LETTER

doi:10.1038/nature19798

Evolution of global temperature over the past two million years

Carolyn W. Snyder[1]

Reconstructions of Earth's past climate strongly influence our understanding of the dynamics and sensitivity of the climate system. Yet global temperature has been reconstructed for only a few isolated windows of time[1,2], and continuous reconstructions across glacial cycles remain elusive. Here I present a spatially weighted proxy reconstruction of global temperature over the past 2 million years estimated from a multi-proxy database of over 20,000 sea surface temperature point reconstructions. Global temperature gradually cooled until roughly 1.2 million years ago and cooling then stalled until the present. The cooling trend probably stalled before the beginning of the mid-Pleistocene transition[3], and predated the increase in the maximum size of ice sheets around 0.9 million years ago[4–6]. Thus, global cooling may have been a precondition for, but probably is not the sole causal mechanism of, the shift to quasi-100,000-year glacial cycles at the mid-Pleistocene transition. Over the past 800,000 years, polar amplification (the amplification of temperature change at the poles relative to global temperature change) has been stable over time, and global temperature and atmospheric greenhouse gas concentrations have been closely coupled across glacial cycles. A comparison of the new temperature reconstruction with radiative forcing from greenhouse gases estimates an Earth system sensitivity of 9 degrees Celsius (range 7 to 13 degrees Celsius, 95 per cent credible interval) change in global average surface temperature per doubling of atmospheric carbon dioxide over millennium timescales. This result suggests that stabilization at today's greenhouse gas levels may already commit Earth to an eventual total warming of 5 degrees Celsius (range 3 to 7 degrees Celsius, 95 per cent credible interval) over the next few millennia as ice sheets, vegetation and atmospheric dust continue to respond to global warming.

Reconstructions of several key climate variables are available with high temporal resolution across past glacial cycles, such as polar temperature, atmospheric greenhouse gas (GHG) concentrations, sea surface temperature (SST), deep-water temperature (DWT) and sea level (see, for example, Extended Data Tables 1–3). Yet global average surface temperature (GAST) has been reconstructed for only a few isolated windows of time[1,2], and continuous reconstructions across glacial cycles remain elusive. The lack of continuous GAST reconstructions has constrained model–data comparisons to particular extreme points in time, such as the Last Glacial Maximum (LGM), but multiple time points are critical for characterizing the uncertainty in relationships estimated from palaeoclimate reconstructions[7,8]. The potential power of a continuous GAST record has been recently demonstrated[8]; GAST was reconstructed for the past 22,000 years (kyr) and used to clarify carbon dioxide's role in driving global climate change across glacial cycles. The present research creates a continuous record of GAST across a much longer timescale.

Previous continuous reconstructions of GAST across glacial cycles used only a single proxy record that was scaled linearly[2,4,9–12] or modelled[13] to estimate global values. This Letter presents a spatially weighted proxy reconstruction of GAST over the past 2 million years (Fig. 1a), estimated using a multi-proxy database compilation of over 20,000 SST point reconstructions from 59 ocean sediment cores (Extended Data Tables 1, 2). This research uses probabilistic simulations across multiple sources of uncertainty to estimate credible intervals at 1-kyr intervals, and validates the new reconstruction against previous estimates. The new GAST reconstruction can provide key insights into several major palaeoclimate questions, including the magnitude and stability of polar amplification, the state dependence of Earth system sensitivity (ESS, see below), and the role of global temperature in the mid-Pleistocene transition (MPT).

A comparison of GAST to Antarctic temperature reconstructions for the past 800 kyr finds that GAST and Antarctic temperature[14] are closely coupled across glacial cycles with a correlation of 0.72 (0.59–0.81, 95% credible interval, hereafter 'interval')—a high correlation given that the GAST reconstruction is estimated independently of the ice core records. There is a linear relationship of 0.61 °C (0.43–0.85 °C, 95% interval) change in GAST for every 1 °C change in Antarctic temperature (Fig. 2a) that does not significantly change over the past 800 kyr (Extended Data Fig. 6a). Some previous research on climate sensitivity over the past 800 kyr has assumed that changes in GAST are similar to half the magnitude of changes in Antarctic temperature[9,12,15]. On the basis of the new GAST reconstruction, there is an 87%

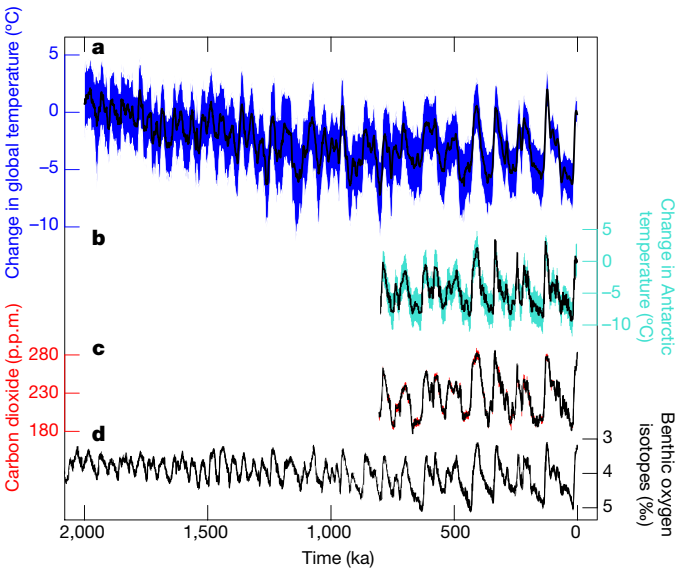


Figure 1 | Reconstruction of global average surface temperature (GAST) over the past 2 million years compared to other key palaeoclimate variables. a, GAST as temperature deviation (in °C) from present (average over 0–5 ka) in blue. **b**, Stacked reconstruction of change in Antarctic temperature[14] (°C) in cyan. **c**, Stacked reconstruction of atmospheric CO_2 concentrations[18] (p.p.m.) in red. **d**, Stack of deep-sea oxygen isotopes[30], $\delta^{18}O$ (‰), in grey. In all panels, the solid black lines show the median estimate and the colour shaded areas show the 95% interval.

[1]Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, California 94305, USA.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

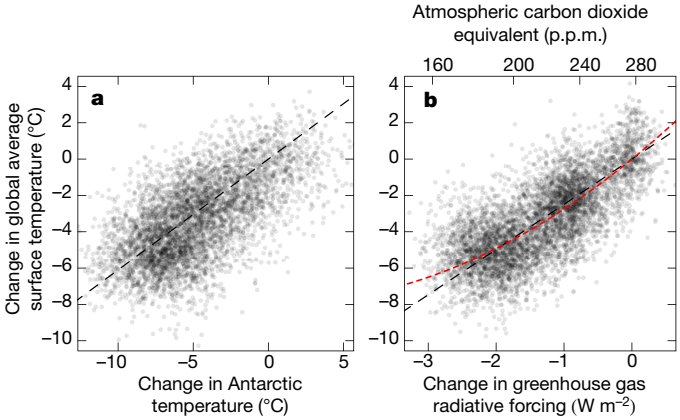


Figure 2 | Relationship of changes in GAST to changes in Antarctic temperature and GHG radiative forcing over the past 800 kyr. **a**, **b**, Each point represents randomly sampled estimates from simulations of GAST plotted against Antarctic temperature[14] (**a**) and GHG radiative forcing[14,17,18] (**b**) over the past 800 kyr. The dashed black line shows the median estimated relationship in °C per °C in **a** and in °C per W m^{-2} in **b**. The red dashed line shows the median estimated quadratic relationship in **b**.

probability that such an assumption underestimates global temperature and thus climate sensitivity.

Polar amplification can be estimated as change in Antarctic temperature for every 1 °C change in GAST, here estimated as 1.6 °C per °C (1.2–2.3 °C per °C, 95% interval). Estimates of polar amplification are complicated because the elevation of ice sheets changes during glacial cycles due to changes in accumulation and isostasy[9,14,16], and current ice sheet glaciological models disagree markedly with the ice topography used in LGM model simulations for Antarctica[9]. Climate models estimate that polar amplification of uncorrected Antarctic temperature will be nonlinear and lower in colder states, ~2 °C per °C for the LGM and ~1.2 °C per °C for future warming[9,16], but ~1.2 °C per °C for both when Antarctic temperature is elevation-corrected[16]. The present research uses a reconstruction of Antarctic temperature that is elevation-corrected[14], but corrections are highly dependent upon uncertain ice sheet assumptions[14,16]. A comparison of the new GAST reconstruction with Antarctic temperature finds a quadratic relationship to not be significant as predicted by the models (Fig. 2a), suggesting the elevation correction of Antarctic temperature is adequate. However, the magnitude of the polar amplification estimate from the new GAST reconstruction is significantly higher than predicted by many models (97.5% probability above 1.2)[9,16]. It is worth noting that those same models underestimate elevation-corrected Antarctic temperature change at the LGM by a factor of 1.2–2.3 (ref. 16).

Other research has assumed that changes in DWT can be used as a direct proxy for GAST[4,9,10] or doubled to estimate GAST[11]. A comparison between GAST and 12 DWT reconstructions from three different methods[4–6] finds highly variable results, with median correlations varying between 0.3 and 0.8 and median linear relationships varying between 1.4 °C and 3.5 °C change in GAST per 1 °C change in DWT (Extended Data Table 3). The observed attenuation of the global temperature signal and the reduced correlation may be caused by deep-water cooling being limited by water's freezing temperature and/or by changes in ocean circulation[4,6,17]. These results demonstrate the high uncertainty in inferring GAST from any single DWT reconstruction.

High-resolution estimates of atmospheric GHG concentrations are also available from Antarctic ice core records[14,18] over the past 800 kyr. The GAST reconstruction reveals a remarkably stable relationship between GAST and GHG radiative forcing[14,17,18] over the past 800 kyr with a correlation of 0.82 (0.66–0.92, 95% interval), stronger than the correlation between GAST and Antarctic temperature (Fig. 2b).

The concept of $S_{[GHG]}$ has been defined[2] as the total global climate response over millennial timescales from changes in ice sheets, dust and vegetation, as well as from the feedbacks included in 'equilibrium climate sensitivity'—water vapour, lapse rate, sea ice, snow cover, clouds and ocean heat uptake, but does not include carbon cycle feedbacks[1,2]. A comparison of the new GAST reconstruction with GHG radiative forcing estimates $S_{[GHG]}$ as 2.5 °C (1.8–3.6 °C, 95% interval) change in GAST per 1 W m^{-2} change in GHG radiative forcing (Fig. 2b), and finds that the relationship does not change significantly over the past 800 kyr (Extended Data Fig. 6b). This $S_{[GHG]}$ estimate translates to a 9 °C (7–13 °C, 95% interval) change in GAST per doubling of atmospheric carbon dioxide (3.7 W m^{-2}), which has often been called ESS[2,19,20].

Attenuation of the $S_{[GHG]}$ relationship is apparent in deep glacial states and a quadratic relationship is found to be a significantly better fit than a linear relationship (Fig. 2b). However, it is unclear whether such a quadratic relationship would apply in warmer states—when the bottom quarter of the record is removed, a quadratic relationship is not significant. Previous research also found $S_{[GHG]}$ to be climate state dependent, and most studies find higher values for the late Quaternary than for the Pliocene[2]. The presence of large ice sheets in the Quaternary is probably a major cause. Yet little research has focused on the potential variation of $S_{[GHG]}$ within the late Quaternary. Masson-Delmotte *et al.*[9] also found attenuation within the late Quaternary in deep glacial states, estimating a parabolic relationship. The observed attenuation of late Quaternary $S_{[GHG]}$ seems to suggest there is a limit to the power of positive climate feedbacks, such as from sea ice and ice sheets, as ice sheet size increases in deep glacial states.

