Data-model discord reveals challenges in reconstructing terrestrial warming of the Pliocene

Ulrich Salzmann[1*], Aisling M. Dolan[2], Alan M. Haywood[2], Wing-Le Chan[3], Daniel J. Hill[4], Jochen Voss[5], Ayako Abe-Ouchi[3,6], Bette Otto-Bliesner[7], Fran Bragg[8], Mark A. Chandler[9], Camille Contoux[10,11], Harry J. Dowsett[12], Anne Jost[11], Youichi Kamae[13], Gerrit Lohmann[14], Daniel J. Lunt[8], Steven J. Pickering[2], Matthew J. Pound[1], Gilles Ramstein[10], Nan A. Rosenbloom[7], Linda Sohl[9], Christian Stepanek[14], Hiroaki Ueda[13], Zhongshi Zhang[15,16]

[1] Department of Geography, Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, UK;

[2] School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds, UK;

[3] Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan;

[4] British Geological Survey, Keyworth, Nottingham, UK;

[5]School of Mathematics, University of Leeds, Leeds, UK;

[6] Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan;

[7] National Center for Atmospheric Research, Boulder, CO, USA;

[8] School of Geographical Sciences, University of Bristol, University Road, Bristol, UK;

[9] Columbia University - NASA/GISS, New York, NY, USA;

[10] LSCE/IPSL, CNRS-CEA-UVSQ, Saclay, France;

[11] Sisyphe, CNRS/UPMC Univ Paris 06, Paris, France;

[12] Eastern Geology and Paleoclimate Science Center, US Geological Survey, Reston, VA, USA;

[13] Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan;

[14] Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany;

[15]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China;

[16]Bjerknes Centre for Climate Research, Bergen, Norway

* corresponding author:

Email: Ulrich.Salzmann@northumbria.ac.uk

Tel: Tel. +44 (0)191 2273874

Uncertainties, bioclimatic ranges and temporal variability in proxy data and climate model outputs are often neglected in data-model comparisons designed to test the predictive ability of climate models, or differentiate between the performance of individual models within an ensemble. Here we use a global data set of confidence-assessed, proxy-based temperature estimates and biome reconstructions to assess the ability of eight models to simulate warm terrestrial climates of the Pliocene. The Late Pliocene, 3.6 to 2.6 million years ago, is an accessible geological interval to understand climate processes of a warmer world [1]. Here we show that model-predicted surface air temperatures reveal a substantial cold bias in the Northern Hemisphere. Particular strong data-model mismatches exist in Northern Russia where differences in mean annual temperatures reach 18 °C. Our model sensitivity tests identify insufficient temporal constraints hampering the accurate configuration of boundary conditions as an important factor impacting on data-model discrepancies. We conclude that in order to allow a more robust evaluation of the ability of current climate models to predict warm climates, future Pliocene data-model comparison studies must focus on orbitally defined time slices.

Our understanding of causes and consequences of global warming relies heavily on climate model simulations conducted under a variety of greenhouse gas emission scenarios [2]. Comparing model simulations of key warm periods in Earth history with contemporaneous geological proxy data is one approach to evaluate the ability of these models to simulate warm, high CO_2-climates which are unprecedented in the more recent past. Existing data-model comparisons (DMCs) demonstrate that climate models are generally able to reproduce past warm climates of the last 65 million years [3-5]. However, a common data-model mismatch in high-latitude temperature estimates suggests that many models seem to underestimate polar amplification [4-6]. This has led to an ongoing controversy about the accuracy of DMC-studies, which might have been biased by uncertainties in estimating temperatures from geological proxies. Recently published proxy-based temperature reconstructions, suggesting tropical-like

28 climates at southern high latitudes ca. 53 Myr ago [7], have intensified the ongoing debate and
29 highlight the need of a systematic assessment in DMC-studies of uncertainty ranges and
30 variability.

31 Here we compare Late Pliocene mean annual surface air temperature (SAT) estimates derived
32 from 45 palaeobotanical sites with simulations of eight fully coupled ocean-atmosphere climate
33 models (Fig. 1). All models have been initialised and run using an established experimental
34 design and protocol, assuming an atmospheric CO_2-concentration of 405 ppmv [8]. From a data
35 perspective our comparison includes the bioclimatic range, temporal variability, a new
36 qualitative assessment of confidence in temperature estimates, and biome reconstructions. For
37 the climate modelling, we consider inter-model differences in temperature predictions, as well
38 as sensitivity to varying boundary conditions such as orbital parameters, atmospheric CO_2-
39 concentrations and prescribed vegetation cover. The additional use of the BIOME4 classification
40 scheme [9, 10] allows a direct comparison between palaeobotanical data and model outputs and
41 therefore reduces potential complicating factors produced by different methods applied to
42 derive temperature estimates from the fossil record. We compare global biome predictions from
43 two selected models and with medium (HadCM3) and high (MIROC4m) climate sensitivity [11, 15].

44 **Multi-model variability and proxy data uncertainties**

45 The zonal averages from each of the eight models (Fig. 1) demonstrate how the models within
46 the ensemble have influenced the multi-model mean (MMM) zonal average. The MIROC4m and
47 COSMOS models show the strongest SAT response in the ensemble (Fig. 1; $\Delta SAT_{\_global}$ 3.46 –
48 3.60). In terms of the global annual mean SAT anomaly, CCSM4 and MRI show the weakest SAT
49 response to the implementation of Pliocene boundary conditions ($\Delta SAT_{\_global}$: 1.9 and 1.8
50 respectively). These results are generally consistent with the spread of climate sensitivity
51 values for the eight models (Supplementary Table S1) which is highest for MIROC4m (4.05 °C)
52 and COSMOS (4.1 °C) and between 2.7 °C and 3.2 °C for all other models. An energy balance

analysis shows that the models with the highest temperatures in the Northern Hemisphere high latitudes, and the highest overall climate sensitivity, exhibit enhanced Arctic feedbacks (Supplementary Fig. S2). Conversely the models (CCSM4 and MRI) with the least polar amplification have suppressed Arctic feedbacks and fail to reproduce a sea-ice free Arctic summer.

If available, we included for our data-model comparison for each palaeobotanical site (Supplementary Table S3) the 1) "bioclimatic range", which is the temperature range under which the reconstructed palaeoflora existed, as derived from modern plant assemblage relationships (Fig. 3d, Supplementary Fig. S1), and 2) "temporal variability", which is the maximum minus the minimum temperature over the time recorded in the geological proxy (Fig. 3e). We have also combined "temporal variability" and "bioclimatic range" (see Supplementary Table S5 and Supplementary Section 6). Following guidance developed by the IPCC [12] we qualitatively assessed the level of "confidence" for each data site.

The Late Pliocene climate shows a generally lower temporal variability than the glacial and interglacials of the Quaternary, although a period of significant global cooling is recorded after ca. 3 Ma. [13]. We have addressed the possibility that this may have introduced a potential cool bias on our dataset, by using two global biome reconstructions from 205 palaeobotanical sites. The biome datasets reconstruct the global vegetation during colder/drier and warmer/wetter intervals within the Late Pliocene stage (Fig. 2). Biomes which remained relatively stable are the warm temperate forests (e.g. in Europe and Asia) as well as the woodlands and forests of central and southern Africa. Palaeobotanical sites which show fluctuations between cooler/drier and warmer/wetter biome types are typically situated near palaeogeographic transitions of major vegetation zones. This includes the northernmost Arctic records from North America which change from boreal forests to tundra shrubs, many sites in western North America which fluctuate between open conifer woodland and temperate xerophytic shrubland, and most palaeorecords from Japan that indicate changes from warm-temperate to slightly colder

79 temperate deciduous forest biomes (Fig. 2). However, our assessment of biome temporal
80 variability demonstrates that the vegetation changes between 3.6-2.6 million years were in
81 many regions relatively minor compared to those of the Quaternary. The warm vegetation
82 dataset predominantly shows the best fit for the majority of model simulations presented here
83 (Supplementary Table S4).

84 **Comparison of Pliocene surface air temperatures from models and data**

85 The difference between model-simulated Pliocene and pre-industrial SATs shows increased
86 temperatures in both hemispheres. The warming is particularly acute at high latitudes, north of
87 the Arctic Circle and in regions of Antarctica where ice was removed in the climate model set-up
88 (Fig. 3a). Proxy-based Pliocene SAT anomalies display a similar trend for the temperate and
89 polar zones of the Northern Hemisphere (Fig. 3b). The few temperature estimates available for
90 the Southern Hemisphere show no consistent large-scale pattern whilst tropical temperatures
91 remained unchanged or experienced cooling during the Late Pliocene.

