**S1. Statistical models for climate model output**

We obtain DMOTs and DGMSSTs from 93 coupled ocean-atmosphere climate model equilibrium simulations that were run with atmospheric CO2 concentrations ranging from 180 ppm to nine times PI (Fig. 2) (Alder and Hostetler, 2014; Galbraith and De Lavergne, 2019; Haywood et al., 2020; Clark et al., 2016; Goudsmit-Harzevoort et al., 2023; Rugenstein et al., 2019; He, 2011; Bereiter et al., 2018). Here we describe the relationship between ΔMOT and ΔGMSST in this synthesis of climate-model output and four regression models used to describe it. The four models were fit using R *stats* package functions *lm*(), *segmented*(), *loess*(), and *smoothing.spline*(), and their associated functions for prediction and diagnostic analyses (R, 2024). R-squared values are given by 1.0 minus the residual sum of squares divided by the total sum of squares. Adjusted R-squared values are calculated as 1.0 minus the fraction with the residual sum of squares divided by n minus p in the numerator and the total sum of squares divided by n minus 1.0 in the denominator, where p is equal to the number of parameters or equivalent degrees of freedom. The adjusted R-squared value, which is lower than the usual R-squared value, thereby includes a penalty for the number of adjustable parameters.

The relationship between ΔMOT and ΔGMSST is broadly linear, with a slope and intercept close to 1.0 and 0.0, respectively (Fig. 2A), a high *R2* value (0.924), highly significant *F*-test statistic (1104, *p* < 0.000) and a relatively low residual standard error (1.184). However, a residual diagnostic analysis of this model reveals a subtle but significant nonlinear lack of fit which can be seen in the residual scatter diagram (Fig. S1B). The nonlinearity in the residual scatter diagram is largely related to a quasi-linear cluster of points that plot below the main cloud of points (Fig. 2A), which include the points with the 10-largest negative residuals. The models that produced these data points are generally a mix of higher-than-present CO2 simulations and sensitivity tests, but not all such experiments produce that pattern of large negative residuals. This eliminates the possibility that there is an underlying “family-of-curves” explanation for the apparent nonlinearity. The residuals are also not normally distributed (Shapiro-Wilk test *W* = 0.932, *p* < 0.001), and in fact are visibly bimodal, imparted by the cluster of negative residuals. Large positive residuals also occur for larger values of ΔMOT and ΔGMSST, and those values also have the largest values of Cook’s Distance, a measure of influence on the fitted regression (Cook, 1977).

Inspection of the scatter diagrams suggests that an alternative to an overall straight-line fit might be a segmented linear fit (Muggeo, 2003), with the individual segments approximately defined by values of ΔGMSST less than 1.0 °C, between 1.0 and 5.0 or 6.0 °C, and greater than 6.0 °C. To test that notion, we used the R *segmented*() function, using 1 to 3 starting breakpoints that ranged between -2.0 to 10.0 °C. The best fit (in terms of goodness-of-fit and residual diagnostics) was a model with two breakpoints, at 1.187 (*s.e*. = 1.307) and 5.290 (*s.e.* = 0.855), and slopes for the lower, middle, and upper segments of 1.000 (*s.e.* = 0.178), 0.575 (*s.e.* = 0.185), and 1.244 (*s.e*. = 0.066), respectively (Fig. S1C). The model fit is slightly better than that for the linear one, with an adjusted *R2*value of 0.939 (vs. 0.924 for the linear one). Although the adjusted *R2*values do not differ much, the Akaike Information Criterion(*AIC*), a measure of the tradeoff between the goodness of fit and the number of adjustable parameters (smaller values are better), is lower here (*AIC* = 281.00) than for the original model (*AIC* = 299.31) despite the larger number of parameters. More importantly, the residual scatter plot shows a smaller indication of non-linearity (Fig. S1D) when compared to the residual scatter plot of the linear fit (Fig. S1B). The distribution of the residuals is still not normal, being a little heavy in the tails, but it is symmetric and shows no signs of bimodality. There are fewer points with extreme values of Cook’s distance.

