

# Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models

Harry J. Dowsett<sup>1\*</sup>, Marci M. Robinson<sup>1</sup>, Alan M. Haywood<sup>2</sup>, Daniel J. Hill<sup>2,3</sup>, Aisling M. Dolan<sup>2</sup>, Danielle K. Stoll<sup>1</sup>, Wing-Le Chan<sup>4</sup>, Ayako Abe-Ouchi<sup>4,5</sup>, Mark A. Chandler<sup>6</sup>, Nan A. Rosenbloom<sup>7</sup>, Bette L. Otto-Bliesner<sup>7</sup>, Fran J. Bragg<sup>8</sup>, Daniel J. Lunt<sup>8</sup>, Kevin M. Foley<sup>1</sup> and Christina R. Riesselman<sup>1</sup>

**In light of mounting empirical evidence that planetary warming is well underway, the climate research community looks to palaeoclimate research for a ground-truthing measure with which to test the accuracy of future climate simulations. Model experiments that attempt to simulate climates of the past serve to identify both similarities and differences between two climate states and, when compared with simulations run by other models and with geological data, to identify model-specific biases. Uncertainties associated with both the data and the models must be considered in such an exercise. The most recent period of sustained global warmth similar to what is projected for the near future occurred about 3.3–3.0 million years ago, during the Pliocene epoch. Here, we present Pliocene sea surface temperature data, newly characterized in terms of level of confidence, along with initial experimental results from four climate models. We conclude that, in terms of sea surface temperature, models are in good agreement with estimates of Pliocene sea surface temperature in most regions except the North Atlantic. Our analysis indicates that the discrepancy between the Pliocene proxy data and model simulations in the mid-latitudes of the North Atlantic, where models underestimate warming shown by our highest-confidence data, may provide a new perspective and insight into the predictive abilities of these models in simulating a past warm interval in Earth history. This is important because the Pliocene has a number of parallels to present predictions of late twenty-first century climate.**

Our understanding of future climate impacts, as well as our ability to adapt to and mitigate effects, relies heavily on the tools with which we explore future scenarios. Numerical models of the climate system have evolved rapidly over the past several decades, in part as a response to the demand for faster and more confident projections of future conditions<sup>1</sup>. As humans have not yet experienced or been able to measure the magnitude of climate change projected for the end of this century, it is difficult to assess the performance of computer models. It has become common practice to hindcast past climate conditions and to verify those efforts using environmental reconstructions based on multiple-proxy palaeodata<sup>2</sup>. The confidence we place in the palaeoestimates is thus paramount to the understanding of model strengths and weaknesses. The most complete reconstruction of a past period of global warmth describes the mid-Piacenzian, an interval within the Pliocene of sustained warmth ~3.3–3.0 million years (Myr) ago, immediately before the intensification of large-scale Northern Hemisphere glaciation<sup>3</sup>. Here, we present an assessment of the confidence determined for each estimate of mean annual sea surface temperature (SST) from 95 sites distributed throughout the mid-Piacenzian global ocean. The formulation for this confidence assessment is presented in the Supplementary Information and all estimates and confidence levels are shown in Supplementary Table S1. Most of our mid-Piacenzian SST estimates are based on quantitative analysis of planktonic foraminiferal faunas from Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) cores<sup>4–6</sup>. Wherever possible,

additional proxies (alkenone and/or Mg:Ca palaeothermometry) provide a more robust understanding of mixed-layer conditions<sup>7</sup>. Furthermore, regional and environmental conditions sometimes allow the inclusion of other biotic proxies (for example, molluscs, bryozoa, diatoms, dinoflagellates, radiolaria and ostracods) that enrich our holistic understanding of the palaeoenvironment<sup>8–12</sup>.

## Mid-Piacenzian SST

Faunal estimates based on planktonic foraminifera were derived using either a factor analytic transfer function<sup>13</sup> or a revised modern analogue technique<sup>14–16</sup> with modifications to allow for extension to Pliocene-age assemblages. Poor carbonate preservation in many regions of the Pacific Ocean made siliceous microfossils (diatoms and radiolaria) a better choice for quantitative temperature estimation in some locations<sup>17–20</sup>. Diatoms were used almost exclusively in the Southern Ocean where they are excellent indicators of the position of the sea-ice margin and Antarctic polar fronts<sup>4,9,20,21</sup>. In marginal marine and shallow-water regions, analyses of mollusc and ostracod assemblages<sup>22–25</sup> provided additional geographic coverage for the dataset.

Wherever possible, independent palaeotemperature methodologies using numerous fossil groups were employed to confirm initial palaeoenvironmental estimates<sup>7</sup>. We included SST estimates derived from Mg:Ca ratios in shallow-dwelling planktonic foraminifer shells as well as the unsaturation index of alkenones (ketones synthesized by haptophyte algae living near the ocean surface) found in raw sediment, both of which have been calibrated to SST.

<sup>1</sup>Eastern Geology and Palaeoclimate Science Center, US Geological Survey, Reston, Virginia 20192, USA, <sup>2</sup>School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK, <sup>3</sup>Climate Change Group, British Geological Survey, Keyworth, Nottingham NG12 5GG, UK, <sup>4</sup>Atmosphere and Ocean Research Institute, University of Tokyo, Tokyo 277-8564, Japan, <sup>5</sup>Research Institute for Global Change, Japan Agency for Marine–Earth Science and Technology, Yokohama 236-0001, Japan, <sup>6</sup>Columbia University—NASA/GISS, New York 10025, USA, <sup>7</sup>National Center for Atmospheric Research, Boulder, Colorado 80305, USA, <sup>8</sup>School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK. \*e-mail: hdowsett@usgs.gov.