Because $S_{[GHG]}$ and ESS are climate state dependent, it is most useful to compare this result to other estimates from the late Quaternary[2]. Rohling *et al.*[2] found a similar ESS estimate of 8.5 °C, but did not include a comparable probabilistic analysis in their estimate. Hansen *et al.*[11,15] both estimated ESS of 6 °C, assuming GAST is half the Antarctic temperature change or twice the DWT change, respectively. The present research finds that there is a 99% probability that ESS for the late Quaternary is higher than 6 °C.

The new GAST reconstruction also can provide insight into the MPT. The causes of the MPT, when the Earth's climate shifted from glacial cycles with periods of about 41 kyr to those with quasi-100-kyr periods, are not well understood and debates continue on the potential linkage between different orbital changes and the quasi-100-kyr cycles[21–24]. Some theories explain the MPT with changes in nonlinear feedbacks internal to the climate system, such as changes to ice sheets, sea ice or ocean circulation[3,21,22,25–28]. An alternative theory is the erosion of continental regolith underneath the ice sheets enabling the growth of thicker ice sheets[3].

Probabilistic breakpoint analysis is used to identify any changes in cooling trends across the past 2 Myr in the new GAST reconstruction, as well as the timing of the cooling trend changes. Such analysis find a strong cooling trend after 2 Myr ago (Ma) that then stops most probably at 1.2 Ma (median estimate), with a 72% probability that GAST cooling stopped by 1.1 Ma and 77% by 0.9 Ma (Fig. 3a). The timing of when the global cooling trend stops roughly corresponds to estimates of the beginning of the broad MPT, which is estimated to occur over the general period of 1.25 to 0.7 Ma based on spectral analysis of oxygen isotopes[3]. Before roughly 1.2 Ma, global temperature cooled gradually by approximately 0.34 °C (0.16–0.62 °C, 95% interval) per 100 kyr (Fig. 3b). However, since 1.2 Ma, GAST stabilized with no significant change in global temperature, 0.007 °C (−0.12 to 0.42 °C, 95% interval) per 100 kyr (Fig. 3c). From 1.2 to 0.5 Ma, the behaviour of GAST across glacial–interglacial cycles gradually shifted to quasi-100-kyr cycles with larger amplitudes of change, as seen in Fig. 1a. Although average GAST did not continue to cool after roughly 1.2 Ma, GAST did show a particularly large amplitude for the glacial cycle at 0.9 Ma of 7 °C (4–10 °C, 95% interval), which is similar in magnitude to more recent, post-MPT glacial cycles. These findings of gradual cooling probably pre-dating the MPT and a gradual shift to quasi-100-kyr

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

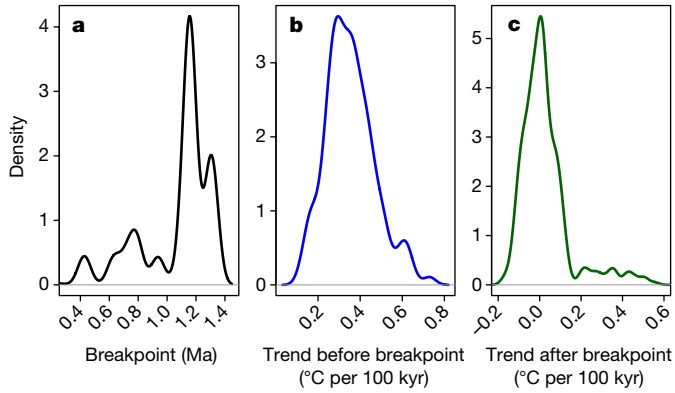


Figure 3 | Probabilistic breakpoint analysis of global temperature trends over the past 2 million years. Shown are empirically fitted frequency distributions for the timing of when the trend in global temperature changes (**a**; in black), the global temperature trend before the breakpoint (**b**; in blue), and the global temperature trend after the breakpoint (**c**; in green).

cycles are consistent with some previous SST and DWT research[6,25,29]. The global cooling trend also is synchronous with the development of the equatorial Pacific cold tongue and bipolar cooling estimated from ~1.8 Ma to ~1.2 Ma (ref. 29). However, GAST does not exhibit the intensified cooling across the MPT seen in some individual SST records and predicted by some MPT theories[25].

Several MPT theories employ increases in ice sheet size to explain the change in nonlinear climate feedbacks at the MPT, and hypothesize that global cooling could be the causal mechanism for such ice sheet growth[21,22,26,27]. Analyses of orbital responses across the MPT similarly suggested that global cooling could have enabled the skipping of obliquity cycles[22,23]. The present research provides evidence of such global cooling before the MPT. However, the global cooling probably pre-dates the rapid ice sheet growth observed at the deep glacial period around 0.9 Ma (refs 4–6) and the development of the first quasi-100-kyr cycle by 300 kyr ago (ka). Thus, either additional explanation is required to explain the lag after global cooling before MPT climate changes or the MPT changes may have been caused by a mechanism not linked to global temperature, such as erosion of continental regolith[3] or orbital changes without internal climate changes[24].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 January; accepted 24 August 2016.
Published online 26 September 2016.

1. Masson-Delmotte, V. *et al.* in *Climate Change 2013: The Physical Science Basis* Ch. 5 (eds Stocker, T. F. *et al.*) 383–464 (Cambridge Univ. Press, 2013).
2. Rohling, E. J. *et al.* Making sense of palaeoclimate sensitivity. *Nature* **491,** 683–691 (2012).
3. Clark, P. U. *et al.* The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO_2. *Quat. Sci. Rev.* **25,** 3150–3184 (2006).
4. Elderfield, H. *et al.* Evolution of ocean temperature and ice volume through the Mid-Pleistocene climate transition. *Science* **337,** 704–709 (2012).
5. Rohling, E. J. *et al.* Sea-level and deep-sea-temperature variability over the past 5.3 million years. *Nature* **508,** 477–482 (2014).
6. Bates, S. L., Siddall, M. & Waelbroeck, C. Hydrographic variations in deep ocean temperature over the mid-Pleistocene transition. *Quat. Sci. Rev.* **88,** 147–158 (2014).
7. Rohling, E., Medina-Elizalde, M., Shepherd, J., Siddall, M. & Stanford, J. Sea surface and high-latitude temperature sensitivity to radiative forcing of climate over several glacial cycles. *J. Clim.* **25,** 1635–1656 (2012).
8. Shakun, J. D. *et al.* Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484,** 49–54 (2012).
9. Masson-Delmotte, V. *et al.* EPICA Dome C record of glacial and interglacial intensities. *Quat. Sci. Rev.* **29,** 113–128 (2010).
10. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292,** 686–693 (2001).
11. Hansen, J., Sato, M., Russell, G. & Kharecha, P. Climate sensitivity, sea level and atmospheric carbon dioxide. *Phil. Trans. R. Soc. Lond. A* **371,** (2013).
12. Chylek, P. & Lohmann, U. Aerosol radiative forcing and climate sensitivity deduced from the last glacial maximum to Holocene transition. *Geophys. Res. Lett.* **35,** L04804 (2008).
13. van de Wal, R. S. W., de Boer, B., Lourens, L. J., Kohler, P. & Bintanja, R. Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7,** 1459–1469 (2011).
14. Parrenin, F. *et al.* Synchronous change of atmospheric CO_2 and Antarctic temperature during the last deglacial warming. *Science* **339,** 1060–1063 (2013).
15. Hansen, J. *et al.* Target atmospheric CO_2: where should humanity aim? *The Open Atmos. Sci. J.* **2,** 217–231 (2008).
16. Masson-Delmotte, V. *et al.* Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dyn.* **26,** 513–529 (2006).
17. Köhler, P. *et al.* What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints on climate sensitivity. *Quat. Sci. Rev.* **29,** 129–145 (2010).
18. Bereiter, B. *et al.* Revision of the EPICA Dome C CO_2 record from 800 to 600kyr before present. *Geophys. Res. Lett.* **42,** 542–549 (2015).
19. Pagani, M., Liu, Z., LaRiviere, J. & Ravelo, A. C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nat. Geosci.* **3,** 27–30 (2010).
20. Lunt, D. J. Earth system sensitivity inferred from Pliocene modelling and data. *Nat. Geosci.* **3,** 60–64 (2010).
21. Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12,** 577–585 (1997).
22. Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313,** 508–511 (2006).
23. Huybers, P. & Wunsch, C. A depth-derived Pleistocene age model: uncertainty estimates, sedimentation variability, and nonlinear climate change. *Paleoceanography* **19,** PA1028 (2004).
24. Imbrie, J. Z., Imbrie-Moore, A. & Lisiecki, L. E. A phase-space model for Pleistocene ice volume. *Earth Planet. Sci. Lett.* **307,** 94–102 (2011).
25. McClymont, E. L., Sosdian, S. M., Rosell-Melé, A. & Rosenthal, Y. Pleistocene sea-surface temperature evolution: early cooling, delayed glacial intensification, and implications for the mid-Pleistocene climate transition. *Earth Sci. Rev.* **123,** 173–193 (2013).
26. Tziperman, E. & Gildor, H. On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. *Paleoceanography* **18,** PA000627 (2003).
27. Abe-Ouchi, A. *et al.* Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume. *Nature* **500,** 190–193 (2013).
28. Ganopolski, A. & Calov, R. The role of orbital forcing, carbon dioxide and regolith in 100 kyr glacial cycles. *Clim. Past* **7,** 1415–1425 (2011).
29. Martínez-Garcia, A., Rosell-Melé, A., McClymont, E. L., Gersonde, R. & Haug, G. H. Subpolar link to the emergence of the modern Equatorial Pacific Cold Tongue. *Science* **328,** 1550–1553 (2010).
30. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography* **20,** PA1003 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements I thank S. Schneider, C. Field, C. Tebaldi, R. Dunbar, K. Caldeira, C. Warshaw, H. Elderfield and R. Samworth for advice and feedback. I am indebted to many scientists for supplying their proxy data records (see Extended Data Tables 1–3), and to NOAA's National Centers for Environmental Information and PANGAEA. This study was supported by a National Science Foundation Graduate Research Fellowship. The views expressed in this article are those of the author and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.W.S. (carolyn.snyder@gmail.com).

Reviewer Information *Nature* thanks E. J. Rohling and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

METHODS

GAST reconstruction. *Overall approach.* This research estimates GAST from local SST proxy-based reconstructions through five steps. First, I collect SST proxy-based reconstructions and estimate proxy uncertainty from a literature review. Second, I interpolate the SST reconstructions to common 1-kyr intervals and estimate dating uncertainty. Third, I estimate average SST values for latitudinal zones using a variety of possible spatial weighting schemes. Fourth, I analyse Paleoclimate Modelling Intercomparison Project (PMIP) model simulations to obtain an estimate of the relationship between average SST over latitudinal zones to GAST. I use the estimated scalar to linearly scale the average changes in SST for latitudinal zones to changes in GAST. Last, I use several approaches to test the validity and sensitivity of this approach.

The biggest challenge in this research is that the primary continuous temperature reconstructions available over the past 2 Myr are mostly from SST proxy records. Available terrestrial temperature reconstructions are too infrequent and limited in spatial distribution to be used for a global reconstruction at this time. Thus, this research develops a method to estimate GAST from a collection of SST estimates. Such a scaling must address the fact that SST records do not adequately cover the entire Earth surface, and that by definition, SST records do not include records of temperature change over land. This is especially important given that temperature change is amplified over land and at the poles relative to the oceans. The goal of this research is to develop a method that is transparent in its assumptions and associated estimates of uncertainty and could be applied in the same way across the past 2 Myr, even though there are far fewer SST reconstructions available farther back in time. Because GAST estimates are often used to investigate questions related to climate sensitivity, the designed approach should not be dependent on any assumptions of climate sensitivity.