The segmented regression implies that there are distinct thresholds (the breakpoints) where the relationship between ΔMOT and ΔGMSST changes. An alternative model that can accommodate smooth variations in slope can be fit via local- or nonparametric-regression approaches. We also considered polynomial models, but these were inferior overall. The third model we considered was fit using the R *loess*() function (Cleveland et al., 1992). The best overall local regression model had a smoothing parameter (*span*) of 0.8, and polynomial degree of 2. The fitted values (Fig. S1E) were broadly similar to the individual slopes in the segmented regression, but the slope varied smoothly instead of abruptly at ΔGMSST values around 1.0 and 6.0 °C. This model fits better than the two previous ones (*AIC* = 277.13; adjusted *R2* = 0.940), and like the segmented-regression model, the residual scatter diagram (Fig. S1F) shows little evidence of nonlinearity and has a similar distribution to the residuals of the segmented-regression model.

Inspection of the loess curve shows smaller-scale variations in slope in regions of the scatter diagram where data are sparse, and increasing the *span* value did not remove those. This is not unexpected given the nature of local regression. Consequently, we considered a fourth model, a smoothing spline (Hastie, 1992). Smoothing splines are often used as the smoothing function in fitting generalized additive models and in some ways can be seen as a less local, more global, approach than a loess smoother. Smoothing splines require the specification of a smoothing parameter (*spar*). Here we examined a range of *spar* values between 0.80 and 1.00, and selected a value of 0.95 as optimal, using the tradeoff between the goodness of fit and the tendency for low values of *spar* to allow the fitted values to “chase” individual data points. This model fit the data best of all (*AIC* = 272.18; adjusted *R2*= 0.944). The fitted values (Fig. S1G) are quite similar to those of the loess local regression model (Fig. S1E). The residual scatter diagram (Fig. S1H) shows little sign of nonlinearity, and again the distribution of the residuals is approximately normal. The residual scatter diagrams of the segmented-regression, loess, and smoothing-spline regressions are quite similar overall (Figs. S1D, S1F, and S1H).

Among the three alternatives to the linear model, the smoothing-spline model has several desirable properties—it fits the data best, with the lowest *AIC* and highest adjusted *R2* (although the differences are quite small), and is quite smooth while still displaying the changes in slope explicitly represented by the segmented regression model. In addition, its associated *predict*() function can be used to extract the local numeric first derivative of the fitted curve, which provides a representation of the HSE. The values of HSE are plotted as the red curve in Figs. 2D and S1G. Given these properties, the smoothing spline regression is our preferred choice for representing the relationship between ΔMOT and ΔGMSST in our synthesis of climate-model results.

**S2. Mg/Ca-based bottom water temperature reconstructions**

For ODP sites 806 and 926, we use the original Mg/Ca data from Lear et al. (2003) and the recent calibrations for *C. wuellerstorfi* (Mg/Ca=1.043Exp(0.118BWT)) and *O. umbonatus* (Mg/Ca= 1.317Exp(0.102BWT)) published by Barrientos et al. (2018). We note that temperature estimates derived from *O. umbonatus* using the new exponential calibration are consistent with those calculated using the calibration of Rathmann et al. (2004).

We use the original published BWT reconstruction from ODP site 607 measured on ostracodes by Dwyer and Chandler (2009) which is uncorrected for changes in seawater Mg/Ca. The original Mg/Ca-based BWT record measured on benthic foraminifera from site 607 (Sosdian and Rosenthal, 2009) includes measurements on *C*. *wuellerstorfi* and *O. umbonatus* using the published calibration Mg/Ca= 0.15\*T+1.16, which we update with the exponential calibration of Barrientos et al. (2018). Since the original record is arguably overprinted by the carbonate ion (∆CO32-) effect (Yu and Broecker, 2010), we have added Mg/Ca measurements of *Uvigerina sp.* (Sosdian and Rosenthal, 2010; Ford et al., 2016), an infaunal benthic foraminifera which has been shown to be insensitive to ∆CO32- effects (Elderfield et al., 2010; Elderfield et al., 2012). We calculate *Uvigerina* *sp.* BWTs using the Elderfield et al. (2012) calibration (Mg/Ca=0.1\*T+0.94). BWTs from site U1313 (Jakob et al., 2020) were recalculated using the Barrientos et al. (2018) calibration. The new compilation presented here, including new Mg/Ca data from *Uvigerina sp.,* demonstrates that the long-term trends are similar for all three species both in the raw Mg/Ca data and estimated BWTs, lending support to the published BWT records (Ford et al., 2016; Sosdian and Rosenthal, 2009). BWTs from site U1313 (Jakob et al., 2020) are based on *O. umbonatus* Mg/Ca and have been also recalculated using the calibration from Barrientos et al. (2018). The warmer Pliocene BOT at these sites are consistent with independent estimates, based on clumped isotope measurements of benthic foraminifera at the nearby IODP site U1308 (Braaten et al., 2023).