92 Point-by-point comparison of SAT anomalies indicate that the models do not sufficiently
93 weaken the SAT gradient from the tropics to the high-latitudes because they underestimate the
94 degree of SAT warming reconstructed by terrestrial proxies in the mid to high-latitudes of the
95 Northern Hemisphere (Fig. 3c). This is particularly seen in Eurasia where temperature
96 differences reach as much as 18°C. These temperature differences are derived from sites some
97 of which have been assessed as 'high' or 'very high' confidence, and there is no evidence to
98 assume a systematic bias in these estimates. The few temperature estimates available for the
99 tropical zone tentatively suggest that the underestimation of SATs in the high latitudes may be
100 accompanied by an overestimation in the low latitudes by 1- 6 °C (Fig. 3c). Recently identified
101 sea surface temperature discrepancies between model predictions and proxy-data estimates in
102 the North Atlantic[14] may be related to a cool bias in model predictions of European SATs
103 identified in this study.

The impact of boundary conditions on the model fit to data

To further examine the potential causes of the identified data-model mismatches we explored the model sensitivity to imposed geological boundary and initial conditions by performing five additional simulations using HadCM3 and MIROC4m (Supplementary Table S2). Overall, the implementation of more recently developed boundary conditions [15, 16] (PRISM3 in Table 1, Supplementary Fig. S3) acts to reduce zonally averaged SATs in the Northern Hemisphere in comparison with the previously used data set (PRISM2 [17] in Table 1). This is likely to be a consequence of increased mountain height[19]. A DMC of ΔSATs for three selected palaeobotanical sites reveal that the implementation of PRISM3 boundary conditions results in a decrease of ΔSATs at the polar and temperate sites by ca. 3°C and 1°C, respectively, whilst at the tropical site ΔSATs increase slightly. In contrast to mean annual SATs, which indicate a greater data-model mismatch after the implementation of PRISM3 boundary conditions, the comparison of polar biomes using Kappa statistics indicate an improvement of model to data fit caused by a further northward extension of the boreal forests (Table 1). Biomes are integrators of climate change and their distribution is controlled by a range of additional climate parameters such as the length of growing season and annual rainfall. Whilst a smaller Greenland ice sheet in the new PRISM3 data set has only regional implications [18, 19], the introduction of a new vegetation cover and update of its physical parameters increased summer temperatures in the Northern Hemisphere resulting in a northward shift of boreal forests.

We also examined the sensitivity of HadCM3 and MIROC4m predictions to imposed orbital configuration and varying atmospheric CO_2-concentrations[20]. Our experiments suggest that orbital parameters which deliver the maximum degree of warmth at 65° N in summer improve the data-model fit for SATs at the high latitudes of the Northern Hemisphere (Table 1), where many of the mismatches occur for sites characterised as 'high' confidence. A further improvement can be achieved by increasing CO_2 concentration in the atmosphere to 450 ppmv, the maximum level suggested by proxy-CO_2 data [8]. However, these changes in boundary

condition increase SATs at high as well as low latitudes resulting in a generally weaker data to model fit in the tropics in all models. The comparison of biome distribution shows that the best data-model fit is achieved under PRISM3 standard Pliocene boundary conditions in the HadCM3 and MIROC4m sensitivity experiments.

Conclusion and Implication

Our DMC identifies a cold bias in models in the Northern Hemisphere (particularly north of 30 °N) demonstrating that, given the boundary conditions we have applied, none of the models used in this study reproduce the magnitude of northern hemisphere high-latitude warming exhibited in this proxy-dataset (Fig. 3c). A tentative data model mismatch may also be evident in the tropical zone where modelled Pliocene SATs appear to be too high, however this is limited by data availability.

Before drawing any conclusions with regard to the ability of climate models to reproduce the Pliocene, the potential causes of mismatches between palaeo-data and –models must be fully understood. Our DMC has identified regions (i.e. northern Russia, North Alaska and northeast Australia) where our data model mismatch is apparent when considering bioclimatic range (Fig. 3d), temporal variability (Fig. 3e) and even a combination of both factors (Fig. 3f). We have qualitatively assessed these sites as 'medium' to 'very high confidence'. The underlying reasons for these large DMC mismatches are still unknown.

Our DMC results also demonstrate that at many localities the spread in model-predicted SAT anomalies, from the model ensemble, is sufficiently large to cause an overlap with the available range of proxy-derived SAT anomalies (highlighted by the purple squares in Fig. 3g). At these localities it is possible to use the proxy data to differentiate between the performance of individual models. However, at this time such a differentiation would be difficult to complete accurately given the only partially quantified factors of bioclimatic range and temporal variability. For example, Figure 3e shows that out of the 14 data localities for which estimates of

155 temporal variability are available, the MMM falls within the proxy-derived temperature range
156 43% of the time. The importance of temporal variability is also highlighted by comparison of
157 biome simulations. Kappa scores (Table 1) for biome reconstructions indicate improved data to
158 model fit with different orbital configurations is possible in a specific latitudinal band (e.g.
159 polar), but often at the cost of a degraded model/data fit in other bands (i.e. the tropics and/or
160 temperate). This is in line with the premise that the proxy data is not consistent with a single
161 orbital configuration, rather the proxy records are representing multiple intervals of time
162 through the mid-Pliocene, which were characterised by different orbital forcing and potentially
163 atmospheric CO_2-concentration as well.

164 Our analysis shows that a tightly constrained time slice proxy reconstruction is necessary in
165 order to reduce the importance of temporal variability in determining any assessment of the
166 relative performance of models within an ensemble. A selection of a time slice [21] has the added
167 advantage of making it possible to provide climate models with values for orbital forcing that
168 are known to be consistent with a given proxy data reconstruction.

169 **Material and Methods**

170 All climate modelling groups followed the Pliocene Model Intercomparison Project (PlioMIP)
171 Experiment 2 design [15], using terrestrial boundary conditions from the "Pliocene Research,
172 Interpretation and Synoptic Mapping" project PRISM [16]. Model specific details on the
173 implementation of the mid-Pliocene boundary conditions, model spin-up and data-model
174 comparison methods can be found in the Supplementary Information. We also incorporate
175 additional Pliocene HadCM3 and MIROC4m sensitivity simulations that explore the effect of
176 prescribing different boundary conditions, such as the effect of adjusting orbital parameters
177 and/or the concentration of atmospheric carbon dioxide (CO_2), which provides either maximum
178 hemispheric forcing and/or maximum potential CO_2-forcing.

179 Two independent methods for data-model comparison have been employed. In the first

approach 45 quantitative temperature estimates from palaeobotanical proxies were used and compared to model predictions (Supplementary Table S3). Temperature estimates have been taken from literature or derived using the Coexistence Approach [22] and Palaeoflora Database [23]. If available, we took the author's interpretation from the original research paper to define climatic ranges for each temperature estimate. The confidence of each data set has been qualitatively assessed in regard to fossil preservation, temperature estimate methods, age control and resolution. Our model data comparison is necessarily qualitative because of the qualitative nature of the confidence assessments. In the second approach (forward modelling) biome reconstructions for the Late Pliocene from 205 palaeobotanical sites (Fig. 2) were compared with outputs of a mechanistically-based biome model (BIOME4; see Supplementary Information) forced by the HadCM3 and MIROC4m sensitivity simulations. BIOME4 model outputs and data have been quantitatively compared using kappa statistics [24]. Further details of the methodology of data synthesis and biome reconstruction are outlined in the Supplementary Information.

Acknowledgements

Financial support was provided by grants to U.S. and A.M.H. from the Natural Environment Research Council, NERC (NE/I016287/1). A.M.D. and A.M.H. acknowledge 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.L. and F.J.B. acknowledge the NERC grant NE/H006273/1. U.S, A.M.D, H.J.D. and A.M.H thank the US Geological Survey John Wesley Powell Center for Analysis and Synthesis. W.-L.C and A.A.-O. acknowledge financial support from the Japan Society for the Promotion of Science and computing resources at the Earth Simulator Center, JAMSTEC. H.J.D. acknowledges the continued support of the US Geological Survey Climate and Land Use Change Research and Development Program; D.J.H. acknowledges the Leverhulme Trust for the award of an

Early Career Fellowship and the National Centre for Atmospheric Research and the British Geological Survey for financial support. G. L. received funding through the Helmholtz research programme PACES and the Helmholtz Climate Initiative REKLIM. C. S. acknowledges financial support from the Helmholtz Graduate School for Polar and Marine Research and from REKLIM. B.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.