To investigate potential approaches to this challenge, this research uses the PMIP global climate models[31,32] because the PMIP model experiments provide estimates of local SST and air surface temperature as well as GAST for both the LGM and the pre-industrial state, thus covering most of the range of temperatures over the past 2 Myr. Because the highest latitude SST reconstructions are at about 60° N/S, I compared the ratio of change in GAST to the average change in SST in all SST grid cells 60° N–60° S. The estimated scaling factors are found to not be correlated with other model features, such as climate sensitivity and LGM GAST (Extended Data Fig. 4). Hargreaves *et al.*[33] similarly find that LGM GAST is not correlated with the model's climate sensitivity values. This is very important, since a main goal of this approach is to develop a method that is not linked to a particular climate model or a particular estimate of climate sensitivity. Another approach would be to scale just from average tropical SST changes, but that would ignore valuable data from high latitude records, would be more uncertain due to the relative larger contribution of the scalar value, and would be more model dependent.

It is important to note that the concept of a regional average SST over latitudes 60° N–60° S does not have direct physical meaning or relevance as it is not a regional average since it does not include the land area in latitudes 60° N–60° S and only averages over SST grid cells. Nor does the described scalar of change in GAST relative to change in SST over 60° N–60° S. However, taking a simple average over latitudinal bands is a stable and transparent way to summarize the available SST reconstructions. Moreover, there is no assumed functional relationship of SST with latitude or with GAST, except for the single coarse scaling metric. This method ensures that the full change in GAST at glacial maxima is captured in the final estimated reconstruction by using the LGM scalar from the PMIP models.

Rohling *et al.*[7] used a somewhat similar approach of a spatial average of SST to estimate global climate sensitivity. They estimated a quadratic relationship of change in SST over latitude from 36 SST reconstructions and integrated that function to calculate a global mean response of SST. They then adjusted for a stronger terrestrial response to scale to a larger estimate of global climate sensitivity, but they did not produce an estimate of GAST.

SST proxy-based reconstructions. This analysis utilizes a multi-proxy database compilation of all available and reliable SST reconstructions that cover at least the past 100 kyr. The SST database includes 61 SST proxy reconstructions from 59 ocean sediment cores: 29 using alkenone unsaturation indices ($U^{K'}_{37}$), 17 using ratios of Mg/Ca in planktonic foraminifera, and 16 based upon microfossil abundances (using transfer functions for planktonic foraminifera and radiolarians) (Extended Data Tables 1, 2; Extended Data Figs 1, 2; Supplementary Data). A multi-proxy approach enables a reduction of the uncertainties and potential biases specific to each proxy method by combining estimates from several independent proxies[34].

Proxy methods have a variety of potential sources of error, including proxy measurement, seasonality, species dependence, productivity, water column depth, mixing, and dissolution and other post-depositional alteration. Estimates of the measurement and calibration errors are available from laboratory and field experiments for the different proxy methods (for example, from alkenone indices[35,36], Mg/Ca ratios[37,38] and species assemblages[39]) and typically range from 1 °C to 3 °C for two standard deviations. However, the published uncertainty estimates often do not include considerations of structural uncertainty from the assumptions of the proxy method. Therefore, I use the upper range of the published values of 3 °C (95% interval) as an estimate for the combined uncertainty for each of the SST proxy methods.

Dating uncertainty and SST reconstruction interpolation. It is imperative that comparisons between palaeoclimate records include the uncertainty in matching which parts of each record occurred at the same point in time[40,41]. To interpolate each SST reconstruction to a common 1-kyr timescale, I estimated a weighted average of the SST reconstruction for each time point on the 1-kyr timescale. The weights are based on the distance in time between the reconstruction value and the time point of interest. The bandwidth is set by the dating uncertainty for that time point. I use the published age scales for the SST reconstructions. I use the estimate of 10 kyr (95% interval) for dating uncertainty from orbital tuning[23,41–43], unless papers provide specific estimates of uncertainty in their timescales. The uncertainty in the estimated interpolated value is estimated from a weighted average of the squared differences between the reconstruction values and the estimated interpolated value. I implement this method using a Nadaraya–Watson kernel-weighted local constant regression, using the function *ksmooth* in the R statistical program (http://stat.ethz.ch/R-manual/R-patched/library/stats/html/ksmooth.html). This method results in larger SST uncertainty during periods of rapid change than during periods of stable SST, thus reflecting the varying potential impact of dating uncertainty.

SST averages over a latitudinal zone. The spatial distribution of the available SST records is too sparse to use spatial statistics to estimate average SST. An alternative approach is to apply a simple assumption of a quadratic change in SST with latitude[7], but such a method adds a large amount of uncertainty when there are fewer available reconstructions. The simplest approach would be a direct average of all SST reconstructions equally, but that ignores the known general amplification of change in SST with latitude and would make the estimate very dependent on the particular distribution of the available reconstructions. The proposed method is a middle ground: SST is first averaged across records within a single latitudinal zone and then the latitudinal zones are summed using applicable spatial weights. Because it is spatially averaged, it is not purely driven by high latitude records, and the reduction in uncertainty from multiple records in a given latitudinal zone is captured.

To explore how the 60° N–60° S average SST estimate varies with different latitudinal zone boundaries, I use 9 different possible configurations of latitudinal zone boundaries used with equal weights in the final ensemble:

- Four zones:

60° N–30° N, 30° N–0°, 0°–30° S, 30° S–60° S (equal degrees);
60° N-25.7° N, 25.7° N–0°, 0°–25.7° S, 25.7° S–60° S (equal areas);
60° N–20° N, 20° N–0°, 0°–20° S, 20° S–60° S;
60° N–35° N, 35° N–0°, 0°–35° S, 35° S–60° S.

- Six zones:

60° N–40° N, 40° N–20° N, 20° N–0°, 0°–20° S, 20° S–40° S, 40° S–60° S (equal degrees);
60° N–35.3° N, 35.3° N–16.8° N, 16.8° N–0°, 0°–16.8° S, 16.8° S–35.3° S, 35.3° S–60° S (equal areas);
60° N–40° N, 40° N–15° N, 15° N–0°, 0°–15° S, 15° S–40° S, 40° S–60° S;
60° N–35° N, 35° N–20° N, 20° N–0°, 0°–20° S, 20° S–35° S, 35° S–60° S;
60° N–30° N, 30° N–15° N, 15° N–0°, 0°–15° S, 15° S–30° S, 30° S–60° S.

The SST proxy records have limited geographical distribution, variable length and variable resolution. In particular, the records are clustered in space and have a non-random spatial distribution (Extended Data Figs 1a, 2). Thus, I explore two approaches in analysing the records. In the first, I include all 61 SST proxy records and weight them equally. In the second approach, I identify locations where there is more than one record within a circle of radius 5° latitude/longitude. I identify 11 such clusters that include 43 proxy records in total. For each cluster, I estimate the mean value over time and the variation across the cores. I plot the resulting 29 records (11 clusters and 18 independent proxy records) for this 'clustered' analysis in Extended Data Fig. 1b.

Because this research is focused on change in SST not absolute values of SST, it is important that the SST reconstructions are normalized to change from present before they are summed within a latitudinal zone. The rest of this research defines present as the mean value 0–5 ka. However, in this particular instance, 8 of the SST reconstructions have their first estimate between 5 and 8 ka. Rather than assume an estimate for 5 ka, this research uses the mean value 0–10 ka when estimating the latitudinal zone averages. Once the weighted average of the zones is estimated, the deviation from present is then recalculated to be the mean value 0–5 ka to ensure consistency.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

This analysis uses Monte Carlo-style simulations to estimate several sources of uncertainty from the proxy reconstruction by adding random error to each reconstruction from the estimated proxy uncertainty and the estimated uncertainty introduced from absolute dating (see previous discussion for those estimates) in each simulation. I explore structural uncertainty in the averaging method by randomly resampling the different latitudinal zonal boundaries. I also randomly resample the proxy cores to explore the uncertainty introduced from a particular set of proxy records. The three general sources of uncertainty—from the proxy reconstructions, from the latitudinal zone boundaries, and from the sample of records—are all major contributors to the final uncertainty distribution. Random resampling of the SST proxy records is the largest contributor to the uncertainty. In total, I calculate 4,000 simulations for time series of average SST over 60° N–60° S. I repeat this approach with the 29 records from the clustering procedure described above, and find similar results for the final GAST estimate (Extended Data Fig. 3a, b).

Scaling from latitudinal SST averages to GAST. As discussed previously, this analysis uses the PMIP global climate models[31,32] to scale regional SST averages to GAST. This is necessary because there are insufficient land temperature proxies available over the past 2 Myr. I use model simulations of the LGM because they most closely compare to the large changes seen in the reconstructions. For a specific LGM climate model simulation, I estimate the change in temperature between the LGM and pre-industrial runs for both SST and surface air temperature. I then estimate the ratio of change in GAST to change in 60° N–60° S average SST to be used as a scaling factor (Extended Data Fig. 4). The analyses are performed for all available model simulations using the PMIP2 database from 30 July 2009 (https://pmip2.lsce.ipsl.fr/database/access/request.shtml;) and the PMIP3 database from CMIP5 archives at PCMDI from 10 December 2015 (http://cmip-pcmdi.llnl.gov/cmip5/data_portal.html). Because of the uncertainty in applying these scalars to estimate GAST, I more than double the standard deviation of the estimates from the sample of nine models (0.14) and use a scaling factor of 1.9 (1.5–2.3, 95% interval), assuming a normal distribution that includes 8 of the 9 models in the middle 67%. The uncertainty in the scaling factor causes approximately a doubling of the standard deviation of the final GAST reconstruction at each point in time.

This research uses 60° N–60° S because that is the maximum extent of the SST reconstructions. However, the high latitude Southern Hemisphere reconstructions are fairly short, and for most of the past 2 Myr, the highest latitude SST reconstruction is at 43° S in contrast to 58° N in the Northern Hemisphere. Sensitivity analysis finds that repeating the analysis with the PMIP models using 50° N–50° S finds a very similar range of scalar factors (the mean value changes by only 1%). Therefore, this research continues to use the scaling factor estimate of 1.9 (1.5–2.3, 95% interval) as equally applying to a range of 60° N–60° S or 50° N–50° S.

To estimate GAST reconstructions, I use Monte Carlo-style simulations to propagate all the previously mentioned sources of uncertainty. I sample from the simulations of 60° N–60° S average SST described previously and apply a scaling factor randomly chosen from the uncertainty distribution to calculate 5,000 simulations for the potential time series of GAST (Fig. 1a). The final GAST simulation ensemble of potential time series includes propagation from simulating the potential errors in each proxy reconstruction, from resampling the proxy records, and from randomly varying the spatial weighting methods and the scaling factors of regional average SST to GAST. When simulating GAST time series, the time series stops when any latitudinal zone no longer has any reconstructions available. Therefore, the lengths of the individual time series within the final simulation GAST ensemble of potential time series (Extended Data Fig. 2c) are reflective of the availability of SST proxy reconstructions (Extended Data Fig. 2b).