At present, 20% of our estimates are based on multiple proxies, with the addition of both Mg:Ca and alkenone palaeothermometry providing independent proxies for comparison at 14 sites each. In total, 24 of the 95 localities have non-faunal temperature proxies, helping to establish a more robust understanding of the palaeoenvironmental setting at these sites.

The present 95 marine locations form a global synthesis of an interval of warm and stable climate (relative to high-amplitude Pleistocene glacial–interglacial cycles) lying between the transition of marine isotope stages M2/M1 (3.264 Myr ago) and G21/G20 (3.025 Myr ago; ref. 26) in the middle part of the Gauss Polarity Chron. This interval ranges from C2An2r (Mammoth reversed polarity) to near the bottom of C2An1 (just above Kaena reversed polarity). This ~240 kyr time slab correlates in part to planktonic foraminiferal zones PL3 (*Sphaeroidinellopsis seminulina* highest-occurrence zone), PL4 (*Dentoglobigerina altispira* highest-occurrence zone) and PL5 (Atlantic) (*Globorotalia miocenica* highest-occurrence zone) or PL5 (Indo-Pacific) (*Globorotalia pseudomiocenica* highest-occurrence zone)<sup>27</sup>.

It occurs before the onset of high-amplitude oxygen-isotope oscillations, which represent a shift towards modern conditions (that is, Northern Hemisphere ice volume increased and glacial–interglacial variation intensified). Within the bounding positive  $\delta^{18}\text{O}$  excursions that mark glacial stages M2 and G20, and excepting glacial stage KM2 ~3.1 Myr ago, benthic foraminiferal oxygen-isotope values in this interval are equal to or isotopically lighter than those measured today<sup>26,28</sup>, making this interval easily distinguishable. Even so, a high degree of isotopic variability dominated by the 41 kyr period of Earth's obliquity is evident within the time slab<sup>28–30</sup>.

The establishment and duration of the time slab (~240 kyr) was originally dictated by limitations in correlating spatially distant data sites<sup>29</sup>. Until a sufficient number of data locations with orbitally tuned chronologies can be established, a dataset of the mean warm phase of climate is a compromise between good spatial coverage and correlation potential<sup>31</sup>. To derive an estimate of the mean warm phase of climate at each site, we evaluated SST estimates in closely spaced time series between 3.264 and 3.025 Myr ago and then averaged short-term warm events (see Supplementary Information) using a warm-peak averaging technique<sup>29,32,33</sup>. The resulting dataset comprises confidence-assessed mean annual SST anomalies for verification of climate-model temperature estimates at individual locations (Fig. 1).

### Assessing confidence

Measures of uncertainty or confidence have become an important focus of Intergovernmental Panel on Climate Change (IPCC) reports. In assessing confidence in the SST-verification data set, we follow guidance developed by the IPCC to define a new confidence metric for evaluating and communicating uncertainty that is applicable across a variety of proxies and proxy methodologies. As is always the case in palaeoclimate studies, some elements of our confidence assessment (for example, stationarity of environmental tolerances) are inherently non-quantifiable<sup>15,33</sup> and therefore cannot support a meaningful calculation of quantitative error. Instead, based on the principles outlined by the IPCC<sup>34</sup>, we have developed a semiquantitative measure of confidence accounting for quality of the age control of the samples at each site, number of samples at each site, fossil preservation and abundance, proxy method or technique used (quantitative or semiquantitative) and performance of the proxy method or technique used (see Supplementary Information). As set out by the IPCC, we must invoke expert subjectivity to arrive at a level of confidence, relying on our understanding of the specifics of individual sites and methods as well as the complexity of the dataset as a whole, while providing a traceable

account of the steps leading to the estimated confidence level. This type of confidence assessment can best be made with an understanding of both modern and palaeoceanographic settings and with an in-depth knowledge of the fossil material and methodologies used for analysis.