Authors Contributions

U.S. synthesised the palaeobotanical proxy data and designed and completed the confidence assessments. U.S., M.J.P., J.V. and H.J.D. carried out the data-model comparisons. A.M.D. and A.M.H. carried out the comparisons of model performance and the BIOME4 simulations. W.-L. C. performed the additional sensitivity experiments using MIROC. D.J.H carried out the energy balance analysis. All other authors performed general circulation model simulations which contributed to the PlioMIP Project and 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. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to U.S.

References

1. Jansen, E., *et al*. in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. & Miller, H.L.) 433–497. (Cambridge Univ. Press, Cambridge, 2007).

2. Randall, D.A., *et al*. in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. & Miller, H.L.) 589–662. (Cambridge Univ. Press, Cambridge, 2007).

3. Masson-Delmotte, V., et al. Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints. *Clim. Dynam*. **26**, 513-529 (2006).

4. Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J. & Hill, D.J. A new global biome reconstruction and data-model comparison for the Middle Pliocene. *Global Ecol. Biogeogr*. **17**, 432-447 (2008).

5. Shellito, C. J., Lamarque, J.-F. & Sloan, L. C. Early Eocene Arctic climate sensitivity to pCO2 and basin geography. *Geophys. Res. Lett*. **36**, L09707 (2009).

6. Huber, M. & Caballero, R. The early Eocene equable climate problem revisited, *Clim. Past* **7**, 603-633 (2011).

7. Bijl, P.K., *et al*. Early Palaeogene temperature evolution of the southwest Pacific Ocean. *Nature* **461**, 776-779 (2009).

8. Pagani, M., Liu, Z., LaRiviere, J., Ravelo, A.C. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geosci.* **3**, 27–30 (2010).

9. Prentice, I.C., *et al*. A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* **19**, 117–134 (1992).

10. Kaplan, J. O., *et al.* Climate change and arctic ecosystems II: modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res. Atmos.* **108**, 8171 (2003).

11. Hansen, J. et al. in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 130-163 (American Geophysical Union, 1984).

12. Mastrandrea, M.D., *et al. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. Intergovernmental Panel on Climate Change (IPCC). Available at http://www.ipcc.ch (2010).

13. McKay, R., *et al.* Antarctic and Southern Ocean influences on Late Pliocene global cooling . *PNAS* **109**, 17, 6423-6428 (2012).

14. Dowsett, H.J., *et al*. Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. *Nature Clim. Change* **2**, 365-371 (2012).

15. Haywood, A.M., *et al.* Pliocene Model Intercomparison Project (PlioMIP): Experimental design and boundary conditions (Experiment2). *Geosci. Model Dev.* **4**, 571-577 (2011).

16. Dowsett, H., *et al*. The PRISM3D Paleoenvironmental Reconstruction. *Stratigraphy*, **7**, 123-139 (2010).

17. Dowsett, H.J. *et al*. Middle Pliocene paleoenvironmental reconstruction: PRISM2. *USGS Open File Report*, **99**-535(1999).

18. Lunt, D.J., Foster, G.L., Haywood, A.M. & Stone, E.J. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels. *Nature* **454**, 1102-1106 (2008).

19. Lunt, D.J., *et al.* On the causes of mid-Pliocene warmth and polar amplification. *Earth Planet. Sc. Lett.* **321-322**, 128-138 (2012).

20. Lisiecki, L.E. & Raymo, M.E. A Pliocene-Pleistocene stack of globally distributed benthic ^{18}O records. *Paleoceanography* **20**, PA1003 (2005).

21. Haywood, A.M., *et al*. On the identification of a Pliocene time slice for data-model comparison. *Phil.Trans.Roy.Soc.A*. (in press).

22. Mosbrugger, V. & Utescher, T. The coexistence approach – a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **134**, 61–86 (1997).

278 23. Utescher, T. Mosbrugger, V. Palaeoflora Database. Available at http://www.palaeoflora.de
279 (2010).

280 24. Salzmann, U., Haywood, A.M. & Lunt, D.J. The past is a guide to the future? Comparing Middle
281 Pliocene vegetation with predicted biome distributions for the twenty-first century. *Phil. Trans.*
282 *Roy. Soc. A.* **367**, 189–204 (2009).

283 25. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **2**, 37–46 (1960).

284 26. Legates, D.R., & Willmott, C.J. Mean seasonal and spatial variability in global surface air
285 temperature. *Theor. Appl. Climatol.* **41**, 11-21 (1990).

286 27. Harrison, S.P. & Prentice, I.C. Climate and CO_2 controls on global vegetation distribution at
287 the last glacial maximum: analysis based on palaeovegetation data, biome modelling and
288 palaeoclimate simulations. *Glob. Change Biol.* **9**, 983–1004 (2003).

289

290 **Figure Legends**

291 Figure 1: **Zonally averaged mean annual surface air temperatures (SAT) in °C**. Plotted
292 zonally averaged SATs from modern observations, the multi-model mean prediction for
293 the pre-industrial era and the Pliocene multi-model mean. Individual Pliocene zonal
294 means are shown for all eight models.

295 Figure 2: **Location of 208 palaeobotanical sites used for data-model comparisons**.
296 Rectangles indicate sites with sufficient resolution and dating control to reconstruct biomes for
297 cold/dry (upper square) and warm/wet (lower square) climate periods or cycles. Circle colour
298 indicates reconstructed biomes for other sites. Red square highlights 45 palaeo sites with
299 temperature estimates used in this study (see Supplementary Table S3 and S4).

300 Figure 3: **Data-model comparison of global temperature estimates**. a) multi-model mean
301 (MMM) SAT anomaly (°C) (Pliocene minus pre-industrial). 3b) proxy-based Pliocene

SAT anomalies (Pliocene absolute SATs minus observed SATs from Legates and Wilmott[26]). 3c) difference between MMM SAT anomaly and proxy-based Pliocene SAT anomalies. 3d) the degree of remaining data-model mismatch, when the known bioclimatic range, 3e) temporal variability and 3f) the climatic range are taken into account. Green circles (3d-3f) indicate where the MMM sits within the range of data. 3g) Purple squares show where at least one model in the ensemble fits within the available range of data at each site (see Supplementary Information).

Tables

Table 1: **Data-model comparison for BIOME4 HadCM3 (PRISM2 and PRISM3) and MIROC (PRISM3)** . Table shows DMC results for annual SATs using regression analysis (R^2), differences in modelled and poxy-based SATs (model minus data) for three selected sites and difference between model and proxy-based mega-biome reconstructions [27] using a warm biased dataset. Kappa values [26] (κ) for global, polar, temperate tropical zones are ranked using a subjective assessment scale, whereby "0" means that the agreement is no better than would be expected by chance and "1" stands for a perfect match.

Model	R^2 (SAT)	Δ°C SAT Model - Data			Kappa Values			
		55.7°N	33.1°N	17.4 °S	Global	Polar	Temperate	Tropic
		L. Baikal	C. Kyushu	Butcher Cr.		> 60° N	60°-23.5°	<23.5° N/S
Pliocene HadCM3 PRISM2	**0.85**	-8.73	2.76	4.98	**0.31**	0.01	0.29	0.03
NH Max Orbit	0.87	-7.97	2.21	5.23	**0.28**	0.30	0.17	0.19
SH Max Orbit	0.85	-8.90	3.10	4.93	**0.30**	0.00	0.26	0.11
NH Max Orbit CO_2 450 ppm	0.87	-7.33	2.67	5.47	**0.31**	0.42	0.21	0.15
SH Max Orbit CO_2 450 ppm	0.85	-8.08	3.20	5.11	**0.33**	0.03	0.26	0.16
Pliocene HadCM3 PRISM3	**0.83**	-11.81	1.81	5.32	**0.30**	0.41	0.23	0.00
NH Max Orbit	0.84	-10.28	0.53	6.05	**0.25**	0.24	0.17	0.14
SH Max Orbit	0.83	-10.62	3.20	5.62	**0.30**	0.23	0.25	0.05
NH Max Orbit CO_2 450 ppm	0.84	-9.74	1.18	6.41	**0.22**	0.20	0.14	0.09
SH Max Orbit CO_2 450 ppm	0.83	-9.99	3.55	5.97	**0.30**	0.35	0.24	0.03
Pliocene MIROC PRISM3	**0.83**	-10.52	1.91	6.45	**0.30**	0.36	0.21	0.06
NH Max Orbit	0.84	-9.22	1.62	6.73	**0.26**	0.17	0.21	0.12
SH Max Orbit	0.83	-10.87	2.31	6.35	**0.29**	0.15	0.21	0.06
NH Max Orbit CO_2 450 ppm	0.84	-8.59	2.08	7.09	**0.25**	0.12	0.22	0.11
SH Max Orbit CO_2 450 ppm	0.83	-9.69	2.82	7.21	**0.29**	0.30	0.20	0.06