The assumption of linear and constant scaling over the past 2 Myr is a weakness in this approach, but it is necessary because of the limited available data[7,15]. One way to explore how that scalar value could potentially change over time is by investigating the PMIP simulations of the mid-Pliocene warm period (mPWP, 3.0–3.3 Ma)[44]. An analysis of cross-model means finds a scalar of 1.6, which is 15% lower than the median estimate from the LGM PMIP experiments and is within the proposed confidence interval. Moreover, the mPWP was much warmer (1.8–3.6 °C warmer than present) with much less ice and sea levels higher by 22 ± 10 m (ref. 44). At the Pleistocene transition at ~2.7 Ma, there was a substantial increase in Northern Hemisphere ice sheets, and by 2.4 Ma, the climate transitioned to ~41-kyr glacial cycles until the MPT[45]. Therefore, the Earth's climate over the past 2 Myr was much more similar to the LGM than to the mPWP. Thus, if the scalar is state dependent, as the model simulations suggest, the value at 2 Ma is likely to be closer to 1.9 than 1.6 and thus well within the proposed interval.

To test the impact of the assumption of a constant linear scalar of the ratio of change in GAST to change in average SST over the latitudinal zones 60° N to 60° S, I analyse two alternative methods. Rather than use the single value of 1.9 estimated from the LGM, I use a moving scalar that is defined by the time points of 1.9 at the LGM and 1.6 at the mPWP. Because the scalar is thought to potentially vary with climate state, I use two different time series to construct the scalar time series: reconstructions of deep sea oxygen isotopes[30] and of relative sea level[5]. I linearly scale each of these time series such that their mean value at the LGM (mean over 19–23 ka) is 1.9 and their mean value at the mPWP (mean over 3–3.3 Ma) is 1.6. The resulting estimated median GAST time series are very similar to the GAST reconstruction estimated from a constant scalar: 0.998 correlation for the deep sea oxygen isotopes method and 0.998 for the relative sea level method. Investigations of polar amplification are also similar: 0.59 (0.45 to 0.79, 95% interval) for the deep sea oxygen isotopes method and 0.59 (0.45 to 0.74, 95% interval) for the relative sea level method, as compared to 0.61 (0.43 to 0.85, 95% interval) for the primary GAST reconstruction. The regressions also find that a quadratic relationship is not significant. The uncertainty estimates of the two alternative approaches are underestimates because they do not include uncertainty in the base reconstructions themselves nor do they include any uncertainty in the LGM and mPWP estimates used to scale the base reconstructions.

Validity testing the GAST reconstruction using particular points in time. To test the validity of the GAST reconstruction, the new record can be compared to previous published reconstructions for points of interest, such as the LGM or the Last Interglacial. The new GAST reconstruction finds global cooling at the LGM (~21 ka) of 6.2 °C (4.5–8.1 °C, 95% interval) from the present value. This estimate is similar to the recent IPCC synthesis of the 'very likely' range (>95% probability) of 3–8 °C (ref. 1). The present analysis finds a higher most-likely value consistent with other recent proxy-based reconstructions of the LGM[17,46,47]. Lower model-based estimates of LGM cooling could be influenced by underestimates of the changes in LGM radiative forcing, such as from changes in dust and vegetation[1,48], or underestimates of climate sensitivity at the LGM. The GAST estimate for maximum warming during the Last Interglacial (~125 ka) is 2.0 °C (0.4–3.6 °C, 95% interval) warmer than present. This result is consistent with recent proxy estimates: sea level rise is likely to have exceeded 8 m above present levels[49], a large warming over Antarctica during the Last Interglacial[50], and a GAST estimate of 1.9 °C warming above pre-industrial levels[51]. The present estimate for maximum warming during the Last Interglacial is higher than some model simulations[52]. Comparisons of the new GAST reconstruction with additional palaeoclimatic reconstructions find strong correlations, as would be expected (Extended Data Table 3).

Validity testing the GAST reconstruction using PMIP model outputs. The PMIP model outputs are used to test the robustness of this research's method for estimating GAST. The question explored is: if the proposed approach of this paper is applied to grid cell values from the models for the same locations as the SST reconstructions, how would the estimated GAST value compare to the model's GAST value? For each of the nine PMIP model simulations available, nearest neighbour classification (via the *knn1* function in the *class* package[53] in the R statistical program, http://cran.r-project.org/web/packages/class/index.html) is used to identify the grid cell closest to each SST proxy reconstruction. The change in SST from present to the LGM at those locations is used as an input to the GAST estimation methods described above. The models' surface air temperature outputs are used to directly estimate GAST for each model. The final estimates of GAST using the 61 SST proxy reconstruction locations are then compared to the models' GAST values (Extended Data Fig. 5). The same procedure is repeated with only the locations of the 5 SST reconstructions that cover the full past 2 Myr (Extended Data Fig. 5). The median estimates for change in GAST at the LGM using PMIP model outputs are very similar (−4.5 °C directly from the models, versus −5.1 °C from the 61 record locations and −4.9 °C from the 5 record locations) when combined across the 9 models, and well within the estimated 95% interval. The method of this paper does find a larger range of uncertainty in GAST, as to be expected when comparing 61 or 5 grid cells to the full surface air temperature model outputs (Extended Data Fig. 5). Based on these 9 models that are available, the median estimates suggest the new GAST reconstruction may overestimate cooling by 10%, but the median estimates vary in direction and magnitude across the individual models and thus such a conclusion is specific to the set of model outputs.

Robustness of the GAST reconstruction to particular sample of cores. Early in the reconstruction the estimate is based on 61 reconstructions, but at 400 ka on only 14 reconstructions, at 800 ka on only 8 reconstructions, and at 2 Ma on only 5 reconstructions (Extended Data Fig. 2). This research uses three different analyses to assess the robustness of the final GAST reconstruction to the particular set of SST reconstructions currently available. First, the final GAST ensemble includes bootstrap Monte Carlo-style simulations that resample the reconstructions before following the methods to calculate GAST. The final ensemble of GAST time series are directly from this bootstrap simulation, and the overall uncertainty thus includes the uncertainty caused by the particular set of 61 reconstructions. Second, the entire methodology, including the bootstrap simulations, is repeated

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

for just the 11 clusters plus 18 individual records, as described above. The entire methodology, including the bootstrap simulations, is also repeated for just the 5 SST reconstructions that cover the full past 2 Myr. Extended Data Fig. 3a, b compares the median estimates from those two variations to the primary GAST reconstruction. Although the reduction in number of reconstructions causes larger uncertainty and potential different variance structures, the median estimates are very similar (0.997 correlation for the clustered version and 0.953 correlation for the 5-record version). Third, as described previously, this analysis's methodology is applied to the PMIP model outputs for both the full 61 reconstructions and just the 5 records that cover the full past 2 Myr and finds GAST estimates consistent with the models' air surface temperature outputs (Extended Data Fig. 5).

DWT reconstructions. This analysis compares GAST to 12 different proxy-based reconstructions of changes in DWT developed from three different methods (Extended Data Table 2). Elderfield *et al.*[4] estimate DWT using Mg/Ca ratios from bottom-dwelling foraminifera. Rohling *et al.*[5] estimate sea level using surface planktonic oxygen isotopes from Mediterranean Sea sediments and then use the sea level estimate to remove the ice-volume effect from the global benthic oxygen stable isotope data to estimate global DWT changes. Bates *et al.*[6] estimate DWT from oxygen isotope records from benthic foraminifera shells using regression analysis for ten different deep-ocean records, using a much simpler model for the relationship between benthic oxygen isotopes and DWT as well as older age models than the other two reconstructions.

GHG radiative forcing. For GHG concentrations from the past 800 kyr, I use stacked reconstructions from Antarctic ice cores[14,18,54]. The same 1-kyr interpolation method is applied to these records as described above for the SST reconstructions. I calculate the radiative forcing changes for CO_2 and methane from the past 800 kyr using the equations from Kohler *et al.*[17] and Hansen *et al.*[11] of total forcing $= \alpha(CO_2 \text{ forcing} + \beta \times \text{methane forcing})$. I apply highly conservative 95% intervals for the parameters in the proposed approximations: α (approximation for N_2O) is 1.12 (1.0–1.24, 95% interval), β (efficacy of methane) is 1.4 (1.0–1.8, 95% interval), and I apply conservative uncertainty of 20% (95% interval) for the overall equation.

Regression analyses. To investigate the relationship between the new GAST reconstruction and other palaeoclimatic reconstructions (for example, Extended Data Table 3 and lines in Fig. 2), I estimate 'GAST sensitivity' (the estimated linear relationship of change in GAST for each unit of change in the palaeoclimate record) and the correlation between the two reconstructions. I first randomly sample a single time series from the GAST simulation ensemble of potential time series (described above) and a single time series from the simulation ensemble for the comparison record. I then normalize each record to be deviations from present, where present is defined as the mean value over 1–5 ka. It is necessary to use 5 ka because not all reconstructions have estimates for 1–3 ka. I use weighted least squares regressions without an intercept, because both records are deviations from present. I test both linear functions and nonlinear, such as quadratic, relationships and I use an ANOVA test to assess the improved fit of alternative functional relationships. I also quantify the correlation of the two time series. I then repeat the analyses for at least 500 random draws of time series from each reconstruction. Because there is high autocorrelation in most palaeoclimatic reconstructions, I include an autoregressive model to evaluate the potential underestimation of error in the regression coefficients that can be caused by autocorrelation[55]. For example, the autoregressive model for GAST as a function of radiative forcing from GHGs does not significantly change the regression coefficient, as predicted by theory, but it does increase the estimated standard error of the regression coefficient from 1.8% to 3.5% of the median value. However, the regression standard error is an insignificant contribution to the overall uncertainty analysis, which estimates an analogous standard error of 19% overall. To assess whether the regression results vary for deep glacial states, I define deep glacial periods as the bottom 25% of the comparison record (less than -2 W m^{-2} for GHG radiative forcing) and repeat the analyses with the two sets of time periods separately. I convert the estimate of change in GAST to change in GHG radiative forcing to ESS by multiplying by 3.7 W m^{-2} (the change in radiative forcing from a doubling of CO_2)[2].

Investigations of ESS are limited by the availability of GHG reconstructions. Because reconstructions of atmospheric CO_2 before 800 ka are highly uncertain and limited in temporal resolution and reconstructions of methane before 800 ka do not exist, this research focuses on analysis of the past 800 kyr of GHG reconstructions. However, I analyse the reconstructions of CO_2 based on boron isotopes[56] that are available for limited time points across the past 2 Myr. I use the equation from Kohler *et al.*[17] to estimate CO_2 radiative forcing. I repeat the regression analyses described above for the limited set of time points to estimate GAST sensitivity to CO_2 radiative forcing and I repeat the analysis separately for 0–1 Ma and 1–2 Ma (Extended Data Table 3 and Extended Data Fig. 7). For the past 1 Myr, the results are substantially similar to the results obtained from a comparison with CO_2 reconstructed from ice cores, with lower correlation and coefficient estimates as would be expected due to the higher uncertainty in the CO_2 reconstruction. Prior to 1 Ma, CO_2 radiative forcing from boron isotopes is poorly correlated with GAST changes, suggesting either a decoupling of GAST and CO_2 before the MPT, or more likely, errors in the CO_2 reconstruction, GAST reconstruction, and/or the relative dating of the records.

Probabilistic breakpoint identification. This research uses probabilistic breakpoint identification to estimate changes in GAST time trends over the past 2 Myr. Breakpoint simulation detects and identifies changes within time series by decomposing the time series into linear trends and breakpoints. This research uses the *bfast* function from the *bfast* package[57] in the R statistical program (http://cran.r-project.org/web/packages/bfast/index.html), which iteratively estimates time trends and break points through a piecewise linear trend to identify optimal values. The *bfast* simulation program is applied independently to 500 randomly selected time series from the final ensemble of GAST time series to estimate the empirically-fitted frequency distributions in Fig. 3. The program also estimates the time trend before and after the breakpoint. No smoothing is used in this analysis.