When reconstructing long-term temperature changes (>1 Ma) from Mg/Ca, it is necessary to account for variations in the seawater Mg/Ca concentration ratio. For the Mg/Ca data from the sites listed above, we follow the correction proposed by Evans and Müller (2012) relating the Mg/Ca of the calcareous shells to the Mg/Ca ratio of seawater through a power equation, where we use the seawater Mg/Ca record from Rosenthal et al. (2022) to correct measured foraminiferal Mg/Ca:

(Mg/Ca)corr = (Mg/Ca)foram\*(5.3/(Mg/Ca)sw)H

where (Mg/Ca)sw is the seawater ratio at the studied time, 5.3 is the modern seawater Mg/Ca ratio, and H is the species-specific power coefficient. We apply the (Mg/Ca)corr to the calibrations for benthic foraminifera and a power coefficient H=0.1 to correct for seawater changes in Mg/Ca, although choosing values anywhere between 0 and 0.4 would have a negligible effect on Pleistocene BWT. The correction for changes in seawater Mg/Ca for the benthic compilations is only about +1°C. We note that the choice of seawater Mg/Ca or the calibration equation has minimal effect on the temperature reconstructions for the past 5 Myr (Meinicke et al., 2021; Rosenthal et al., 2022).

The smoothed BWT reconstruction from Cramer et al. (2011) is based on a compilation of Mg/Ca records derived from six species or genera of benthic foraminifera. Because *O. umbonatus* is present throughout the study interval (0-60 Ma), Cramer et al. (2011) inferred Mg/Ca offsets between *O.* *umbonatus* and the other taxa and then applied two published Mg/Ca calibrations derived from core-top *O.* *umbonatus* to the entire data set (their equations 7a and 7b). It is not clear why the calibrations differ, but they lead to significant differences in the estimated amount and rate of cooling over the last 10 Myr. We use the equation 7b reconstruction in Cramer et al. (2011) based on Rathmann et al. (2004), since it is supported by the newer calibration of Barrientos et al. (2018).

**A collage of graphs and charts

AI-generated content may be incorrect.**

**Figure S1.** Scatter diagrams and fitted curves showing the relationship between ΔMOT and ΔGMSST from the climate-model output synthesis for four statistical models (left column, panels **A**, **C**, **E**, and **G**) and residual scatter diagrams for each model (right column, panels **B**, **D**, **F**, and **H**). Data shown here differ from Figure 2 in showing the temperature range extending to 20oC. Results shown are from PMIP3 model runs for LGM boundary conditions (Braconnot et al., 2012), model runs for CO2 <270 ppm with no ice sheets and with LGM ice sheets (Galbraith and De Lavergne, 2019), model runs for the last deglaciation (Alder and Hostetler, 2015), model runs for CO2 >420 ppm (Clark et al., 2016), LongRunMIP model runs for CO2 >2x PI (Rugenstein et al., 2019), model runs for CO2 ranging from 1x to 9x PI (Goudsmit-Harzevoort et al., 2023), PlioMIP2 model runs with CO2 = 400 ppm (Haywood et al., 2020), and model runs for CO2 >405 ppm (Galbraith and De Lavergne, 2019). In the left column of panels, the solid blue lines show the fitted values, the dashed blue lines are the 2s confidence intervals of the fitted values (illustrating the uncertainty in the curves), and the dashed gray lines are the 2s prediction intervals (illustrating the variability of the data about the fitted curves). In the right column of panels, the values plotted are the ordinary residuals and the fitted values from each model, while the blue lines are loess curves (with span = 0.90 and degree = 1) that summarize the pattern in the residual scatter diagrams. The red dots and lines in panel **C** show the estimated breakpoints in the segmented regression (dots) and their uncertainties. The red line in panel **G** shows the calculated first derivative of the fitted values, i.e. the heat storage efficiency (HSE = ΔMOT/ΔGMSST). R-squared values are given by 1.0 minus the residual sum of squares divided by the total sum of squares. Adjusted R-squared values are calculated as 1.0 minus the fraction with the residual sum of squares divided by n minus p in the numerator and the total sum of squares divided by n minus 1.0 in the denominator, where p is equal to the number of parameters or equivalent degrees of freedom. The adjusted R-squared value, which is lower than the usual R-squared value, thereby includes a penalty for the number of adjustable parameters.