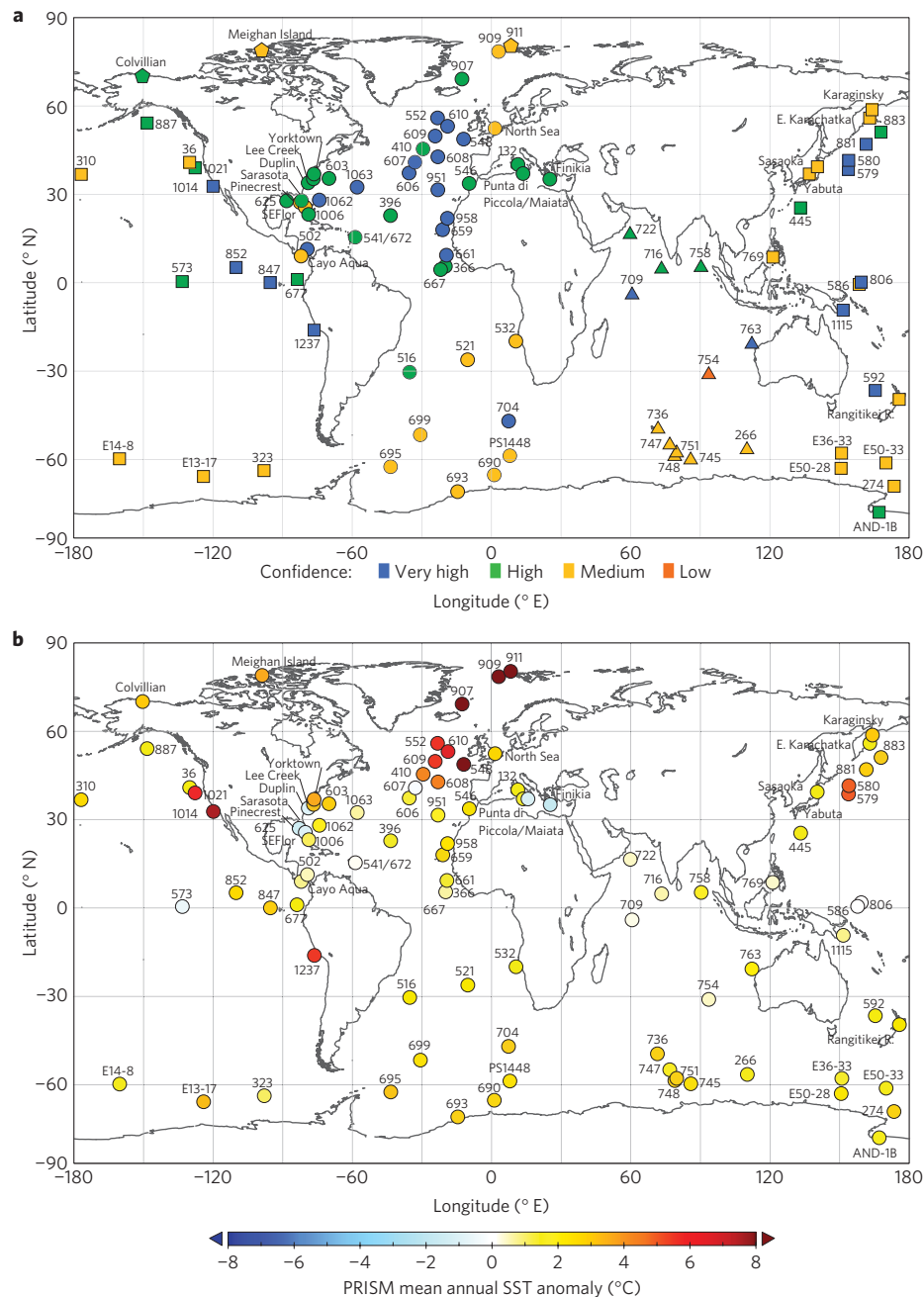
The chronological confidence is both a measure of quality of age control and how well the Pliocene Research Interpretation and Synoptic Mapping (PRISM) interval can be identified in a specific sedimentary sequence. An orbitally tuned isotopic record or a complete record of palaeomagnetic reversals through the Pliocene earns a higher confidence. Conversely, a small number of biochronologic events, fossil first or last appearances calibrated to the palaeomagnetic timescale in another region, earns lower confidence in the absence of local palaeomagnetic stratigraphy.

We also weight the assessment based on the number of samples analysed within the time slab, fossil preservation and abundance. In short, more samples equate to a better chance of characterizing temperature variability across the time slab. Similarly, more specimens equate to a higher probability that rare elements of the fauna or flora are represented. At least 300 specimens per sample are required to attain statistical confidence and avoid potential bias in the assemblage<sup>35</sup>. Even in samples with high abundances, preservation is important because dissolution can selectively remove less-robust species, altering the assemblage and thereby producing biased temperature results. In general, warm-water species of foraminifera are most fragile and their preferential dissolution drives temperature estimates cooler.

Methods of determining SST vary with fossil group and region and therefore are hard to compare. A quantitative estimate is considered better than a qualitative estimate simply because it is reproducible. The performance of the method is considered as part of the confidence assessment. For example, the performance of transfer functions is based on communality, the measure of how well the fossil assemblage can be described by the modern factors (assemblages). Performance of the modern analogue technique is based on the multivariate distance between the fossil assemblage and its closest modern analogues, and the agreement in temperature among those closest analogues<sup>36</sup>. Temperature estimates from geochemical methodologies (that is Mg:Ca and alkenone) are included or excluded based on sample characteristics that may limit analysis. For example, an alkenone-based SST estimate is not included if the total alkenone concentration of that sample falls below a threshold of reproducibility.

Different palaeotemperature proxies measure different aspects of temperature by sampling the marine environment at various spatial and temporal resolutions, further complicated by effects unique to each signal carrier and method. Therefore, our multiple-proxy analysis is done on a site-by-site basis, taking into account the full range of palaeoenvironmental information derived from a complete assessment of a fossil assemblage and allied geochemical proxies, to determine the overall quality of the temperature estimate<sup>37,38</sup>. Slight differences between multiple-proxy estimates from a single site strengthen the confidence of the overall site estimate, compared with an estimate from a single proxy.

**Very-high-confidence (VHC) example.** DSDP site 552 in the North Atlantic is among the group of sites attaining the highest confidence level because every piece of information, proxy and method is available in excellent condition and in good agreement. Chronology at this site is based on palaeomagnetism<sup>39</sup>, with all magnetochrons recorded in the sediment core, key biostratigraphic events and a high-resolution oxygen-isotope record clearly indicating the mid-Piacenzian interval<sup>40</sup>. Planktonic foraminifera in 22 samples are abundant and well preserved, and transfer-function communalities are very high (0.894–0.991; ref. 41). SST estimates are provided by multiple proxies with



**Figure 1 | Confidence estimates and temperature anomalies from PRISM dataset. a,** Distribution of verification dataset sites; symbols indicate region (Atlantic, bullets; Pacific, squares; Indian, triangles; Arctic, pentagons) and colours indicate level of confidence. **b,** Mean annual SST anomalies at 95 confidence-assessed Pliocene localities; see Supplementary Table S1.