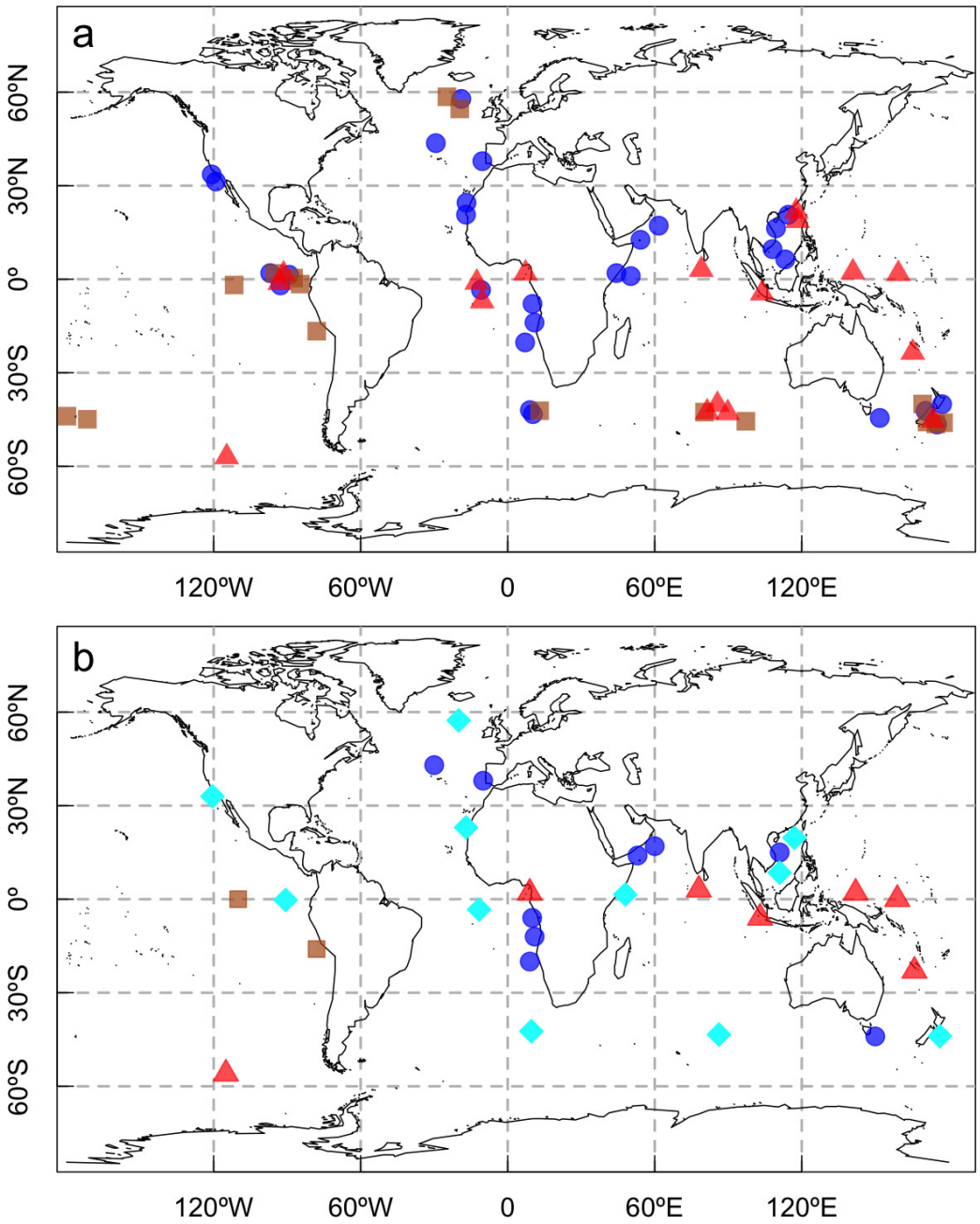
Data and code availability. Supplementary Methods includes R code for key methods described in the paper. Supplementary Data includes the new GAST reconstruction at 2.5%, 5%, 25%, 50%, 75%, 95% and 97.5% likelihood values, and the 61 SST reconstructions used to create the GAST reconstruction, including a detailed summary table.

31. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum — Part 1: experiments and large-scale features. *Clim. Past* **3,** 261–277 (2007).
32. Harrison, S. P. *et al.* Climate model benchmarking with glacial and mid-Holocene climates. *Clim. Dyn.* **43,** 671–688 (2014).
33. Hargreaves, J. C., Annan, J. D., Yoshimori, M. & Abe-Ouchi, A. Can the Last Glacial Maximum constrain climate sensitivity? *Geophys. Res. Lett.* **39,** L24702 (2012).
34. Mix, A. C., Bard, E. & Schneider, R. Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quat. Sci. Rev.* **20,** 627–657 (2001).
35. Müller, P. J., Kirst, G., Ruhland, G., Von Storch, I. & Rossel-Mele, A. Calibration of the alkenone paleotemperature index $U_{37}^{K'}$ based on core-tops from the eastern South Atlantic and the global ocean (60°N-60°S). *Geochim. Cosmochim. Acta* **62,** 1757–1772 (1998).
36. Herbert, T. D. Review of alkenone calibrations (culture, water column, and sediments). *Geochem. Geophys. Geosyst.* **2,** http://dx.doi.org/10.1029/2000GC000055 (2001).
37. Mashiotta, T. A., Lea, D. W. & Spero, H. J. Glacial-interglacial changes in Subantarctic sea surface temperature and δ^{18}O-water using foraminiferal Mg. *Earth Planet. Sci. Lett.* **170,** 417–432 (1999).
38. Elderfield, H. & Ganssen, G. Past temperature and δ^{18}O of surface ocean waters inferred from foraminiferal Mg/Ca ratios. *Nature* **405,** 442–445 (2000).
39. Barrows, T. T., Juggins, S., De Deckker, P., Calvo, E. & Pelejero, C. Long-term sea surface temperature and climate change in the Australian-New Zealand region. *Paleoceanography* **22,** PA2215 (2007).
40. Haam, E. & Huybers, P. A test for the presence of covariance between time-uncertain series of data with application to the Dongge Cave speleothem and atmospheric radiocarbon records. *Paleoceanography* **25,** PA001713 (2010).
41. Lin, L., Khider, D., Lisiecki, L. E. & Lawrence, C. E. Probabilistic sequence alignment of stratigraphic records. *Paleoceanography* **29,** 976–989 (2014).
42. Martinson, D. G. *et al.* Age dating and the orbital theory of the ice ages: development of a high-resolution 0 to 300000-year chronostratigraphy. *Quat. Res.* **27,** 1–29 (1987).
43. Huybers, P. Glacial variability over the last two million years: an extended depth-derived age model, continuous obliquity pacing, and the Pleistocene progression. *Quat. Sci. Rev.* **26,** 37–55 (2007).
44. Haywood, A. M. Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project. *Clim. Past* **9,** 191–209 (2013).
45. Bell, D. B., Jung, S. J. A. & Kroon, D. The Plio-Pleistocene development of Atlantic deep-water circulation and its influence on climate trends. *Quat. Sci. Rev.* **123,** 265–282 (2015).
46. Kageyama, M. *et al.* The Last Glacial Maximum climate over Europe and western Siberia: a PMIP comparison between models and data. *Clim. Dyn.* **17,** 23–43 (2001).
47. Ballantyne, A. P., Lavine, M., Crowley, T. J., Liu, J. & Baker, P. B. Meta-analysis of tropical surface temperatures during the Last Glacial Maximum. *Geophys. Res. Lett.* **32,** L05712 (2005).
48. Schneider von Deimling, T., Ganopolski, A., Held, H. & Rahmstorf, S. How cold was the Last Glacial Maximum? *Geophys. Res. Lett.* **33,** L14709 (2006).
49. Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C. & Oppenheimer, M. Probabilistic assessment of sea level during the last interglacial stage. *Nature* **462,** 863–867 (2009).
50. Sime, L. C., Wolff, E. W., Oliver, K. I. C. & Tindall, J. C. Evidence for warmer interglacials in East Antarctic ice cores. *Nature* **462,** 342–345 (2009).
51. Turney, C. S. & Jones, R. T. Does the Agulhas Current amplify global temperatures during super-interglacials? *J. Quat. Sci.* **25,** 839–843 (2010).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