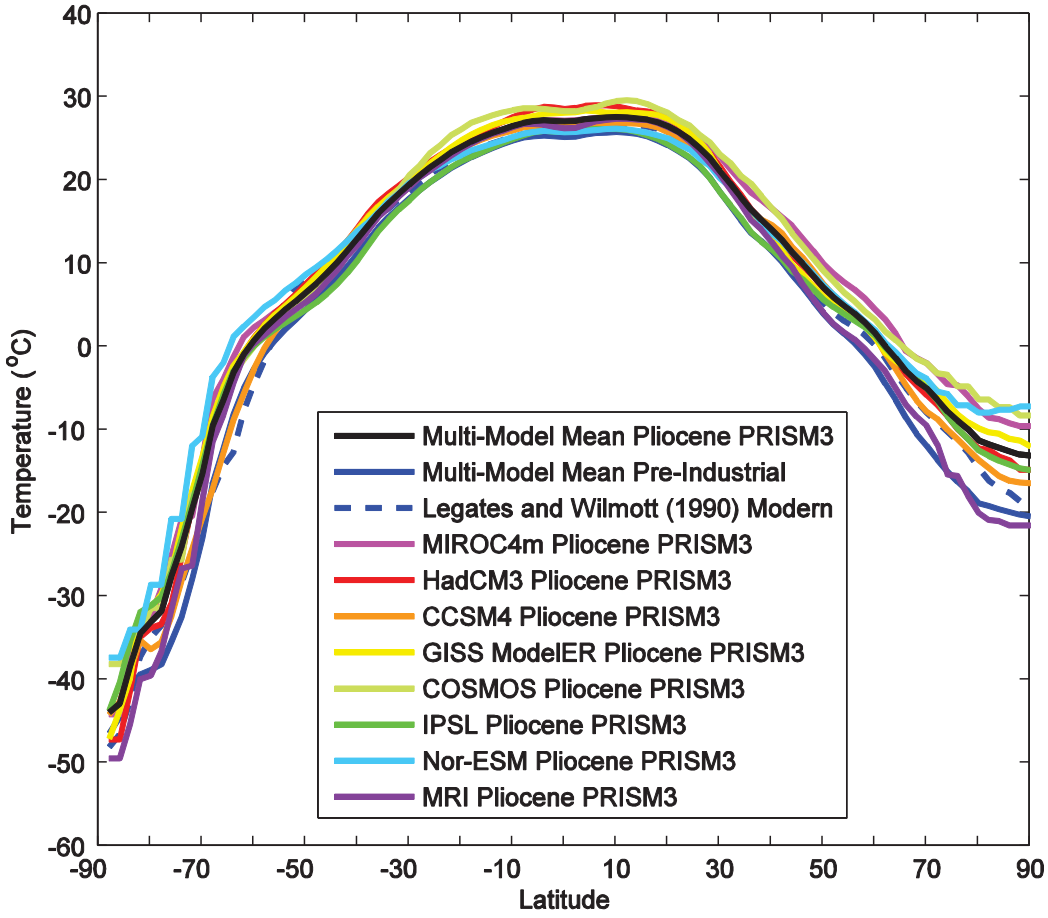


Figure 1: **Zonally averaged mean annual surface air temperatures (SAT) in °C**

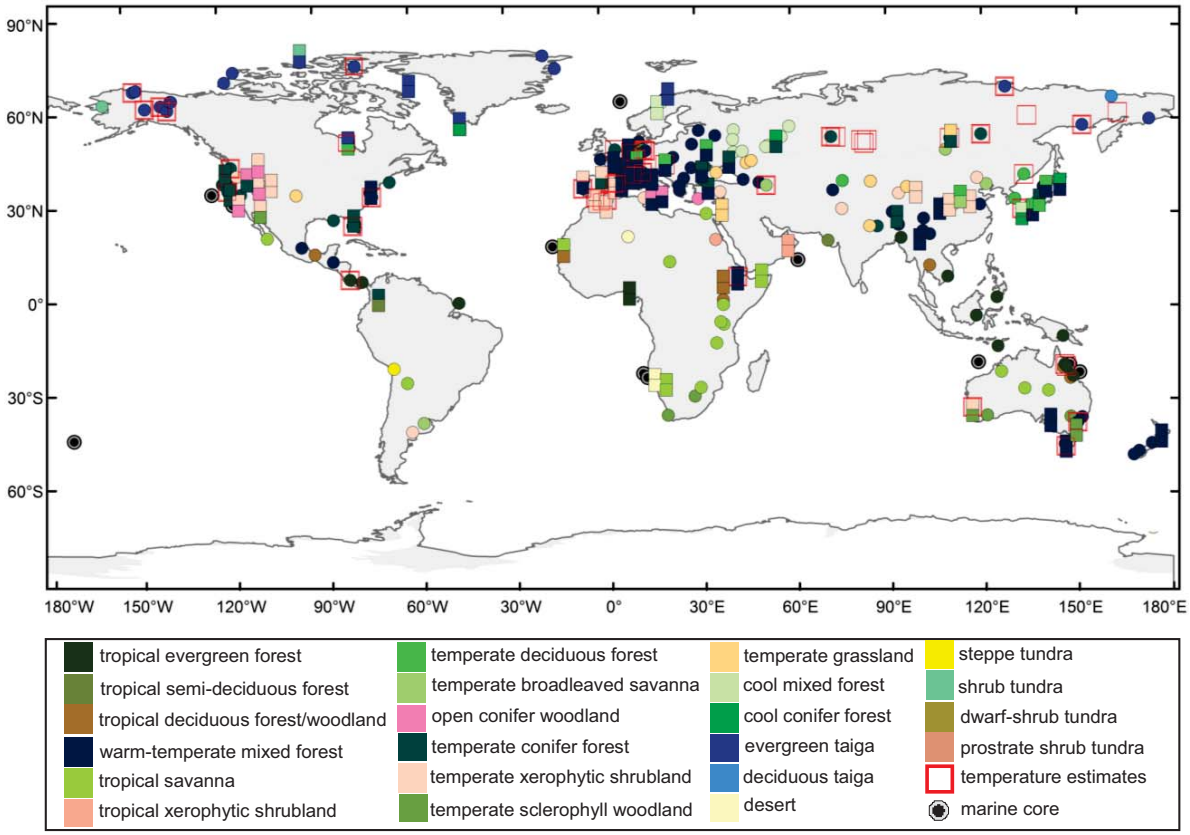


Figure 2: **Location of 208 palaeobotanical sites used for data-model comparisons**