**A graph of a wave

AI-generated content may be incorrect.**

**Figure S2.** **A**. DMOT reconstructions based on HSE = 0.7 and HSE = 0.3 (reproduces Figure 5B) fall within the 1s uncertainty on our DMOT reconstruction based on HSE = 0.5. **B.** Our high-resolution d18Osw (violet) reconstruction with 1s uncertainty compared to long-term (401-kyr running average) d18Osw for the three HSE scenarios shown in **A**. The differences during the early Pleistocene from our preferred 0.5-1 HSE scenario are small (<0.1 per mil) and the high-resolution d18Osw for the two bracketing HSE scenarios fall within the 1s uncertainty of high-resolution d18Osw based on 0.5-1 HSE scenario. Panels G and H address the timing question by comparing **C.** Sensitivity of DMOT in our preferred scenario (HSE increased from 0.5 to 1 between 1.5 Ma and 0.9 Ma) to one scenario where HSE increased between 1.2 Ma and 0.9 Ma and another where it increased from 1.7 Ma to 0.7 Ma. **D.** Sensitivity of our d18Osw reconstruction to different durations of the HSE increase from 0.5 to 1.

A diagram of a number of different age

AI-generated content may be incorrect.

**Figure S3.** Sensitivity of our d18Osw reconstruction (blue lines) to the same magnitude of our diagenetic correction (0.375‰) as used for that reconstruction but that tracks our DMOT reconstruction (red lines). The faster rate of change across the MPT slightly increases the trend in d18Osw over this interval (more depleted before 1 Ma, more enriched after 1 Ma), but the changes are insignificant (≤ 0.1‰).

**References**

Alder, J. and Hostetler, S. W.: Global climate simulations at 3000 year intervals for the last 21 000 years with the GENMOM coupled atmosphere-ocean model, Climate of the Past Discussions, 10, 2925-2978, 2014.

Alder, J. R. and Hostetler, S. W.: Global climate simulations at 3000-year intervals for the last 21 000 years with the GENMOM coupled atmosphere-ocean model, Climate of the Past, 11, 449-471, 10.5194/cp-11-449-2015, 2015.

Barrientos, N., Lear, C. H., Jakobsson, M., Stranne, C., O'Regan, M., Cronin, T. M., Gukov, A. Y., and Coxall, H. K.: Arctic Ocean benthic foraminifera Mg/Ca ratios and global Mg/Ca-temperature calibrations: New constraints at low temperatures, Geochimica et Cosmochimica Acta, 236, 240-259, 10.1016/j.gca.2018.02.036, 2018.

Bereiter, B., Shackleton, S., Baggenstos, D., Kawamura, K., and Severinghaus, J.: Mean global ocean temperatures during the last glacial transition, Nature, 553, 39-44, 10.1038/nature25152, 2018.

Braaten, A. H., Jakob, K. A., Ho, S. L., Friedrich, O., Galaasen, E. V., De Schepper, S., Wilson, P. A., and Meckler, A. N.: Limited exchange between the deep Pacific and Atlantic oceans during the warm mid-Pliocene and Marine Isotope Stage M2 “glaciation”, Clim. Past, 19, 2109-2125, 10.5194/cp-19-2109-2023, 2023.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nature Climate Change, 2, 417-424, 10.1038/nclimate1456, 2012.