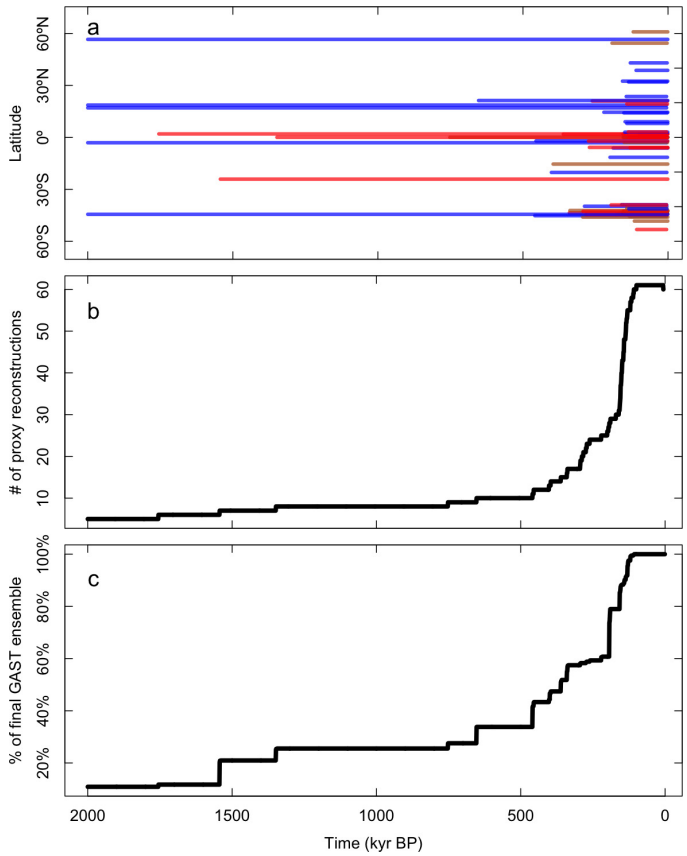
52. Otto-Bliesner, B. L. *et al.* How warm was the last interglacial? New model–data comparisons. *Phil. Trans. R. Soc.* **371,** http://dx.doi.org/10.1098/rsta.2013.0097 (2013).
53. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn (Springer, 2002).
54. Loulergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH_4 over the past 800,000 years. *Nature* **453,** 383–386 (2008).
55. Abraham, B. & Ledolter, J. *Introduction to Regression Modeling* (Duxbury Press, 2006).
56. Honisch, B., Hemming, G., Archer, D., Siddal, M. & McManus, J. Atmospheric carbon dioxide concentration across the Mid-Pleistocene Transition. *Science* **324,** 1551–1554 (2009).
57. Verbesselt, J., Hyndman, R., Newnham, G. & Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **114,** 106–115 (2010).
58. de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F. & Beaufort, L. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature* **433,** 294–298 (2005).
59. Lea, D. W. The 100 000-yr cycle in tropical SST, greenhouse forcing, and climate sensitivity. *J. Clim.* **17,** 2170–2179 (2004).
60. Lea, D. W. *et al.* Paleoclimate history of Galapagos surface waters over the last 135,000yr. *Quat. Sci. Rev.* **25,** 1152–1167 (2006).
61. Medina-Elizalde, M. & Lea, D. W. The mid-Pleistocene transition in the tropical Pacific. *Science* **310,** 1009–1012 (2005).
62. Mohtadi, M. *et al.* Late Pleistocene surface and thermocline conditions of the eastern tropical Indian Ocean. *Quat. Sci. Rev.* **29,** 887–896 (2010).
63. Nürnberg, D., Muller, A. & Schneider, R. R. Paleo-sea surface temperature calculations in the equatorial east Atlantic from Mg/Ca ratios in planktic foraminifera: A comparison to sea surface temperature estimates from $U_{37}^{K'}$, oxygen isotopes, and foraminiferal transfer function. *Paleoceanography* **15,** 124–134 (2000).
64. Oppo, D. W. & Sun, Y. B. Amplitude and timing of sea-surface temperature change in the northern South China Sea: dynamic link to the East Asian monsoon. *Geology* **33,** 785–788 (2005).
65. Pahnke, K., Zahn, R., Elderfield, H. & Schulz, M. 340,000-year centennial-scale marine record of southern hemisphere climatic oscillation. *Science* **301,** 948–952 (2003).
66. Rickaby, R. E. M. & Elderfield, H. Planktonic foraminiferal Cd/Ca: paleonutrients or paleotemperature? *Paleoceanography* **14,** 293–303 (1999).
67. Russon, T. *et al.* Inter-hemispheric asymmetry in the early Pleistocene Pacific warm pool. *Geophys. Res. Lett.* **37,** L11601 (2010).
68. Saraswat, R., Nigam, R., Weldeab, S., Mackensen, A. & Naidu, P. D. A first look at past sea surface temperatures in the equatorial Indian Ocean from Mg/Ca in foraminifera. *Geophys. Res. Lett.* **32,** L24605 (2005).
69. Wei, G. J., Deng, W. F., Liu, Y. & Li, X. H. High-resolution sea surface temperature records derived from foraminiferal Mg/Ca ratios during the last 260 ka in the northern South China Sea. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **250,** 126–138 (2007).
70. Weldeab, S., Lea, D. W., Schneider, R. R. & Andersen, N. 155,000 years of West African monsoon and ocean thermal evolution. *Science* **316,** 1303–1307 (2007).
71. Brathauer, U. & Abelmann, A. Late Quaternary variations in sea surface temperatures and their relationship to orbital forcing recorded in the Southern Ocean (Atlantic sector). *Paleoceanography* **14,** 135–148 (1999).
72. Kandiano, E. S., Bauch, H. A. & Muller, A. Sea surface temperature variability in the North Atlantic during the last two glacial-interglacial cycles: comparison of faunal, oxygen isotopic, and Mg/Ca-derived records. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **204,** 145–164 (2004).
73. Labeyrie, L. *et al.* Hydrographic changes of the Southern Ocean (southeast Indian sector) over the last 230 kyr. *Paleoceanography* **11,** 57–76 (1996).
74. Pisias, N. G. & Mix, A. C. Spatial and temporal oceanographic variability of the eastern equatorial Pacific during the late Pleistocene: evidence from Radiolaria microfossils. *Paleoceanography* **12,** 381–393 (1997).
75. Weaver, P. P. E., Carter, L. & Neil, H. L. Response of surface water masses and circulation to Late Quaternary climate change east of New Zealand. *Paleoceanography* **13,** 70–83 (1998).
76. Weaver, P. P. E. *et al.* Combined coccolith, foraminiferal, and biomarker reconstruction of paleoceanographic conditions over the past 120 kyr in the northern North Atlantic (59°N, 23°W). *Paleoceanography* **14,** 336–349 (1999).
77. Bard, E. Climate shock — Abrupt changes over millennial time scales. *Phys. Today* **55,** 32–38 (2002).
78. Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone palaeothermometry. *Nature* **385,** 707–710 (1997).
79. Clemens, S. C., Prell, W. L., Sun, Y., Liu, Z. & Chen, G. Southern Hemisphere forcing of Pliocene $\delta^{18}O$ and the evolution of Indo-Asian monsoons. *Paleoceanography* **23,** PA4210 (2008).
80. Herbert, T. D., Peterson, L. C., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328,** 1530–1534 (2010).
81. Dubois, N. *et al.* Millennial-scale variations in hydrography and biogeochemistry in the Eastern Equatorial Pacific over the last 100 kyr. *Quat. Sci. Rev.* **30,** 210–223 (2011).
82. Eglinton, G. *et al.* Molecular record of secular sea surface temperature changes on 100-year timescales for glacial terminations I, II and IV. *Nature* **356,** 423–426 (1992).
83. Horikawa, K., Minagawa, M., Murayama, M., Kato, Y. & Asahi, H. Spatial and temporal sea-surface temperatures in the eastern equatorial Pacific over the past 150 kyr. *Geophys. Res. Lett.* **33,** L13605 (2006).
84. Lawrence, K. T., Liu, Z. H. & Herbert, T. D. Evolution of the eastern tropical Pacific through Plio-Pleistocene glaciation. *Science* **312,** 79–83 (2006).
85. Lawrence, K. T., Herbert, T. D., Brown, C. M., Raymo, M. E. & Haywood, A. M. High-amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography* **24,** PA2218 (2009).
86. Müller, P. J., Cepek, M., Ruhland, G. & Schneider, R. R. Alkenone and coccolithophorid species changes in late Quaternary sediments from the Walvis Ridge: implications for the alkenone paleotemperature method. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **135,** 71–96 (1997).
87. Pahnke, K. & Sachs, J. P. Sea surface temperatures of southern midlatitudes 0-160 kyr BP. *Paleoceanography* **21,** PA2003 (2006).
88. Pelejero, C., Grimalt, J. O., Heilig, S., Kienast, M. & Wang, L. J. High-resolution U^{K}_{37} temperature reconstructions in the South China Sea over the past 220 kyr. *Paleoceanography* **14,** 224–231 (1999).
89. Pelejero, C., Calvo, E., Barrows, T. T., Logan, G. A. & De Deckker, P. South Tasman Sea alkenone palaeothermometry over the last four glacial/interglacial cycles. *Mar. Geol.* **230,** 73–86 (2006).
90. Rostek, F., Bard, E., Beaufort, L., Sonzogni, C. & Ganssen, G. Sea surface temperature and productivity records for the past 240 kyr in the Arabian Sea. *Deep-Sea Res.* **44,** 1461–1480 (1997).
91. Sachs, J. P. & Anderson, R. F. Fidelity of alkenone paleotemperatures in southern Cape Basin sediment drifts. *Paleoceanography* **18,** 1082 (2003).
92. Schneider, R. R., Muller, P. J. & Ruhland, G. Late Quaternary surface circulation in the east equatorial South Atlantic: evidence from alkenone sea surface temperatures. *Paleoceanography* **10,** 197–219 (1995).
93. Sicre, M. A. *et al.* Biomarker stratigraphic records over the last 150 kyears off the NW African coast at 25°N. *Org. Geochem.* **31,** 577–588 (2000).
94. Villanueva, J., Grimalt, J. O., Cortijo, E., Vidal, L. & Labeyrie, L. Assessment of sea surface temperature variations in the central North Atlantic using the alkenone unsaturation index $U_{37}^{k'}$. *Geochim. Cosmochim. Acta* **62,** 2421–2427 (1998).
95. Yamamoto, M., Oba, T., Shimamune, J. & Ueshima, T. Orbital-scale anti-phase variation of sea surface temperature in mid-latitude North Pacific margins during the last 145,000 years. *Geophys. Res. Lett.* **31,** L16311 (2004).
96. Yamamoto, M., Yamamuro, M. & Tanaka, Y. The California current system during the last 136,000 years: response of the North Pacific High to precessional forcing. *Quat. Sci. Rev.* **26,** 405–414 (2007).
97. Zhao, M. X., Huang, C. Y., Wang, C. C. & Wei, G. J. A millennial-scale $U_{37}^{K'}$ sea-surface temperature record from the South China Sea (8°N) over the last 150 kyr: monsoon and sea-level influence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **236,** 39–55 (2006).
98. Masson-Delmotte, V. *et al.* Atmospheric science: GRIP deuterium excess reveals rapid and orbital-scale changes in Greenland moisture origin. *Science* **309,** 118–121 (2005).
99. Masson-Delmotte, V. *et al.* Past temperature reconstructions from deep ice cores: relevance for future climate change. *Clim. Past* **2,** 145–165 (2006).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



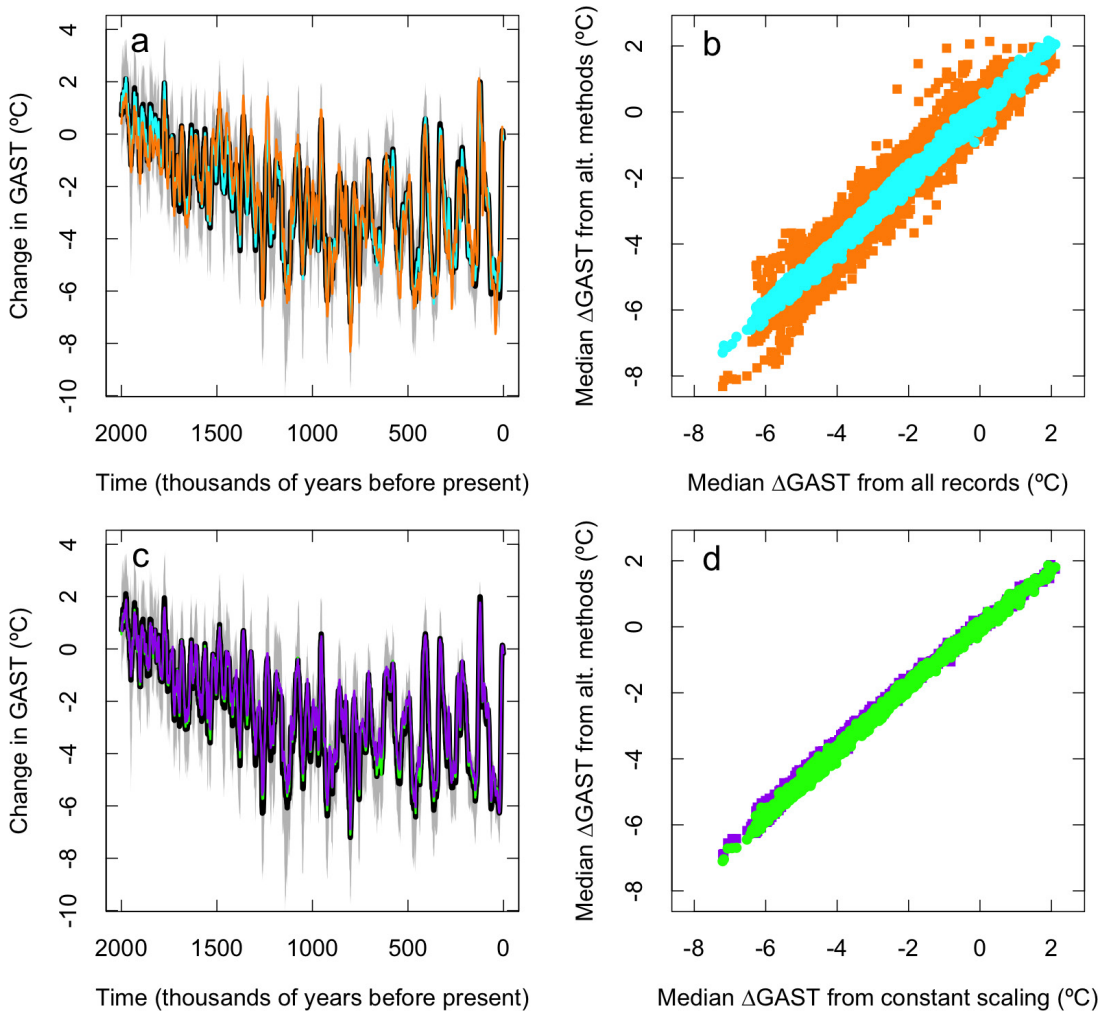
Extended Data Figure 1 | Spatial distribution of the SST proxy reconstructions used in this analysis. a, All 61 SST records, with methods as follows: from alkenone indices, blue circles; from Mg/Ca ratios, red triangles; and from species assemblage methods, brown squares. **b**, Repeated after clustering records within 5° latitude/longitude of each other, with the 11 clusters in cyan diamonds and the remaining 18 records as in **a**.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