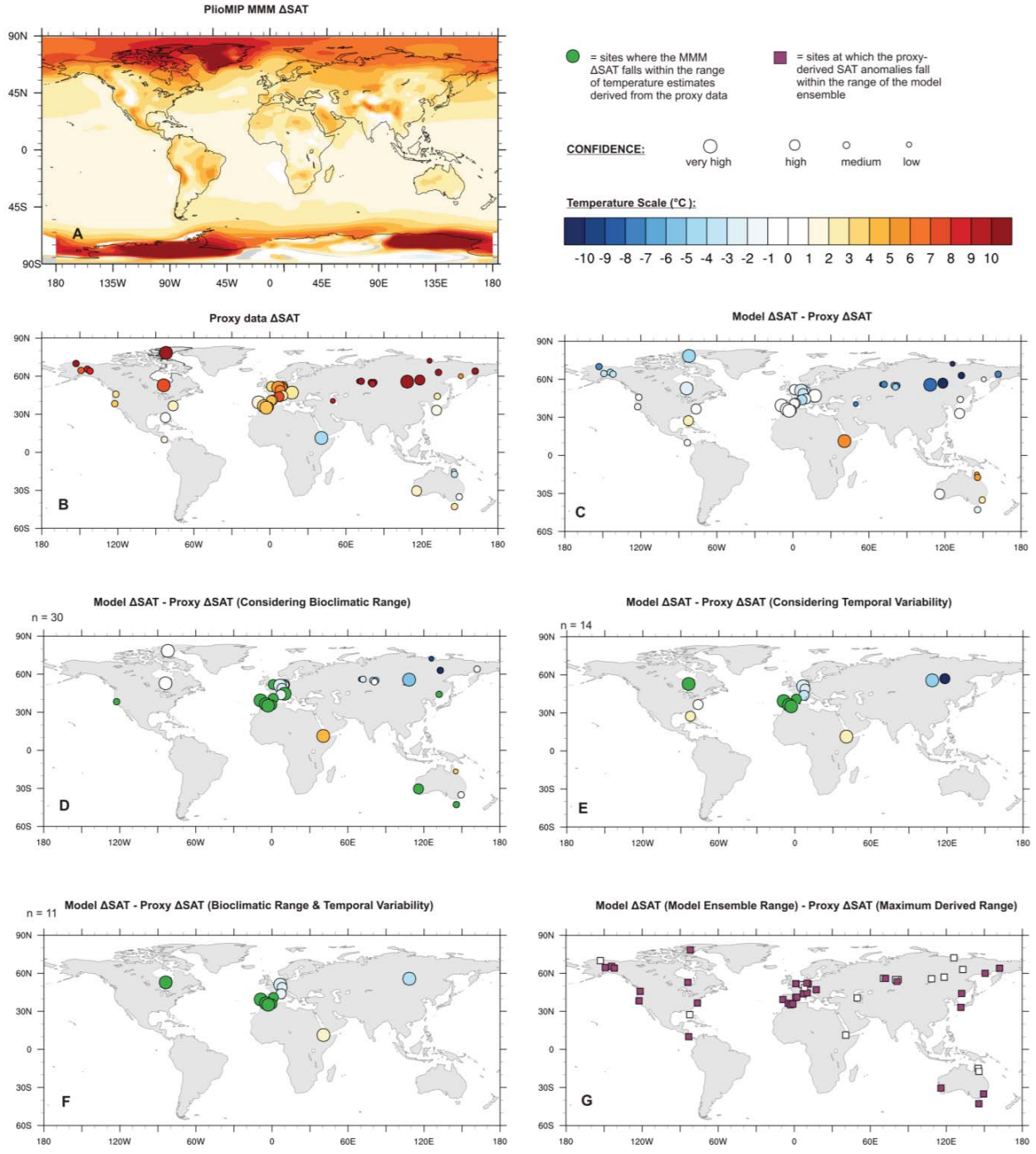


Figure 3: **Data-model comparison of global temperature estimates.**

Supporting Information for:

Data-model discord reveals challenges in reconstructing terrestrial warming of the Pliocene

Ulrich Salzmann[1*], Aisling M. Dolan[2], Alan M. Haywood[2], Wing-Le Chan[3], Daniel J. Hill[4], Jochen Voss[5], Ayako Abe-Ouchi[3,6], Bette Otto-Bliesner[7], Fran Bragg[8], Mark A. Chandler[9], Camille Contoux[10,11], Harry J. Dowsett[12], Anne Jost[11], Youichi Kamae[13], Gerrit Lohmann[14], Daniel J. Lunt[8], Steven J. Pickering[2], Matthew J. Pound[1], Gilles Ramstein[10], Nan A. Rosenbloom[7], Linda Sohl[9], Christian Stepanek[14], Hiroaki Ueda[13], Zhongshi Zhang[15,16]

[1] Department of Geography, Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, UK; [2] School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds, UK; [3] Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan; [4] British Geological Survey, Keyworth, Nottingham, UK; [5]School of Mathematics, University of Leeds, Leeds, UK; [6] Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan; [7] National Center for Atmospheric Research, Boulder, CO, USA; [8] School of Geographical Sciences, University of Bristol, University Road, Bristol, UK; [9] Columbia University - NASA/GISS, New York, NY, USA; [10] LSCE/IPSL, CNRS-CEA-UVSQ, Saclay, France; [11] Sisyphe, CNRS/UPMC Univ Paris 06, Paris, France; [12] Eastern Geology and Paleoclimate Science Center, US Geological Survey, Reston, VA, USA; [13] Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan; [14] Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany; [15]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; [16]Bjerknes Centre for Climate Research, Bergen, Norway

1. Description of climate models and outputs

Pliocene climate model runs (Supplementary Table S1) were initialised following the protocol of the Pliocene Modeling Intercomparison Project (PlioMIP), which incorporates the latest version of the US Geological Survey PRISM data set of boundary conditions (PRISM3) [1, 2] . Following this experimental design, atmospheric carbon dioxide was set to 405 ppmv in all of the climate model simulations. However, due to the lack of geological proxy data that can be used to robustly estimate methane

(CH_4) and other trace gasses, dust emissions and aerosols for the Late Pliocene, these boundary conditions were left unchanged from pre-industrial conditions. Therefore a caveat associated with this experimental design is that the incorrect specification of these trace gasses, aerosols and dust emissions for the late Pliocene, may over- (or under-) estimate the degree of data model discord.

Simulated Pliocene surface air temperatures (SATs) were derived from the final 30 years of each model run and re-gridded on a standard 2°×2° lat/long grid to facilitate the production of a multi-model mean (MMM). The MMM has been calculated by taking an average of the eight models (detailed in Supplementary Table S1); HadCM3, MIROC4m, CCSM4, GISS Model ER, COSMOS, IPSLCM5A, MRI-GCM 2.3, NorESM-L. No model weighting has been applied in the creation of the MMM. Model ensemble means for the Southern Hemisphere higher than 50 degrees south are not considered reliable because of inconsistencies in the prescribed land/sea mask between the eight models (Supplementary Table S1).

2. Description of BIOME4 vegetation model

BIOME4 is a development of the BIOME3 model [3]. It is a coupled carbon and water flux model, which predicts global steady state vegetation distribution, structure and biogeochemistry. BIOME4 was developed from the physiological constraints influencing the distribution of different plant functional types (PFT). Twelve plant functional (PFT) types, each with a very specific range of bioclimatic limits, are represented in BIOME4, ranging from Arctic to tropical flora. BIOME4 determines which of 28 biomes is most likely to occur in a grid square based on computed biogeochemical variables [3, 4].

The model is forced by long-term averages of monthly mean temperature, precipitation and sunshine. Atmospheric carbon dioxide concentrations must be specified and information on soil texture and depth must also be provided. To force the vegetation model, a standard anomaly method is employed (as in [5-7]). This takes into account the systematic error in the climate model relative to present-day observations of climate [8]. Owing to the lack of sufficient observational data over

Antarctica, the anomaly method cannot be employed in this region and there, the absolute values from the climate model are used to force BIOME4.

3. Sensitivity experiments

To examine the potential bias that orbital and CO_2 forcing could introduce, orbital parameters and CO_2 levels were changed in a suite of HadCM3 and MIROC4m sensitivity experiments (see Supplementary Table S2). Orbital parameters were set to deliver the maximum amount of incoming solar radiation at the top of the atmosphere at specific latitudes (either 65°N for the Northern Hemisphere or 80°S for the Southern Hemisphere) during summer within the Pliocene. The orbital configuration in the HadCM3 sensitivity experiments is thus representative of time points in the Pliocene where extremes in summer insolation were shown in the astronomical solution of Laskar et al. [9]. The values of eccentricity, precession and obliquity are shown for reference in Supplementary Table S2. In some sensitivity experiments, CO_2 levels were also increased to 450 ppmv (from the standard PRISM3 value of 405 ppmv) to assess the impact on our data-model comparison (DMC) of specifying a higher level of CO_2 (thus promoting more climate warming). Late Pliocene atmospheric CO_2 concentrations of 450 ppmv are the maximum level suggested by proxy-CO_2 data [10].

In contrast to the previous iteration of the boundary condition data set (PRISM2), PRISM3 which was used in the PlioMIP protocol, specifies the height of the western cordillera of both North and South America to be at approximately modern altitude (in contrast to a 50% reduction in PRISM2). There are also substantial changes in the vegetation scheme used and in the distribution of ice on Greenland and Antarctica. Experiments using HadCM3, where the model boundary conditions were changed from PRISM2 to PRISM3, highlight the combined effect of changes in land surface on climate diagnostics such as SAT and biome type.