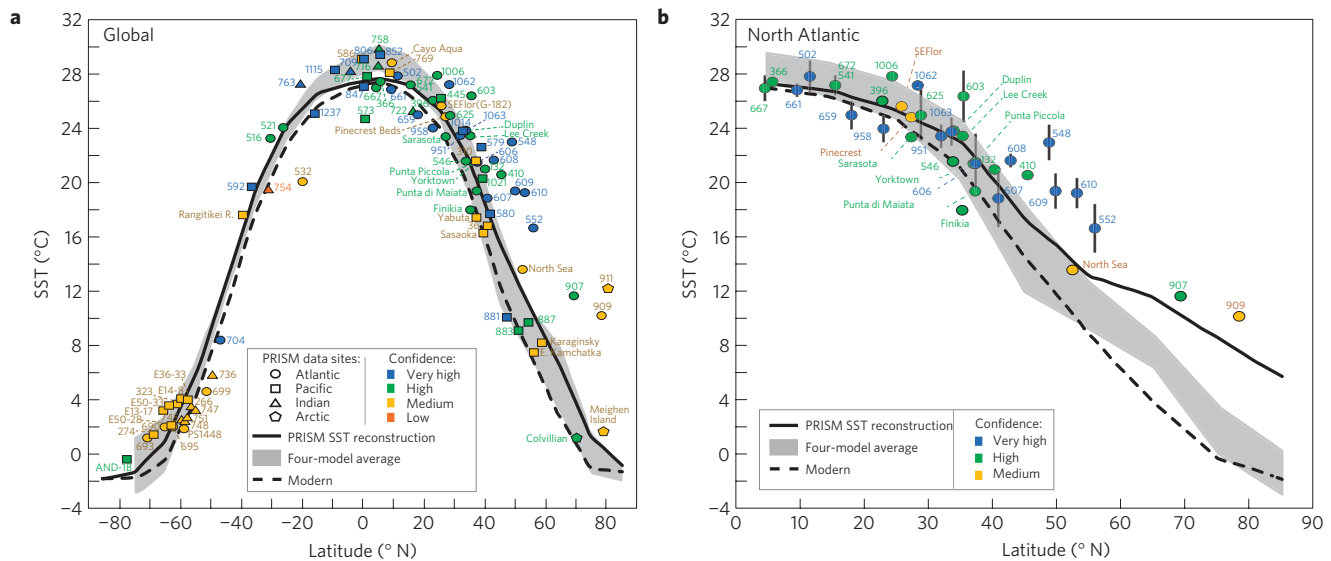
Mg:Ca and alkenone-derived estimates, indicating mean annual conditions, bracketed by faunal cold- and warm-season estimates<sup>7</sup>.

**High-confidence example.** At ODP site 677 in the eastern equatorial Pacific, alkenone-derived SST estimates from 28 samples fall between faunal based cold- and warm-season estimates from 16 samples. The modern analogue technique performs well on the foraminifer assemblages, returning multivariate distances between fossil and modern assemblages of less than 0.28. However, chronology at this site is based solely on integrated biochronology<sup>42</sup>, and is therefore less confidently resolved than at site 552.

**Medium-confidence example.** SST estimates from site E36-33 in the Southern Ocean are based on diatom assemblages and their

relationship to modern assemblages and summer temperatures near the Antarctic polar front. It is because this palaeothermometry method is semiquantitative that this site earns a medium, rather than high, confidence rating. The diatoms from this site are abundant with good to excellent preservation and the chronology is based on a fully resolved record of palaeomagnetic reversals augmented by five key biostratigraphic events<sup>43</sup>.

**Low-confidence example.** A single site in the SST-verification data set (ODP site 754 in the Indian Ocean) is categorized as low confidence and is included in the verification dataset solely for the purposes of comparison with higher levels of confidence. At site 754, chronology is based on only two biostratigraphic events<sup>44</sup>.



**Figure 2 | Data and model mean annual temperature profiles.** **a**, Verification dataset SST estimates superimposed on zonally averaged multimodel mean annual SST from four climate-model simulations shown as grey band with width equal to  $\pm 2\sigma$ . Symbols indicate ocean basin; colours indicate level of confidence. Zonally averaged modern mean annual SST (ref. 46) shown as dashed line, PRISM3 reconstruction<sup>3,6</sup> shown as solid line. **b**, same as in **a** except North Atlantic region only, with  $\pm 2\sigma$  SST variability bars on deep-sea sites with better than 22 kyr sample resolution and multiple warm peaks in the SST-estimate time series.

The planktonic foraminifera from 30 samples are abundant and well preserved. When assemblage data were applied to a newly developed Indian Ocean transfer function, however, the resulting communalities ranged from 0.03 to 0.20. This very poor performance by the transfer function indicates a situation for which there is no analogue and the SST estimates provided by the transfer function command little confidence.

## Results

Twenty-seven of the 95 localities were ranked with VHC and 32 with high confidence. All ocean basins contain sites with confidence levels ranging from VHC to medium confidence. The regional distribution of confidence in Pliocene SST estimates is shown in Fig. 1a.

The highest percentage of VHC sites is found in the North Atlantic Ocean where they are confined to latitudes below 60° N. VHC sites form a transect from the Caribbean Sea (DSDP site 502) to the northeast Atlantic (DSDP sites 548, 552 and ODP site 610). These VHC sites illustrate an ever-increasing temperature anomaly with increasing latitude (Fig. 1b). This warming trend is extended further north to the Arctic Ocean by including high-confidence site 907 and medium-confidence sites 909 and 911.