Clark, P. U., Shakun, J. D., Marcott, S. A., Mix, A. C., Eby, M., Kulp, S., Levermann, A., Milne, G. A., Pfister, P. L., Santer, B. D., Schrag, D. P., Solomon, S., Stocker, T. F., Strauss, B. H., Weaver, A. J., Winkelmann, R., Archer, D., Bard, E., Goldner, A., Lambeck, K., Pierrehumbert, R. T., and Plattner, G. K.: Consequences of twenty-first-century policy for multi-millennial climate and sea-level change, Nature Climate Change, 6, 360-369, 10.1038/nclimate2923, 2016.

Cleveland, W. S., Grosse, E., and Shyu, W. M.: Local regression models, in: Statistical Models in S, edited by: Chambers, J. M., and Hastie, T. J., Chapman and Hall, <https://doi.org/10.1201/9780203738535>, 1992.

Cook, R. D.: Detection of influential observations in linear regression, Technometrics, 19, 10.2307/1268249, 1977.

Cramer, B. S., Miller, K. G., Barrett, P. J., and Wright, J. D.: Late Cretaceous-Neogene trends in deep ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal geochemistry (delta O-18 and Mg/Ca) with sea level history, Journal of Geophysical Research-Oceans, 116, 10.1029/2011jc007255, 2011.

Dwyer, G. S. and Chandler, M. A.: Mid-Pliocene sea level and continental ice volume based on coupled benthic Mg/Ca palaeotemperatures and oxygen isotopes, Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences, 367, 157-168, 10.1098/rsta.2008.0222, 2009.

Elderfield, H., Ferretti, P., Greaves, M., Crowhurst, S., McCave, I. N., Hodell, D., and Piotrowski, A. M.: Evolution of ocean temperature and ice volume through the mid-Pleistocene climate transition, Science, 337, 704-709, 10.1126/science.1221294, 2012.

Elderfield, H., Greaves, M., Barker, S., Hall, I. R., Tripati, A., Ferretti, P., Crowhurst, S., Booth, L., and Daunt, C.: A record of bottom water temperature and seawater delta(18)O for the Southern Ocean over the past 440 kyr based on Mg/Ca of benthic foraminiferal Uvigerina spp, Quaternary Science Reviews, 29, 160-169, 10.1016/j.quascirev.2009.07.013, 2010.

Evans, D. and Müller, W.: Deep time foraminifera Mg/Ca paleothermometry: Nonlinear correction for secular change in seawater Mg/Ca, Paleoceanography, 27, PA4205, 10.1029/2012pa002315, 2012.

Ford, H. L., Sosdian, S. M., Rosenthal, Y., and Raymo, M. E.: Gradual and abrupt changes during the Mid-Pleistocene Transition, Quaternary Science Reviews, 148, 222-233, <https://doi.org/10.1016/j.quascirev.2016.07.005>, 2016.

Galbraith, E. and de Lavergne, C.: Response of a comprehensive climate model to a broad range of external forcings: relevance for deep ocean ventilation and the development of late Cenozoic ice ages, Climate Dynamics, 52, 653-679, 10.1007/s00382-018-4157-8, 2019.

Goudsmit-Harzevoort, B., Lansu, A., Baatsen, M. L. J., von der Heydt, A. S., de Winter, N. J., Zhang, Y. R., Abe-Ouchi, A., de Boer, A., Chan, W. L., Donnadieu, Y., Hutchinson, D. K., Knorr, G., Ladant, J. B., Morozova, P., Niezgodzki, I., Steinig, S., Tripati, A., Zhang, Z. S., Zhu, J., and Ziegler, M.: The Relationship Between the Global Mean Deep-Sea and Surface Temperature During the Early Eocene, Paleoceanography and Paleoclimatology, 38, 10.1029/2022pa004532, 2023.

Hastie, T. J.: Generalized additive models, in: Statistical Models in S, edited by: Chambers, J. M., and Hastie, T. J., Chapman and Hall, 10.1201/9780203738535, 1992.