4. Palaeobotanical dataset and temperature estimates

We used an updated version of the Late Pliocene (3.6-2.6 Ma) palaeobotanical TEVIS dataset[11]. The dataset integrates marine and terrestrial vegetation data derived from fossil pollen, leaves, wood and

palaeosol carbonate, whereby the authors' interpretation of palaeobotanical data, taken from the original research papers, was used and translated into the 28-type land-cover classification scheme of BIOME4 [12]. For the updated version presented here, we removed from the original dataset two re-dated sites from the Tjoernes Section in Iceland [12] and Sirius Group, Antarctica [13] and added data from 9 sites in Canada[14], Mexico [15], Russia [16, 17], Germany [18, 19], France [20], Portugal [21] and Turkey [22].

Uncertainties in estimating Pliocene palaeo-altitudes and locating the source of proxy data can significantly bias comparisons of temperature estimates. We therefore excluded terrestrial information from marine sites more than 250 km away from the mainland and estimates for sites with a modern altitude above 1000 m a.s.l. with the exception of one site for which temperatures have been adjusted to sea level by the original authors (Supplementary Table S3).

5. Assessing range and confidence of proxy data-based temperature estimates

Here, we discuss the various sources of uncertainty in our proxy temperature estimates. These are associated with temporal variability, bioclimatic range and additional unquantifiable uncertainties. We also discuss the process we use for assessing a qualitative indicator of our confidence in each data point.

5.1 Bioclimatic range and temporal variability

The bioclimatic range of a temperature estimate was derived (where available) from the fossil record using quantitative and semi-quantitative methods such as:

a) Climate Leaf Analysis Multivariate Program (CLAMP) [23] which uses the physiognomy of fossil leaves to determine past climates

b) Coexistence Approach (CA) [24], which uses the climatic requirements of the Nearest Living Relative (NLR) of fossil taxa to reconstruct the past climatic range (Supplementary Figure S1a)

c) Semi-quantitative methods using the climatic range of the nearest modern analogue vegetation distribution

d) Multi-proxy measurements that combine the above listed palaeobotanical methods with other palaeoclimate proxies, such as oxygen isotopes.

In our DMC study we either used published temperature estimates or, if not available, we applied the Coexistence Approach to generate temperature estimates from the fossil record (Supplementary Table S3). The quantitative and semi-quantitative temperature estimate techniques generally produce a temperature range rather than one absolute value (see Supplementary Figure S1a/b). Such ranges represent the climate interval in which all taxa of reconstructed palaeovegetation can co-exist. The bioclimatic range has a lower and upper limit of tolerance beyond which the nearest living relatives of the fossil assemblage cannot exist.

In addition to the bioclimatic range, we also included the temporal variability of a temperature estimate (Supplementary Figure S1b) in our DMC (where available). This takes into account the variability of the reconstructed vegetation in response to climate change over the time period covered by the fossil record (e.g. orbitally controlled cold and warm cycles). By including the temporal variability into our DMC we addressed uncertainties in age determination of the fossil record. However, it should be noted that depending on the quality and temporal resolution of the fossil record, temporal variability and bioclimatic ranges are not available for all sites. Generally, such sites have been assigned lower confidence levels (see also following sections).

5.2 Assessing the qualitative confidence of a temperature estimate

The quality of geological archives can strongly vary between sites and depends on a number of external factors, such as taphonomy, sedimentation rate, depositional environment and availability of datable material. These factors often impact on the quality of the proxy data and their use for environmental reconstruction. In order to address the variation in quality associated with our temperature estimates, we qualitatively assessed the level of confidence for each data point

(Supplementary Table S3). The criteria used to assign each data point to one of the four confidence levels (very high, high, medium and low) were: a) age control, b) resolution, c) fossil preservation and d) temperature estimate methods used. Temperature estimate methods and age control were the most important parameters impacting on the confidence levels of proxy data. Established quantitative methods have a generally higher confidence than semi-quantitative estimates using nearest living relatives. Temperature estimates with the highest confidence level are (i) typically well dated, using for example, radiometric or oxygen isotope dating methods, (ii) show an excellent fossil preservation that allows multi-proxy analyses, (iii) use quantitative methods such as CA or CLAMP and (iv) have a high resolution which allows the reconstruction of Late Pliocene mean annual temperature changes over several cold and warm cycles. Temperature estimates with a medium to low confidence typically have a poorer age control, which for example, is based on relative dating using similarities of fossil assemblages; (i) have a low diversity related to poor fossil preservation and (ii) use semi-quantitative temperature estimation techniques. The confidence level should be used in addition to the indicated temperature range as a guide to assess the robustness of a temperature estimate (see Fig. 3). Temperature estimates from sites with a high confidence level might be more accurate and reliable than other fossil records with lower confidence. However, we found a very good consistency between temperature estimates derived from high and lower confidence sites, in particular in high latitude Northeast Asia and North America (see Fig. 3b), indicating that impact of the level of confidence on DMC results appears to be rather low.

5.3 Additional unquantifiable uncertainties

Our DMC study focuses on two quantifiable sources of total uncertainty; bioclimatic range and temporal variability, because we consider that these are likely to have the largest impact on the total uncertainty of proxy temperature estimations. However, there are additional uncertainties in data and model derived temperature estimates which could potentially impact on the DMC results, but which are unquantifiable.

One additional uncertainty is caused by the fact that a model produces an average temperature over the size of one model grid cell, whereas data is collected from a single point in space. However, we consider this to have a rather small impact on the uncertainty ranges if compared to the temporal and bioclimatic ranges described above. Grid cell uncertainties gain importance in high altitude regions. However, most of our palaeobotanical sites are at altitudes below 350 m.a.s.l. (see Table S3) and we excluded any data points from our data set with a modern altitude above 1000 m. a.s.l. If we assume a) a Pliocene latitudinal temperature gradient of ca. 0.6 °C per degree in latitude[25] and b) that we further reduced uncertainties by using interpolation to the exact proxy location within the model grid cell and c) that the variation within the grid cell is also further reduced by the fact that our proxy based temperatures were derived from the predominant vegetation covering a wider area of the grid cell, then we have no reasons to assume that the uncertainty imposed by the size of a 2° x 2° grid cells exceeds 1°C.

Other additional uncertainties, such as methodological errors, might have a higher impact on the DMC results, although their uncertainties can hardly be quantified and vary between sites and methods applied. For example, CLAMP shows a tendency to produce for the Neogene temperature estimates that are generally colder (likely range 1-2°C) than CA[18]. The accuracy of the CA estimates, which is highest for temperature-related parameters and has been indicated to be usually in the range of 1–2 °C[26], has been recently questioned by a study[27], which identified remarkable inaccuracies in temperature reconstruction of warmer lowland and cooler vegetation at higher elevations. We therefore excluded any data points from our data set with a modern altitude above 1000 m. a.s.l.

6. Performing a data-model comparison on surface air temperature

To illustrate the differences between each model in the PlioMIP ensemble, zonal means were calculated by averaging the individual models across latitudinal bands (Fig. 1). The multi-model zonal mean is also shown on Fig. 1.

175 In order to compare MMM-predicted with proxy-based estimates of SAT both the modelled absolute
176 temperatures (Table 1) and the modelled temperature anomalies from pre-industrial have been used
177 (Fig. 3). Initially a visual comparison between the SAT anomalies was undertaken (Fig. 3c). Modelled
178 temperature anomalies from the nearest grid square to the location of the proxy-data site were
179 compared to data-based SATs (anomalies calculated to the Legates and Wilmott modern
180 observational dataset[28]) in order to determine the spatial differences or similarities between the
181 data-based and modelled anomalies.

182 The match between the absolute SAT data and the modelled SATs were evaluated in terms of
183 their correlation using a simple linear regression (r^2; Table 1). It should be noted that this
184 metric does not take into account any uncertainties in the proxies, and as such it should be
185 interpreted with caution. Furthermore, statistical comparisons of absolute data and modelled
186 SATs can be biased by the overriding effect of latitude on SAT, which can lead to a misleading level of
187 agreement between the models and data. Therefore, testing the commonality between the SAT
188 anomalies (mid-Pliocene minus modern/pre-industrial) produced by the model and the data provide
189 a fairer test of the model [29]. For all panels in Figure 3, the SAT anomalies of the proxy-data and model
190 have been used.