In the North Pacific Ocean, a combination of VHC and high-confidence sites documents both Pliocene warmth and the path of the Kuroshio current (Fig. 1). In the Southern Hemisphere, although no VHC sites are found south of ~45° S, a high-confidence site is found as far as 77° S in the Ross Sea. The circumpolar distribution of medium-confidence sites in the Southern Ocean (medium confidence because the reconstruction method is not quantitative) lends robust support to the poleward displacement of the Antarctic polar front zone and decreased sea-ice distribution during the mid-Piacenzian relative to present-day conditions<sup>19</sup>.

As our highest-confidence sites are not geographically restricted, they can serve as a starting point for formulating global hypotheses. For example, upwelling zones in the North, equatorial and South Pacific, and in the North Atlantic off North Africa, are represented by VHC sites showing warmer-than-modern SST. This indicates a commonality among mid-Piacenzian upwelling regions and argues for a system-wide phenomenon of warmer nutrient-rich upwelling waters across the globe.

## Comparison of the multimodel mean to Pliocene SST

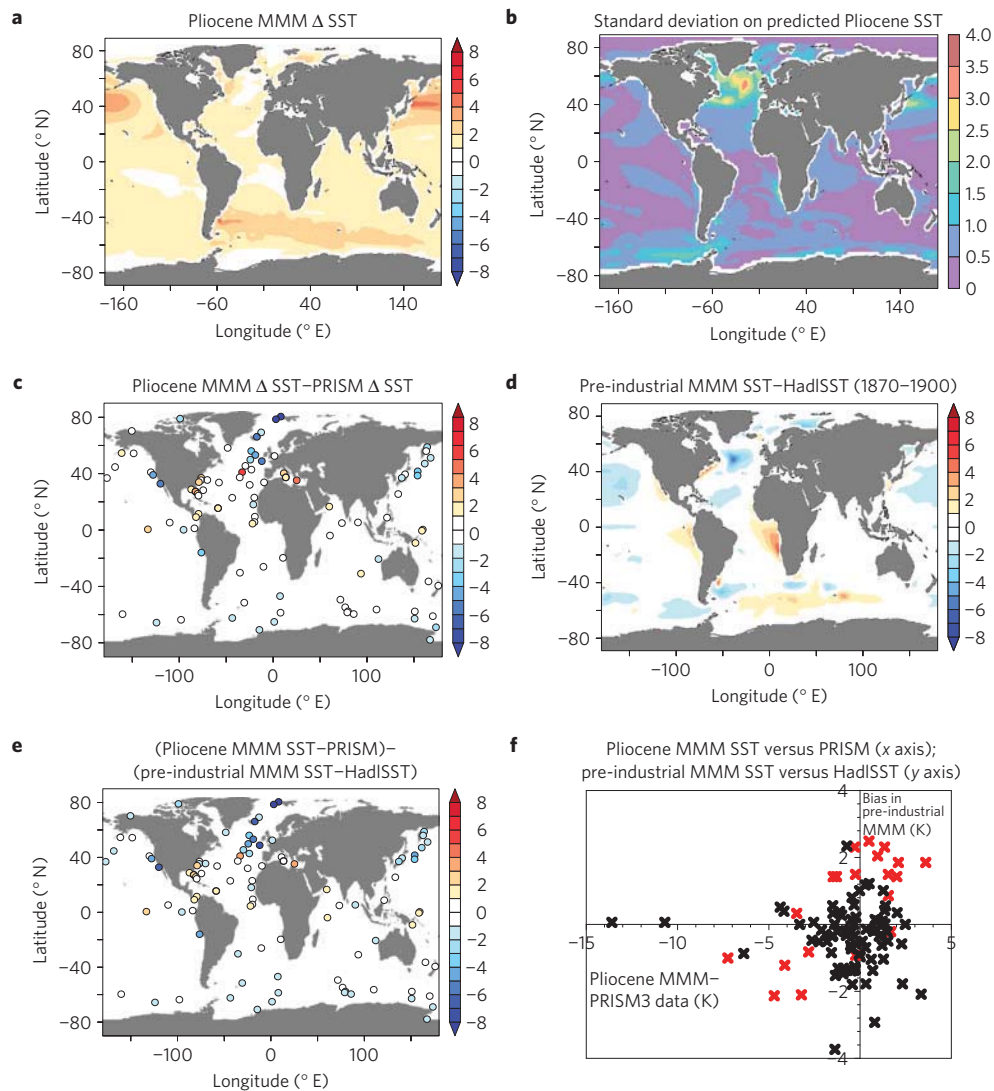
To assess the performance of Pliocene climate-model simulations, we compare zonally averaged multimodel mean annual surface temperature results from four leading climate models (the National Center for Atmospheric Research CCSM4; the Hadley Centre HadCM3; the University of Tokyo/Japan Agency for Marine–Earth Science and Technology MIROC4m and NASA Goddard Institute for Space Studies GISS-ER) all initialized and run using an identical experimental design and protocol<sup>45</sup> (see Supplementary Table S2) to the confidence-assessed SST-verification data (Fig. 2). The comparison also includes zonally averaged modern SST (ref. 46) and the globally interpolated PRISM3 SST reconstruction<sup>3</sup>.

The overall shape of the pole-to-pole temperature profile is nearly identical across modern, PRISM3 reconstruction and the multimodel mean representations of the surface ocean. The similarity of the zonally averaged PRISM3 reconstruction to the others is instructive as it emphasizes the commonalities between Pliocene, modern and near-future climates, relative to warm periods of the more distant past. The models are qualitatively in close agreement with each other in the Southern Hemisphere, but less so in the Northern Hemisphere.