Haywood, A. M., Tindall, J. C., Dowsett, H. J., Dolan, A. M., Foley, K. M., Hunter, S. J., Hill, D. J., Chan, W. L., Abe-Ouchi, A., Stepanek, C., Lohmann, G., Chandan, D., Peltier, W. R., Tan, N., Contoux, C., Ramstein, G., Li, X. Y., Zhang, Z. S., Guo, C. C., Nisancioglu, K. H., Zhang, Q., Li, Q., Kamae, Y., Chandler, M. A., Sohl, L. E., Otto-Bliesner, B. L., Feng, R., Brady, E. C., von der Heydt, A. S., Baatsen, M. L. J., and Lunt, D. J.: The Pliocene Model Intercomparison Project Phase 2: large-scale climate features and climate sensitivity, Climate of the Past, 16, 2095-2123, 10.5194/cp-16-2095-2020, 2020.

He, F.: Simulating transient climate evolution of the last deglaciation with CCSM3, The University of Wisconsin - Madison, 2011.

Jakob, K. A., Wilson, P. A., Pross, J., Ezard, T. H. G., Fiebig, J., Repschlager, J., and Friedrich, O.: A new sea-level record for the Neogene/Quaternary boundary reveals transition to a more stable East Antarctic Ice Sheet, Proceedings of the National Academy of Sciences of the United States of America, 117, 30980-30987, 10.1073/pnas.2004209117, 2020.

Lear, C. H., Rosenthal, Y., and Wright, J. D.: The closing of a seaway: ocean water masses and global climate change, Earth and Planetary Science Letters, 210, 425-436, 10.1016/s0012-821x(03)00164-x, 2003.

Meinicke, N., Reimi, M. A., Ravelo, A. C., and Meckler, A. N.: Coupled Mg/Ca and clumped isotope measurements indicate lack of substantial mixed layer cooling in the western Pacific warm pool during the last 5 Million years, Paleoceanography and Paleoclimatology, 36, e2020PA004115, 10.1029/2020pa004115, 2021.

Muggeo, V. M. R.: Estimating regression models with unknown break-points, Statistics in Medicine, 22, 3055-3071, 2003.

R: A Language and Environment for Statistical Computing: <https://www.R-project.org/>, last

Rathmann, S., Hess, S., Kuhnert, H., and Mulitza, S.: Mg/Ca ratios of the benthic foraminifera <i>Oridorsalis umbonatus</i> obtained by laser ablation from core top sediments:: Relationship to bottom water temperature -: art. no. Q12013, Geochemistry Geophysics Geosystems, 5, 10.1029/2004gc000808, 2004.

Rosenthal, Y., Bova, S., and Zhou, X. L.: A user guide for choosing planktic foraminiferal Mg/Ca-temperature calibrations, Paleoceanography and Paleoclimatology, 37, e2022PA004413, 10.1029/2022pa004413, 2022.

Rugenstein, M., Bloch-Johnson, J., Abe-Ouchi, A., Andrews, T., Beyerle, U., Cao, L., Chadha, T., Danabasoglu, G., Dufresne, J. L., Duan, L., Foujols, M. A., Frolicher, T., Geoffroy, O., Gregory, J., Knutti, R., Li, C., Marzocchi, A., Mauritsen, T., Menary, M., Moyer, E., Nazarenko, L., Paynter, D., Saint-Martin, D., Schmidt, G. A., Yamamoto, A., and Yang, S. T.: LongRunMIP: Motivation and Design for a Large Collection of Millennial-Length AOGCM Simulations, Bulletin of the American Meteorological Society, 100, 2551-2570, 10.1175/bams-d-19-0068.1, 2019.

Sosdian, S. and Rosenthal, Y.: Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions, Science, 325, 306-310, 10.1126/science.1169938, 2009.

Sosdian, S. and Rosenthal, Y.: Response to Comment on "Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions", Science, 328, 10.1126/science.1186768, 2010.

Yu, J. M. and Broecker, W. S.: Comment on "Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions", Science, 328, 10.1126/science.1186544, 2010.