191 In Figures 3d and 3e, we show the impact on the DMC when we consider the proxy-derived
192 bioclimatic range (available at 30 sites) and temporal variability (available at 14 sites). We
193 demonstrate (using green circles) sites at which the MMM SAT anomaly falls within the
194 range associated with the proxy data. Where both ranges are available (11 sites), we
195 combine the bioclimatic range and temporal variability to create a total "climatic range" (r_{total}).
196 From any given fossil sample only a range of possible values of the SAT can be determined and
197 there is a deviation (X_{bc}) between true SAT and the center of the range of possible temperatures.
198 This deviation represents a form of measurement error and the magnitude of X_{bc} is described by
199 the bioclimatic range r_{bc}. Similarly, the mean temperature fluctuates over the time period covered

200 by the fossil record and at a given time there will be a deviation (X_{tv}) between the current mean
201 temperature and the long term average of the SAT. The magnitude of X_{tv} is described by the
202 temporal variability r_{tv}. Using this notation, the true SAT at a given time and location can be
203 described as the value reported in the SAT column of Supplementary Table S3b, plus $X_{bc} + X_{tv}$.
204 Since X_{bc} describes the position of the actual temperature within the range of temperatures
205 allowed by the fossil record, whereas X_{tv} is determined by the point in time an observation
206 belongs to, it is reasonable to assume that both quantities are random and independent.

207 The numerical values for r_{bc} and r_{tv}could be interpreted in different ways. The values could
208 either be seen as hard limits for the corresponding temperature range (for example, the value of r_{bc}
209 could be interpreted to mean that we always have $-r_{bc} \leq X_{bc} \leq r_{bc}$), or as soft limits (for
210 example, by assuming that the random value X_{tv} satisfies $-r_{tv} \leq X_{tv} \leq r_{tv}$ with a given
211 probability (e.g. $P(-r_{tv} \leq X_{tv} \leq r_{tv}) = 95\%$)). The same interpretations are possible for r_{total}.
212 Similarly, different assumptions are possible for the distributions of the random values X_{bc} and
213 X_{tv}. In particular if the ranges are interpreted as hard limits, the random values could be assumed
214 to be uniformly distributed on the range of possible values, or the distributions could be assumed
215 to be more concentrated near the centre of the ranges (e.g. by assuming a normal distribution).

216 To represent the choices outlined above, we tested two different methods for combining the
217 individual ranges into a single value. For the first method, we assume that X_{bc} and X_{tv} are
218 independent and normally distributed with variances such that $P(-r_{bc} \leq X_{bc} \leq r_{bc}) = 95\%$ and
219 $P(-r_{tv} \leq X_{tv} \leq r_{tv}) = 95\%$. In this case, the values r_{bc} and r_{tv} correspond to 1.96 standard
220 deviations of X_{bc} and X_{tv}, respectively, and the corresponding value for r_{total} is given by:

221 $$r_{total} = \sqrt{r_{bc}^2 + r_{tv}^2}$$

222 This method seems the most natural approach (Supplementary Table S3b; Method 1). For the

second method, we assume that X_{bc} and X_{tv} are uniformly distributed on the intervals $[-r_{bc}, r_{bc}]$ and $[-r_{tv}, r_{tv}]$ respectively. For easy comparison to the first method, we again choose r_{total} to be the value such that $P(-r_{total} \leq X_{bc} + X_{tv} \leq r_{total}) = 95\%$ (Supplementary Table S3b; Method 2 (95%)). Since the probability of 95% in this approach was chosen rather arbitrarily, we also tested a variant of Method 2, where the probability 90% was used instead (Supplementary Table S3b, Method 2 (90%)). The resulting total ranges for the methods tested are very similar, and thus the method for combining r_{bc} and r_{tv} into a total range has little effect on the results. Therefore, in Figure 3f of our analysis, we use Method 1 described here to derive r_{total}, and highlight sites where the MMM SAT anomaly falls within r_{total} with green circles.

Figure 3g presents a summary of the data in Figures 3c to 3f and includes all proxy-data localities and the full range of SAT anomalies derived from the PlioMIP ensemble. Sites are coded as purple squares if any of the proxy-data available (both with and without ranges) falls within the model ensemble range. If this is the case, we argue that the data can be used as a means of discriminating between models within the ensemble. However, such an undertaking without knowledge of the magnitude of r_{bc} and r_{tv} would likely generate incorrect conclusions regarding individual model performance. White squares in Figure 3f demonstrate sites where the available proxy-data (i.e. with or without ranges) and the model ensemble do not overlap. One particular region where this occurs is in North East Asia/Siberia, which shows that the models are consistently too cold in comparison to the data.

7. Performing a data-model comparison on biome reconstructions

Global biome maps were compared numerically by employing the ArcView 3.x extension for kappa statistics [30]. The Kappa statistic measures the degree of agreement between predicted and observed categorisations of a dataset or map, while correcting for agreement that occurs by chance [31]. Kappa statistics have already been successfully applied for comparing Quaternary and Pliocene biome

reconstructions [11, 32, 33]. Kappa values (κ) are ranked using a subjective assessment scale between 0 and 1, whereby "0" means that the agreement is no better than would be expected by chance and "1" stands for a perfect match. Kappa results strongly depend on the number of different classes selected and are rarely comparable across studies. Therefore, here we interpret our kappa values in isolation from previous studies [e.g. 7, 34]. For comparing biome reconstructions from fossil data with model simulations, we grouped the 28 biomes into broader units (mega-biomes[35]) to avoid the minimum number of sample points per category becoming too low for meaningful Kappa statistics.

8. PlioMIP Experiment 2 Energy Balance Analysis

By following the methods of Heinemann et al. [36] and Lunt et al. [37], we can model the energy balance components of each of the PlioMIP Experiment 2 simulations by approximating the temperature, T, at each latitude using the planetary albedo α, the effective longwave emissivity ϵ, and the implied net meridional heat transport divergence, H. Where $SW_{TOA}^{\downarrow}$ is the incoming shortwave radiation at the top of the atmosphere and σ is the Stefan-Boltzmann constant,

$$T = \frac{(SW_{TOA}^{\downarrow}(1-\alpha) - H)^{1/4}}{\epsilon\sigma}$$

Due to their small changes relative to their absolute values, Pliocene warming can be approximated by a linear combination of changes in emissivity, albedo and heat transport. However, by assessing clear sky radiation components within the simulations, these components can be further broken down into the impact of changes in atmospheric greenhouse gases ($\Delta T_{gg\epsilon}$), cloud emissivity ($\Delta T_{c\epsilon}$), cloud albedo ($\Delta T_{c\alpha}$) and clear sky albedo ($\Delta T_{cs\alpha}$). At latitudes where changes in topography occur between the Pliocene and control simulations, the impact of these changes in surface altitude (ΔT_{topo}) must also be accounted for. Therefore,

$$\Delta T_{Plio} = \Delta T_{gg\epsilon} + \Delta T_{c\epsilon} + \Delta T_{c\alpha} + \Delta T_{cs\alpha} + \Delta T_{H} + \Delta T_{topo}$$

The two simulations that show the greatest changes in the high latitudes of the Northern Hemisphere (MIROC and COSMOS) have the greatest climate sensitivity. Both also show the largest changes due to feedbacks in each of the albedo and emissivity components, even where the feedback is negative (Supplementary Figure S3). The complete loss of summer sea-ice and strength of Arctic feedbacks explain why these models show the greatest polar amplification of all the PlioMIP simulations (Supplementary Figure S4). Conversely CCSM and MRI show the least polar amplification, despite having climate sensitivities close to the best estimates of IPCC[38]. These simulations seem to have suppressed positive albedo feedbacks and, particularly in the case of MRI, enhanced negative feedbacks at key latitudes of the Northern Hemisphere.

9. References

1. Haywood, A.M., *et al.* Pliocene Model Intercomparison Project (PlioMIP): Experimental design and boundary conditions (Experiment2). *Geosci. Model Dev.* **4**, 571-577 (2011).

2. Dowsett, H., *et al*. The PRISM3D Paleoenvironmental Reconstruction. *Stratigraphy*, **7**, 123-139 (2010).

3. Kaplan, J.O. *Geophysical applications of vegetation modeling*. PhD thesis, Lund University, Lund (2001).

4. Kaplan, J.O. *et al.* Climate change and arctic ecosystems II: Modeling, paleodata-model comparisons, and future projections. *J. Geophys. Res.–Atmos.* **108**, 8171 (2003).