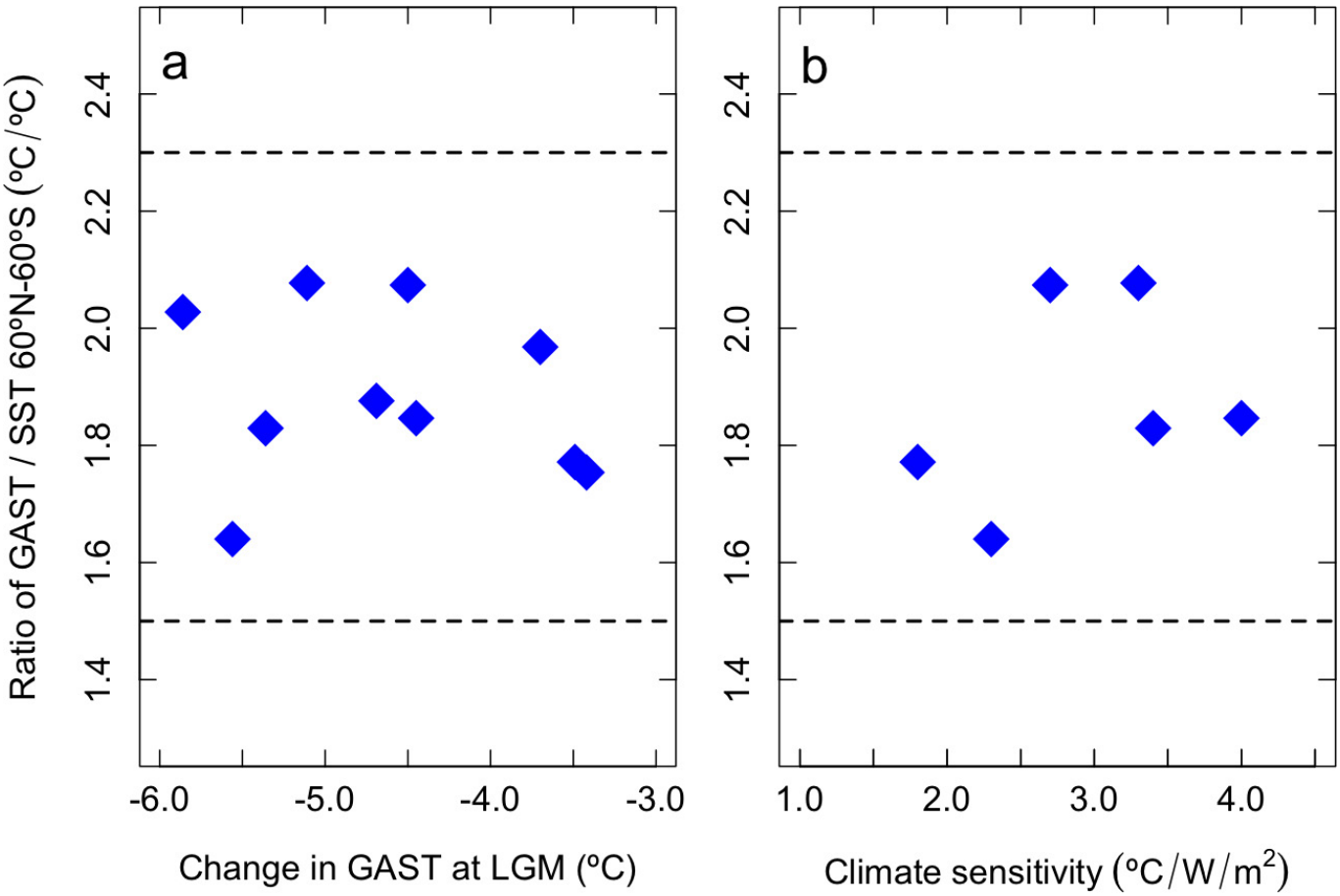
Extended Data Figure 2 | Temporal distribution of the 61 SST proxy reconstructions used in this analysis. a, Reconstruction length versus latitude, colours as in Extended Data Fig. 1. **b**, Empirical cumulative distribution function for lengths of the SST proxy reconstructions. **c**, Empirical cumulative distribution function for lengths of GAST time series in the final simulation ensemble of potential GAST time series.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



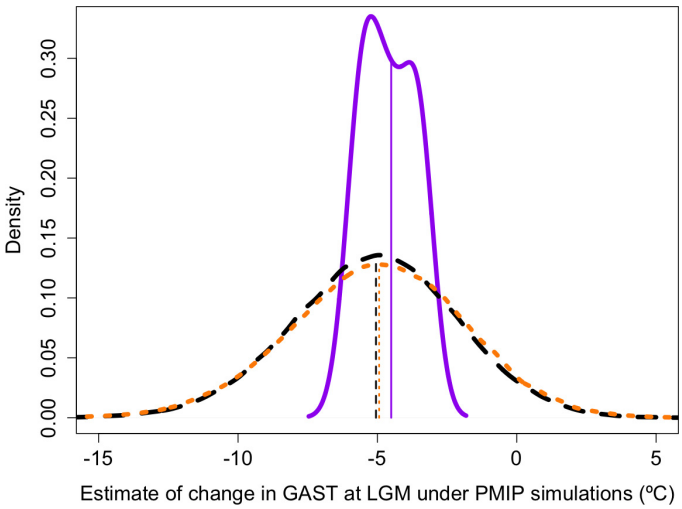
Extended Data Figure 3 | Comparison of different methods used to estimate GAST. a, The primary GAST estimate (using 61 proxy reconstructions) is plotted as a function of time, with the median in black and the 95% interval in grey. The GAST estimation method is repeated for a clustering of the data (11 clusters and 18 individual reconstructions), with the median shown in cyan, and for only the 5 proxy reconstructions that cover the past 2 Myr, with the median shown in orange. **b**, The median time series from each alternative method are plotted against the primary median GAST estimate, with the clustered version in cyan circles and the 5-record version in orange squares. **c**, The primary GAST estimate is plotted as a function of time, with the median in black and the 95% interval in grey. An alternative GAST estimation method using a time-varying scalar based on the deep-sea oxygen isotopes median estimate is shown in green, and another estimation method based on the relative sea level median estimate is shown in purple. **d**, The median time series from each alternative method is plotted against the primary median GAST estimate, with the reconstruction scaled using deep sea oxygen isotopes shown in green circles and the reconstruction scaled using relative sea level shown in purple squares.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



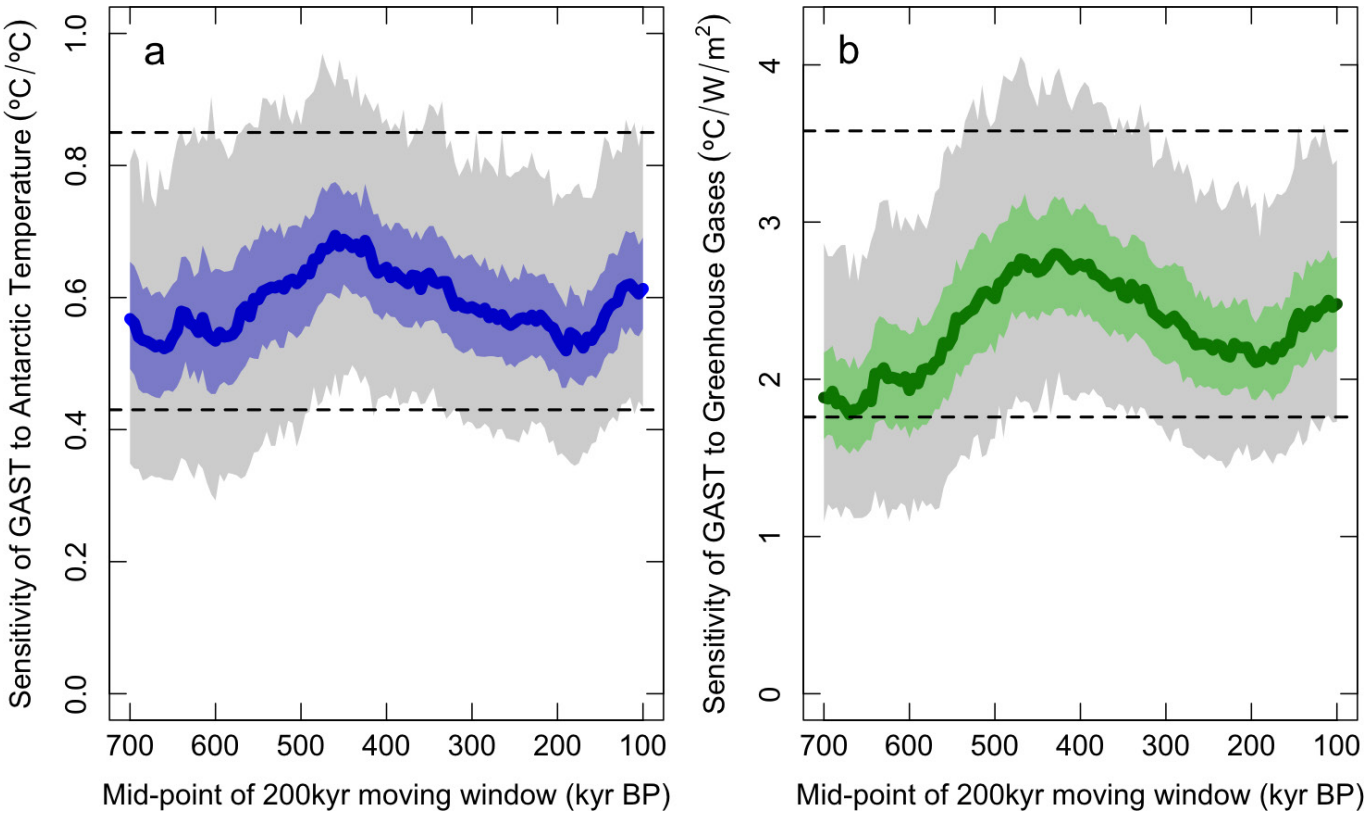
Extended Data Figure 4 | Estimates of the ratio of change in GAST to change in average SST. a, **b**, Scatter plots show the dependency of the ratio of change in GAST to change in average SST over the latitudinal zone 60° N to 60° S from PMIP2 and PMIP3 climate model simulations[31,32] as a function of change in GAST at the LGM (**a**) and of model climate sensitivity (**b**). The climate sensitivity estimates (in °C per W m^{-2}) are from ref. 33. Dashed lines show the scalar range used in this analysis.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