A PRISM3 gridded Pliocene-SST reconstruction was created by extrapolating from and interpolating between a subset of the verification data (86 of the 95 sites)<sup>3</sup>; hence, individual site data representing local conditions should not be expected to fall squarely on the zonally averaged PRISM3 gradient. For example, site 532 registers significantly cooler mean annual SST than the overall reconstruction or the models because this site monitors upwelling in the Benguela current and is therefore cooler than other estimates at the same latitude. Most upwelling sites show similar cool offsets. The verification data points, particularly the VHC sites, should instead be used to identify gradients in the data not mirrored in the zonal gradients.

The models show good agreement with the data in the Southern Hemisphere and in the Southern Ocean. Likewise, the  $2\sigma$  variability about the multimodel mean brackets the verification data points in the North Pacific (Fig. 2a). Low-latitude warming away from upwelling regions has long been a feature of simulated Pliocene climate primarily owing to increased CO<sub>2</sub> used in these modelling





**Figure 3 | Data-model comparison.** **a**, Multimodel mean (MMM) annual SST anomaly ( $^{\circ}\text{C}$ ), Pliocene minus pre-industrial. **b**, Standard deviation of modelled SSTs. **c**, Anomaly as **a** minus PRISM3 SST anomaly. **d**, Difference between the multimodel mean pre-industrial SST prediction and HadISST dataset (averaged between 1870 and 1900) showing biases for the pre-industrial era. **e**, Same as **c** except removing pre-industrial biases shown in **d**. **f**, Scatter plot showing no correlation between model differences to HadISST and model differences to PRISM3 SSTs. Red crosses highlight North Atlantic data sites south of Iceland.

experiments. The VHC verification sites, although showing a range of low-latitude temperature, clearly support warmer tropics.

One clear difference between the model simulations and the verification data is immediately apparent. Neither the multimodel mean nor outputs from any of the individual models (Supplementary Figs S2 and S3) capture the amount of warming shown by the VHC verification data in the North Atlantic (Figs 1 and 2b). These sites clearly show increased anomalies with increasing latitude, probably amplified by reduced sea ice in the Northern Hemisphere and associated positive feedbacks, but the models do not capture the decreasing thermal gradient documented by multiple proxies even when variability in the Pliocene estimates is taken into account (Fig. 2b). This may be linked to differences in the position of the Gulf Stream–North Atlantic Drift in the models compared with the geological reconstruction. The models show a high degree of variation in the North Atlantic south of Iceland ( $\sim 10^{\circ}\text{C}$  in the mean annual SST at  $2\sigma$ ; see also Supplementary Fig. S3).

Previous analyses of model performance have identified limitations in the ability to reproduce the pre-industrial-era<sup>1</sup> SST. In this Pliocene comparison, however, the observed discrepancy in

the North Atlantic is larger and of a different nature than the error observed in pre-industrial simulations carried out by the same models. An analysis of model performance in simulating reconstructed Pliocene versus observed pre-industrial North Atlantic SSTs indicates that the discrepancy noted in our Pliocene analysis is probably not solely associated with errors in the simulated pre-industrial state, but instead may provide new insights into the predictive abilities of these models, above and beyond what is already known (Fig. 3).

### Implications and outlook

The IPCC has traditionally focused on multimodel mean climate scenarios to depict future climate conditions and has developed protocols to define confidence in those projections<sup>1</sup>. We have taken the same concepts and adapted them to a comprehensive set of SST estimates based on multiple proxies from the last sustained period of global warmth similar to that projected for the end of the present century. The resulting confidence-assessed verification SST dataset represents the highest concentration of data focused on a period significantly warmer than the pre-industrial period

and offers a test to Pliocene climate simulations. More notably, the verification dataset can be used to assess model behaviour regionally and at specific locations. Regional or site-specific levels of confidence provided by the verification dataset can be used as guidance for climate-simulation improvement.

The verification dataset is in many ways similar to multimodel mean climate scenarios of the same time period produced by four leading climate models. However, the confidence-assessed verification data show warming in the North Atlantic and Arctic oceans beyond what is simulated by the climate models used, implying more warmth from ocean heat transport than the models reproduce. Furthermore, the verification dataset (and faunal assemblage and geochemical data it is derived from) recognizes a system-wide phenomenon of warmer nutrient-rich upwelling regions during the mid-Piacenzian that is not present in the multimodel mean.

Although this analysis has identified data/model discord in the North Atlantic, we recognize inherent uncertainties in modelling Pliocene SSTs. For example, at present we are unable to constrain with certainty a number of critical forcing mechanisms and boundary conditions that climate models require to simulate Pliocene SSTs with the same degree of skill that they are able to simulate modern SSTs (for example, atmospheric CO<sub>2</sub> concentrations as well as other trace gases such as CH<sub>4</sub>, aerosol and dust loading, and orbital forcing). Reducing uncertainties wherever possible and thereby improving our skill in simulating regional characteristics of Pliocene climate is of prime importance to our understanding of this potential window into late twenty-first century climate. We consider our characterization of Pliocene SST data in terms of confidence levels a critical step in reducing uncertainties in the ground-truth data.

