029/2010RG000345

Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., et al. (2008). Target atmospheric $CO_2$: Where should humanity aim? *The Open Atmospheric Science Journal*, *2*(1), 217–231. 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 & Engineering Sciences*, *371*(2001), 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., et al. (2013). On the identification of a Pliocene time slice for data–model comparison. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *371*(2001), 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., et al. (2020). The Pliocene model intercomparison project phase 2: Large-scale climate features and climate sensitivity. *Climate of the Past*, *16*(6), 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*(8), e2021PA004375. https://doi.org/10.1029/2021PA004375

Hoffert, M. I., & Covey, C. (1992). Deriving global climate sensitivity from palaeoclimate reconstructions. *Nature*, *360*(6404), 573–576. https://doi.org/10.1038/360573a0

Hollis, C. J., Dunkley Jones, T., Anagnostou, E., Bijl, P. K., Cramwinckel, M. J., Cui, Y., et al. (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*(7), 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., et al. (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., et al. (2012). The geological record of ocean acidification. *Science*, *335*(6072), 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., et al. (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*(5), 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., et al. (2023). Meridional heat transport in the DeepMIP Eocene ensemble: Non-$CO_2$ and $CO_2$ effects. *Paleoceanography and Paleoclimatology*, *38*(8), e2022PA004607. https://doi.org/10.1029/2022PA004607

Kiehl, J. T., & Shields, C. A. (2013). Sensitivity of the Palaeocene–Eocene thermal maximum climate to cloud properties. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *371*(2001), 20130093. https://doi.org/10.1098/rsta.2013.0093

Lear, C. H., Coxall, H. K., Foster, G. L., Lunt, D. J., Mawbey, E. M., Rosenthal, Y., et al. (2015). Neogene ice volume and ocean temperatures: Insights from infaunal foraminiferal Mg/Ca paleothermometry. *Paleoceanography*, *30*(11), 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*(5451), 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*(5), 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 $\delta^{18}$O records. *Paleoceanography*, *20*(1), PA1003. https://doi.org/10.1029/2004PA001071

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., et al. (2018). World ocean atlas 2018, volume 1: Temperature.

Lunt, D. J., Bragg, F., Chan, W.-L., Hutchinson, D. K., Ladant, J.-B., Morozova, P., et al. (2021). DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data. *Climate of the Past*, *17*(1), 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., et al. (2016). Palaeogeographic controls on climate and proxy interpretation. *Climate of the Past*, *12*(5), 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., et al. (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*(2), 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., et al. (2015). Plio-Pleistocene climate sensitivity evaluated using high-resolution CO2 records. *Nature*, *518*(7537), 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., et al. (2020). Lessons from a high-$CO_2$ world: An ocean view from ∼3 million years ago. *Climate of the Past*, *16*(4), 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*(6), 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., et al. (2022). Cenozoic evolution of deep ocean temperature from clumped isotope thermometry. *Science*, *377*(6601), 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 $CO_2$, 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*(11), 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*, *117*(D8), D08101. 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 of the United States of America*, *117*(41), 25302–25309. https://doi.org/10.1073/pnas.2003914117

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*(7884), 239–244. https://doi.org/10.1038/s41586-021-03984-4

Pagani, M. (2002). The alkenone-$CO_2$ proxy and ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences*, *360*(1793), 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 Geoscience*, *3*(1), 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*(5421), 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. *Nature Communications*, *9*(1), 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*(2), 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*(5), 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*(1), 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 $\delta^{18}$O-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*(12), 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*(5), 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., et al. (2022). Comparison and synthesis of sea-level and deep-sea temperature variations over the past 40 million years. *Reviews of Geophysics*, *60*(4), 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*(5), 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*(26), 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*(17), 3633–3643. https://doi.org/10.1016/S0016-7037(97)00181-6

Sagoo, N., Valdes, P., Flecker, R., & Gregoire, L. J. (2013). The early Eocene equable climate problem: Can perturbations of climate model parameters identify possible solutions? *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *371*(2001), 20130123. https://doi.org/10.1098/rsta.2013.0123

Schmittner, A., Urban, N. M., Shakun, J. D., Mahowald, N. M., Clark, P. U., Bartlein, P. J., et al. (2011). Climate sensitivity estimated from temperature reconstructions of the Last Glacial Maximum. *Science*, *334*(6061), 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., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), 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*(6659), 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*(4), 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., et al. (2020). Past climates inform our future. *Science*, *370*(6517), eaay3701. https://doi.org/10.1126/science.aay3701

Tierney, J. E., Zhu, J., Li, M., Ridgwell, A., Hakim, G. J., Poulsen, C. J., et al. (2022). Spatial patterns of climate change across the Paleocene–Eocene thermal maximum. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(42), 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*(4), 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*(18), 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*(6), e2020JC017130. https://doi.org/10.1029/2020JC017130

Westerhold, T., Marwan, N., Drury, A. J., Liebrand, D., Agnini, C., Anagnostou, E., et al. (2020). An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science*, *369*(6509), 1383–1387. https://doi.org/10.1126/science.aba6853

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*(3), 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*(5517), 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*(7176), 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*(2), 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*(13–14), 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., et al. (2022). Early Eocene ocean meridional overturning circulation: The roles of atmospheric forcing and strait geometry. *Paleoceanography and Paleoclimatology*, *37*(3), e2021PA004329. https://doi.org/10.1029/2021PA004329

Zhu, J., Poulsen, C. J., & Tierney, J. E. (2019). Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Science Advances*, *5*(9), eaax1874. https://doi.org/10.1126/sciadv.aax1874

Ziveri, P., Thoms, S., Probert, I., Geisen, M., & Langer, G. (2012). A universal carbonate ion effect on stable oxygen isotope ratios in unicellular planktonic calcifying organisms. *Biogeosciences*, *9*(3), 1025–1032. https://doi.org/10.5194/bg-9-1025-2012