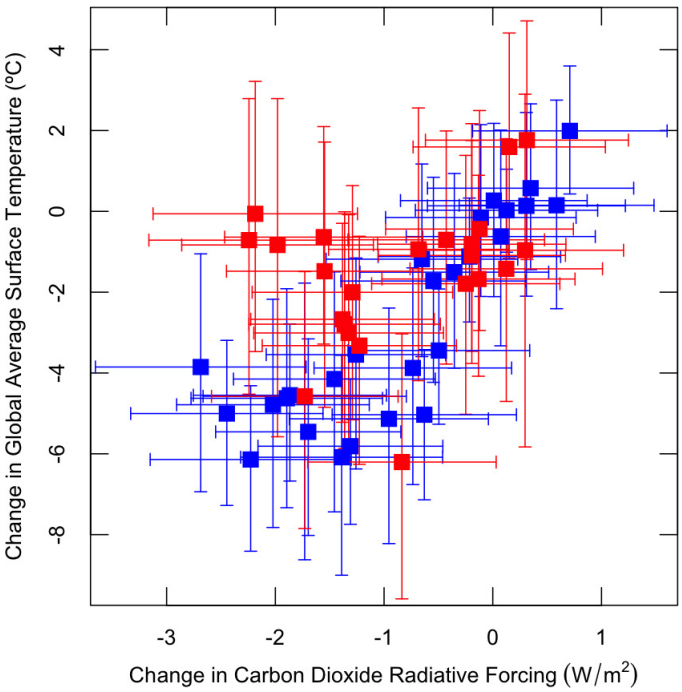
Extended Data Figure 5 | Estimating change in GAST at the LGM using simulations drawn from PMIP model outputs. The solid, purple line is the empirically fitted frequency distribution (shown in density on the *y* axis) of GAST estimated from the full air surface temperature outputs from the 9 PMIP models. The dashed, black line is the distribution of GAST estimated using the method in the present paper and the PMIP SST outputs drawn from only the locations of the 61 proxy reconstructions. The short-dashed, orange line is the same analysis completed for only the 5 proxy reconstructions that cover the past 2 Myr. The thin vertical lines are the medians of each distribution.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 6 | The dependence of coupling relationships over time for GAST on changes in Antarctic temperature and GHG radiative forcing. a, **b**, Regression results of change in GAST as a function of change in Antarctic temperature[14] (**a**) and of change in GHG radiative forcing[17,18,54] (**b**) are calculated for moving 200-kyr-long time windows every 5 kyr. The solid line shows the median estimates, with the coloured and grey-shaded areas showing the 50% and 95% intervals, respectively. The dashed lines show the 95% intervals calculated from the entire time series.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



Extended Data Figure 7 | Comparison of changes in GAST to changes in CO_2 radiative forcing. Boron-isotope-based proxy reconstruction of CO_2 from refs 17, 56. Blue points are from 0–1 Ma, red points are from 1–2 Ma, and error bars show 95% intervals.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 1 | Database of SST proxy reconstructions based on Mg/Ca ratio and species assemblages used in estimating GAST

Citation	Latitude (+N/-S)	Longitude (+E/-W)	Start Date (yr BP)	End Date (yr BP)
Mg/Ca Ratio SST Proxy Reconstructions				
de Garidel-Thoron et al. 2005 [58]	2	142	7,000	1,755,000
Lea 2004 [59]	2	-91	1,000	361,000
Lea et al. 2006 [60]	0	-92	1,200	135,100
Mashiotta et al. 1999 [37]	-43	80	2,700	293,700
Mashiotta et al. 1999 [37]	-56	-115	8,110	108,450
Medina-Elizalde and Lea 2005 [61]	0	159	4,300	1,348,000
Mohtadi et al. 2011 [62]	-6	103	0	131,400
Nürnberg et al. 2000 [63]	-2	-12	240	274,630
Nürnberg et al. 2000 [63]	-6	-11	1,000	271,000
Oppo and Sun 2005 [64]	20	118	1,943	142,571
Pahnke et al. 2003 [65]	-45	175	1,951	340,835
Rickaby and Elderfield 1999 [66]	-40	85	4,750	196,710
Rickaby and Elderfield 1999 [66]	-44	90	3,600	283,670
Russon et al. 2010 [67]	-23	166	0	2,438,000
Saraswat et al. 2005 [68]	3	78	5,434	137,333
Wei et al. 2007 [69]	20	117	1,740	261,300
Weldeab et al. 2007 [70]	2	9	360	155,420
Species Assemblages SST Proxy Reconstructions				
Barrows et al. 2007 [39]	-42	170	3,830	142,712
Barrows et al. 2007 [39]	-46	175	5,401	133,995
Brathauer and Abelmann 1999 [71]	-43	12	387	338,519
Kandiano et al. 2004 [72]	54	-20	2,375	193,065
Labeyrie et al. 1996 [73]; Rickaby and Elderfield 1999 [66]	-46	96	4,590	149,250
Martinson et al. 1987 [42]	-44	80	750	294,000
Pisias and Mix 1997 [74]	0	-96	360	752,047
Pisias and Mix 1997 [74]	-3	-83	1,914	151,051
Pisias and Mix 1997 [74]	0	-110	880	144,763
Pisias and Mix 1997 [74]	0	-86	1,479	154,653
Pisias and Mix 1997 [74]	-16	-78	785	466,520
Weaver et al. 1998 [75]	-44	-172	0	170,000
Weaver et al. 1998 [75]	-46	172	0	116,705
Weaver et al. 1998 [75]	-45	179	170	120,880
Weaver et al. 1998 [75]	-43	-178	0	154,600
Weaver et al. 1999 [76]	60	-23	0	120,000

Data are taken from refs 37, 39, 42, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75 and 76.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 2 | Database of SST proxy reconstructions based on alkenone indices used in estimating GAST

Citation	Latitude (+N/-S)	Longitude (+E/-W)	Start Date (yr BP)	End Date (yr BP)
Alkenone Indices SST Proxy Reconstructions				
Bard 2002 [77]	38	-10	0	110,100
Bard et al. 1997 [78]	0	46	0	151,500
Bard et al. 1997 [78]	3	50	0	149,200
Clemens et al. 2008 [79]; Herbert et al. 2008 [80]	19	116	4,178	2,155,910
Dubois et al. 2011 [81]	2	-91	1,070	99,290
Eglinton et al. 1992 [82]	21	-18	736	653,042
Herbert et al. 2010 [80]	-2	-12	1,290	445,000
Herbert et al. 2010 [80]	17	60	7,110	3,330,480
Herbert et al. 2010 [80]	-3	-91	5,550	5,069,892
Horikawa et al. 2006 [83]	0	-95	1,100	154,097
Lawrence et al. 2006 [84]	-3	-91	5,230	5,089,802
Lawrence et al. 2009 [85]	58	-17	0	4,012,230
Martinez-Garcia et al. 2010 [29]	-43	9	0	3,642,410
Muller et al. 1997 [86]	-20	9	3,300	402,900
Pahnke and Sachs 2006 [87]	-46	175	1,950	156,360
Pahnke and Sachs 2006 [87]	-40	178	3,320	135,050
Pelejero et al. 1999 [88]	15	111	0	221,100
Pelejero et al. 1999 [88]	8	112	0	142,290
Pelejero et al. 2006 [89]	-42	170	3,570	288,540
Pelejero et al. 2006 [89]	-44	150	5,045	459,632
Rostek et al. 1997 [90]	14	53	2,300	152,600
Pahnke and Sach 2006 [87]; Sachs and Anderson 2003 [91]	-41	9	6,080	160,000
Schneider et al. 1995 [92]	-6	10	400	189,000
Schneider et al. 1995 [92]	-12	11	1,300	200,600
Sicre et al. 2000 [93]	25	-16	5,000	144,000
Villanueva et al. 1998 [94]	43	-30	2,900	285,900
Yamamoto et al. 2004, 2007 [95,96]	32	-119	870	157,048
Yamamoto et al. 2004, 2007 [95,96]	34	-122	3,186	136,475
Zhao et al. 2006 [97]	9	110	905	148,886

Data are taken from refs 29, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96 and 97.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Extended Data Table 3 | Comparisons of GAST with other important palaeoclimate reconstructions

Record Compared to GAST	GAST Sensitivity		Correlation	
	Median	(95% Interval)	Median	(95% Interval)
Total Greenhouse Gas Radiative Forcing[14,17,18,54]	2.5 °C/W/m^2	(1.8, 3.6)	0.82	(0.66, 0.92)
CO_2 Radiative Forcing from ice cores[17,18]	3.0 °C/W/m^2	(2.1, 4.5)	0.82	(0.63, 0.92)
CO_2 Radiative Forcing from Boron-isotopes, 0-1Ma[17,56]	2.4 °C/W/m^2	(1.6, 3.4)	0.74	(0.62, 0.84)
CO_2 Radiative Forcing from Boron-isotopes, 1-2Ma[17,56]	1.3 °C/W/m^2	(0.57, 2.2)	0.28	(-0.10, 0.55)
Antarctic Temperature stack[14]	0.61 °C/°C	(0.43, 0.85)	0.72	(0.59, 0.81)
Greenland Temperature-GRIP[98]	0.31 °C/°C	(0.23, 0.43)	0.89	(0.83, 0.92)
Greenland Temperature-NGRIP[99]	0.27 °C/°C	(0.19, 0.35)	0.91	(0.87, 0.93)
Deep-Sea Oxygen Isotope[30]	-4.2 °C/‰	(-5.6, -3.0)	-0.85	(-0.94, -0.58)
Deep-Water Temperature from Southwest Pacific (Core 1123)[4]	2.2 °C/°C	(1.5, 3.2)	0.64	(0.35, 0.78)
Deep-Water Temperature from Mediterranean Sea[5]	1.7 °C/°C	(0.8, 2.9)	0.47	(0.14, 0.64)
Deep-Water Temperature from Temperate North Atlantic (DSDP 607)[6]	1.5 °C/°C	(1.0, 2.2)	0.48	(0.36, 0.64)
Deep-Water Temperature from East Tropical Atlantic (ODP 659)[6]	1.5 °C/°C	(1.0, 2.1)	0.84	(0.71, 0.90)
Deep-Water Temperature from Indian Ocean (ODP 758)[6]	3.2 °C/°C	(2.0, 4.6)	0.27	(0.14, 0.42)
Deep-Water Temperature from Equatorial Eastern Pacific (ODP 849)[6]	3.0 °C/°C	(2.1, 4.2)	0.85	(0.78, 0.89)
Deep-Water Temperature from Sub-Arctic North Atlantic (ODP 980, ODP 981)[6]	1.4 °C/°C	(1.0, 1.9)	0.75	(0.40, 0.85)
Deep-Water Temperature from Sub-Antarctic South Atlantic (ODP 1090)[6]	2.1 °C/°C	(1.4, 3.0)	0.77	(0.61, 0.82)
Deep-Water Temperature from South China Sea (ODP 1143)[6]	2.2 °C/°C	(1.5, 3.1)	0.79	(0.64, 0.85)
Deep-Water Temperature from South China Sea (ODP 1148)[6]	3.5 °C/°C	(2.4, 5.0)	0.54	(0.42,0.64)
Deep-Water Temperature from Equatorial Eastern Pacific composite (V19-30, ODP 677, ODP 846)[6]	2.4 °C/°C	(1.6, 3.3)	0.60	(0.37, 0.69)
Deep-Water Temperature from Southwest Pacific (Core 1123)[6]	1.9 °C/°C	(1.3, 2.6)	0.81	(0.58, 0.93)

'GAST sensitivity' is the estimated linear relationship of change in GAST for each unit of change in the palaeoclimate record. Data are taken from refs 4, 5, 6, 14, 17, 18, 30, 54, 56, 98 and 99.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.