Received 29 November 2011; accepted 15 February 2012;  
published online 18 March 2012

## References

- Randall, D. A. *et al.* in *IPCC Climate Change 2007: The Physical Science Basis* (eds Solomon, D. *et al.*) 589–662 (Cambridge Univ. Press, 2007).
- Jansen, E. *et al.* in *IPCC Climate Change 2007: The Physical Science Basis* (eds Solomon, D. *et al.*) 433–497 (Cambridge Univ. Press, 2007).
- Dowsett, H. *et al.* The PRISM3D paleoenvironmental reconstruction. *Stratigraphy* **7**, 123–139 (2010).
- Dowsett, H. J., Barron, J. & Poore, H. R. Middle Pliocene sea surface temperatures: A global reconstruction. *Mar. Micropaleontol.* **27**, 13–25 (1996).
- Dowsett, H. J., Chandler, M. A., Cronin, T. M. & Dwyer, G. S. Middle Pliocene sea surface temperature variability. *Paleoceanography* **20**, 1–8 (2005).
- Dowsett, H. J., Robinson, M. M. & Foley, K. M. Pliocene three-dimensional global ocean temperature reconstruction. *Clim. Past* **5**, 769–783 (2009).
- Robinson, M. M., Dowsett, H. J., Dwyer, G. S. & Lawrence, K. T. Reevaluation of mid-Pliocene North Atlantic sea surface temperatures. *Paleoceanography* **23**, PA3213 (2008).
- Cronin, T. M. *et al.* Microfaunal evidence for elevated Pliocene temperatures in the Arctic Ocean. *Paleoceanography* **8**, 161–173 (1993).
- Barron, J. A. in *Proc ODP, Sci Results* Vol. 145 (eds Rea, D. K., Basov, I. A., Scholl, D. W. & Allan, J. F.) 43–53 (Ocean Drilling Program, College Station, 1995).
- Johnson, A. L. A. *et al.* Comparative sclerochronology of modern and mid-Pliocene (c. 3.5 Ma) *Aequipecten opercularis* (Mollusca, Bivalvia): An insight into past and future climate change in the north-east Atlantic region. *Palaeogeogr. Palaeoclimatol.* **284**, 164–179 (2009).
- Knowles, T., Taylor, P. D., Williams, M., Haywood, A. M. & Okamura, B. Pliocene seasonality across the North Atlantic inferred from cheilostome bryozoans. *Palaeogeogr. Palaeoclimatol.* **277**, 226–235 (2009).
- Valentine, A., Johnson, A. L. A., Leng, M. J., Sloane, H. J. & Balson, P. S. Isotopic evidence of cool winter conditions in the mid-Piacenzian (Pliocene) of the southern North Sea Basin. *Palaeogeogr. Palaeoclimatol.* **309**, 9–16 (2011).
- Dowsett, H. J. The development of a long-range foraminifer transfer function and application to Late Pleistocene North Atlantic climatic extremes. *Paleoceanography* **6**, 259–273 (1991).
- Waelbroeck, C. *et al.* Improving past sea surface temperature estimates based on planktonic fossil faunas. *Paleoceanography* **13**, 272–283 (1998).
- Dowsett, H. J. & Robinson, M. M. Application of the modern analogue technique (MAT) of sea surface temperature estimation to middle Pliocene North Pacific planktic foraminifer assemblages. *Palaeontol. Electron.* **1**, [http://palaeo-electronica.org/1998\\_1/dowsett/text.pdf](http://palaeo-electronica.org/1998_1/dowsett/text.pdf) (1998).
- Dowsett, H. J. Faunal re-evaluation of mid-Pliocene conditions in the western equatorial Pacific. *Micropaleontology* **53**, 447–456 (2007).
- Hays, P. E., Pisias, N. G. & Roelofs, A. K. Paleoceanography of the eastern equatorial Pacific during the Pliocene: A high-resolution radiolarian study. *Paleoceanography* **4**, 57–73 (1989).
- Barron, J. A. Pliocene paleoclimatic interpretation of DSDP Site 580 (NW Pacific) using diatoms. *Mar. Micropaleontol.* **20**, 23–44 (1992).
- Barron, J. A. Diatom constraints on the position of the Antarctic Polar Front in the middle part of the Pliocene. *Mar. Micropaleontol.* **27**, 195–213 (1996).
- Barron, J. A. *et al.* in *Paleoclimate and Evolution with Emphasis on Human Origins* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 197–212 (Yale Univ. Press, 1995).
- Barron, J. A. Diatom constraints on sea surface temperatures and sea ice distribution during the middle part of the Pliocene. *USGS Open-File Rep.* **96-713**, 1–45 (1996).
- Cronin, T. M. & Dowsett, H. J. A quantitative micropaleontologic method for shallow marine paleoclimatology: Application to Pliocene deposits of the western North Atlantic Ocean. *Mar. Micropaleontol.* **16**, 117–147 (1990).
- Gladenkov, Y. B., Barinov, K. B., Basilian, A. E. & Cronin, T. M. Stratigraphy and paleoceanography of Pliocene deposits of Karaginsky Island, eastern Kamchatka, USSR. *Quat. Sci. Rev.* **10**, 239–245 (1991).
- Cronin, T. M. Pliocene shallow water paleoceanography of the North Atlantic Ocean based on marine ostracodes. *Quat. Sci. Rev.* **10**, 175–188 (1991).
- Ikeya, N. & Cronin, T. M. Quantitative analysis of Ostracoda and water masses around Japan: Application to Pliocene and Pleistocene paleoceanography. *Micropaleontology* **39**, 263–281 (1993).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, 1–17 (2005).
- Wade, B. S., Pearson, P. N., Berggren, W. A. & Pälike, H. Review and revision of Cenozoic tropical planktonic foraminiferal biostratigraphy and calibration to the geomagnetic polarity and astronomical time scale. *Earth Sci. Rev.* **104**, 111–142 (2011).
- Shackleton, N. J., Crowhurst, S., Hagelberg, T., Pisias, N. G. & Schneider, D. A. in *Proc ODP, Sci Results* Vol. 138 (eds Pisias, N. G., Mayer, L. A., Janecek, T. R., Palmer-Julson, A. & van Andel, T. H.) 73–101 (Ocean Drilling Program, College Station, 1995).
- Dowsett, H. J. & Poore, R. Z. Pliocene sea surface temperatures of the North Atlantic Ocean at 3.0 Ma. *Quat. Sci. Rev.* **10**, 189–204 (1991).
- Naish, T. *et al.* Obliquity-paced Pliocene West Antarctic ice sheet oscillations. *Nature* **458**, 322–328 (2009).
- Robinson, M. M. & Dowsett, H. J. ePRISM: A case study in multiple proxy and mixed temporal resolution integration. *Stratigraphy* **7**, 177–187 (2010).
- Dowsett, H. J. & Robinson, M. M. Stratigraphic framework for Pliocene paleoclimate reconstruction: The correlation conundrum. *Stratigraphy* **3**, 53–64 (2006).
- Dowsett, H. J. in *Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies* (eds Williams, M., Haywood, A. M., Gregory, J. & Schmidt, D. N.) 459–480 (Geological Society of London, 2007).
- Mastrandrea, M. D. *et al.* *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties* (IPCC, 2010).
- Ortiz, J. D. in *Encyclopedia of Quaternary Science* (eds Elias, S. A. *et al.*) 1692–1699 (Elsevier, 2007).
- Hutson, W. H. The Agulhas Current during the Late Pleistocene: Analysis of modern faunal analogs. *Science* **207**, 64–66 (1980).
- Dowsett, H. J., Robinson, M. M., Stoll, D. K. & Foley, K. M. Mid-Piacenzian mean annual sea surface temperature analysis for data-model comparisons. *Stratigraphy* **7**, 189–198 (2010).
- Dowsett, H. J. *et al.* Sea surface temperatures of the mid-Piacenzian Warm Period: A comparison of PRISM3 and HadCM3. *Palaeogeogr. Palaeoclimatol.* **309**, 83–91 (2011).
- Zimmerman, H. B. *et al.* in *Init Rpts DSDP* Vol. 81 (eds Roberts, D. G. *et al.*) 861–875 (US Govt. Printing Office, 1984).
- Curry, W. B. & Miller, K. G. in *Proc ODP, Sci Results* Vol. 108 (eds Ruddiman, W. *et al.*) 157–166 (Ocean Drilling Program, 1989).
- Dowsett, H. J. & Poore, R. Z. A new planktic foraminifer transfer function for estimating Pliocene–Holocene paleoceanographic conditions in the North Atlantic. *Mar. Micropaleontol.* **16**, 1–23 (1990).
- Jenkins, D. G. & Houghton, S. D. *et al.* in *Proc ODP, Sci Results* Vol. 111 (ed. Becker, K.) 289–293 (Ocean Drilling Program, College Station, 1989).
- Keany, J. Paleoclimatic trends in early and middle Pliocene deep-sea sediments of the Antarctic. *Mar. Micropaleontol.* **3**, 35–49 (1978).
- Peirce, J. *et al.* *Proc ODP, Init Rpts* 121, 191–236 (Ocean Drilling Program, College Station, 1989).
- Haywood, A. *et al.* Pliocene Model Intercomparison Project (PlioMIP): Experimental design and boundary conditions (Experiment 2). *Geosci. Model Dev.* **4**, 571–577 (2011).
- Reynolds, R. W. & Smith, T. M. A high-resolution global sea surface temperature climatology. *J. Clim.* **8**, 1571–1583 (1995).

## Acknowledgements

H.J.D., M.M.R., D.K.S. and K.M.F. acknowledge the continued support of the US Geological Survey Climate and Land Use Change Research and Development Program; C.R.R. thanks the US Geological Survey Mendenhall Postdoctoral Fellowship Program; H.J.D., M.M.R. and M.A.C. thank the US Geological Survey John Wesley Powell Center for Analysis and Synthesis; M.A.C. acknowledges support from the NASA Climate Modelling Program. A.M.H. acknowledges financial support from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 278636. D.J.H. undertook his contribution as part of a Leverhulme Early Career Fellowship (ECF-2011-2015), co-financially supported by the British Geological Survey and National Centre for Atmospheric Science; D.J.L. and F.J.B. acknowledge the Natural Environment Research Council grant NE/H006273/1. B.L.O.-B. and N.A.R. recognize the National Center for Atmospheric Research is sponsored by the US National Science Foundation and computing resources were provided by the Climate Simulation Laboratory at the National Center for Atmospheric Research's Computational and Information Systems Laboratory sponsored by the National Science Foundation and other agencies. This research used samples

and/or data provided by the Integrated Ocean Drilling Program. This is a product of the PRISM Project and the Pliocene Model Intercomparison Project.

## Author contributions

H.J.D., M.M.R., K.M.F. and D.K.S. developed the mid-Piacenzian SST-verification data set and carried out the data-model comparison. H.J.D., M.M.R. and C.R.R. designed and completed the confidence assessments. D.J.L., F.J.B., A.M.H., M.A.C., W.-L.C., A.A.-O., B.L.O.-B. and N.A.R. carried out the general circulation model simulations and analysis. H.J.D., M.M.R., A.M.H. and B.L.O.-B. were involved in the study design. A.M.H., D.J.H. and A.M.D. carried out the comparison of model performance for the pre-industrial versus Pliocene. All authors discussed the results and commented on the manuscript.

## Additional information

The authors declare no competing financial interests. Supplementary information accompanies this paper on [www.nature.com/natureclimatechange](http://www.nature.com/natureclimatechange). Reprints and permissions information is available online at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.J.D.