5. Haxeltine, A. & Prentice, I.C. BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constrains, resource availability, and competition among plant functional types. *Global Biogeochem. Cy.* **10**, 693-709 (1996).

6. Texier, D., *et al.* Quantifying the role of biosphere- atmosphere feedbacks in climate change: coupled model simulations for 6000 years BP and comparison with paleodata for northern Eurasia and Africa. *Clim. Dynam.* **13**, 865–882 (1997).

293 7. Salzmann, U., Haywood, A.M., & Lunt, D.J. The past is a guide to the future? Comparing Middle
294 Pliocene vegetation with predicted biome distributions for the twenty-first century. *Phil. Trans. R.*
295 *Soc. A* **367**, 189-204 (2009).

296 8. Leemans, R. & Cramer, W. The IIASA database for mean monthly values of temperature,
297 precipitation and cloudiness of a global terrestrial grid. *International Institute for Applied Systems*
298 *Analysis (IIASA)*. **RR-91-18** (1991).

299 9. Laskar, J. *et al.* A long-term numerical solution for the insolation quantities of the Earth. *Astron.*
300 *Astrophys.* **428**, 261-285 (2004).

301 10. Pagani, M., Liu, Z., LaRiviere, J., Ravelo, A.C. High Earth-system climate sensitivity determined
302 from Pliocene carbon dioxide concentrations. *Nature Geosci*. **3**, 27–30 (2010).

303 11. Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J., Hill, D.J. A new global biome reconstruction
304 and data-model comparison for the Middle Pliocene. *Global Ecol. Biogeogr.* **17**, 432-447 (2008).

305 12. Verhoeven, K., Louwye, S. Eiriksson, J. & De Schepper, S. A new age model for the Pliocene-
306 Pleistocene Tjornes section on Iceland: Its implication for the timing of North Atlantic-Pacific
307 palaeoceanographic pathways. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **309**, 33-52 (2011).

308 13. Wilson, G.S., *et al.* The Mount Feather Diamicton of the Sirius Group: an accumulation of
309 indicators of Neogene Antarctic glacial and climatic history. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*
310 **182**, 117–131 (2002).

311 14. Gao, C., *et al.* Glaciation of North America in the James Bay Lowland, Canada, 3.5 Ma. *Geology*
312 **40**,75-978 (2012).

313 15. Velasco-de León, M.P., Spicer, R.A. & Steart, D.C. Climatic reconstruction of two Pliocene floras
314 from Mexico. *Palaeobiodivers. Palaeoenviron.* **90**, 99–110 (2010).

315 16. Blokhina, N. & Bondarenko, O. Woody plant assemblages and palaeoenvironments in the Pliocene
316 of Pavlovskaya Depression (Southern Primory'e). *Acta Palaeontol. Rom.* **4**, 23-35 (2004).

17. Enikeev, F.I. The Late Cenozoic of northern Transbaikalia and paleoclimates of southern East Siberia. *Russ. Geol. Geophys.* **49**, 602-610 (2008).

18. Uhl, D., *et al*. Cenozoic paleotemperatures and leaf physiognomy: A European perspective. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **248**, 24-31 (2007).

19. Mai, D.H. The floral change in the Tertiary of the Rhön mountains (Germany). *Acta Palaeobot.* **47**, 135–143 (2007).

20. Teodoridis, V., Kvaček, Z. & Uhl, D. Late Neogene palaeoenvironment and correlation of the Sessenheim-Auenheim floral complex. *Palaeodiversity* **2**, 1-17 (2009).

21. Vieira M., Poças E., Pais J. & Pereira D. Pliocene flora from S. Pedro da Torre deposits (Minho, NW Portugal). *Geodiversitas* **33**, 71-85 (2011).

22. Yavuz-Isik, N. & Toprak, V. Palynostratigraphy and vegetation characteristics of Neogene continental deposits interbedded with the Cappadocia ignimbrites (Central Anatolia, Turkey). *Int. J.Earth Sci.* **99**, 1887-1897 (2010).

23. Wolfe, J.A. A method of obtaining climatic parameters from leaf assemblages. *U.S. Geol. Surv. Bull.* **2040**, 1-71 (1993).

24. Mosbrugger, V. & Utescher, T. The coexistence approach – a method for quantitative reconstructions of Tertiary terrestrial palaeoclimate data using plant fossils. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **134**, 61–86 (1997).

25. Fauquette, S., *et al*. Latitudinal climatic gradients in the Western European and Mediterranean regions from the Mid-Miocene (c. 15 Ma) to the Mid-Pliocene (c. 3.5 Ma) as quantified from pollen data. [Williams, M., Haywood, A.M., Gregory, F.J., Schmidt, D.N. (eds.)] *Deep–Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies*. The Micropalaeontological Society, Special Publications. The Geological Society, London, 481–502 (2007).

26 Bruch, A. A., T. Utescher, V. Mosbrugger, I. Gabrielyan, Ivanov, D.A. Late Miocene climate in the circum-Alpine realm--a quantitative analysis of terrestrial palaeofloras. *Palaeogeogr., Palaeoclimatol.,*

342 *Palaeoecol*. **238**, 270-280 (2006).

343 27 Grimm, G. W., & Denk, T. Reliability and resolution of the coexistence approach - A revalidation
344 using modern-day data. *Rev. Palaeobot. Palyno*. **172**, 33-47 (2012).

345 28. Legates, D.R., & Willmott, C.J. Mean seasonal and spatial variability in global surface air
346 temperature. *Theor. Appl. Climatol.* **41**, 11-21 (1990).

347 29 Haywood, A. M. & Valdes, P. J. Modelling middle Pliocene warmth: Contribution of atmosphere,
348 oceans and cryosphere. *Earth Planet. Sci. Lett*. **218**, 363-377 (2004).

349 30. Jenness, J. & Wynne J.J. Cohen's Kappa and classification table metrics 2.0: an ArcView 3x
350 extension for accuracy assessment of spatially explicit models. *U.S. Geol. Survey Open-File Report*
351 **2005-1363** (2005).

352 31. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **2**, 37–46 (1960).

353 32. Prentice, I.C., *et al.* A global biome model based on plant physiology and dominance, soil
354 properties and climate. *J. Biogeogr.* **19**, 117-134 (1992).

355 33. Harrison, S.P., *et al.* Intercomparison of simulated global vegetation distributions in response to 6
356 kyr BP orbital forcing. *J. Climate* **11**, 2721-2742 (1998).

357 34. Monserud, R.A. & Leemans, R. Comparing global vegetation maps with the Kappa statistic. *Ecol.*
358 *Model.* **62**, 275-293 (1992).

359 35. Harrison, S.P. & Prentice, I.C. Climate and CO_2 controls on global vegetation distribution at the last
360 glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate
361 simulations. *Glob. Change Biol.* **9**, 983-1004 (2003).

362 36. Heinemann, M., et. al.: Warm Paleocene/Eocene climate as simulated in ECHAM5/MPI-OM, *Clim.*
363 *Past* **5**, 785-802 (2009).

364 37. Lunt, D.J., et. al.: A model-data comparison for a multi-model ensemble of early Eocene
365 atmosphere-ocean simulations: EoMIP. *Clim. Past Discuss*. **8**, 1229-1273 (2012).

366 38. Randall, D.A., et. al.: Climate Models and Their Evaluation. In: Climate Change 2007: The Physical
367 Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the
368 Intergovernmental Panel on Climate Change [Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M.,
369 Averyt, K.B., Tignor, M. and Miller, H.L. (eds.)]. Cambridge University Press, Cambridge, United
370 Kingdom and New York, NY, USA (2007).

371 39. Haywood, A.M. et al. Large-scale features of Pliocene climate: results from the Pliocene Model
372 Intercomparison Project. *Clim. Past*. **9**, 191-209 (2013).

373 40. Neale, R.B., et al. The mean climate of the Community Atmosphere Model (CAM4) in forced SST
374 and fully coupled experiments. *J. Climate*, In Press, (2013)

375 41. Smith, R. D., *et al*. Parallel Ocean Program (POP) reference manual, Ocean component of the
376 Community Climate System Model (CCSM). *Tech. Rep.* **LAUR-10-01853**, Los Alamos National
377 Laboratory, Los Alamos, NM. (Available online at http://www.ccsm.ucar.edu/models/ccsm3.0/pop)
378 (2010).

379 42. Danabasoglu, G., *et al.*: The CCSM4 Ocean Component. *J. Climate* **25**, 1361–1389 (2012).

380 43. Holland, M. M., et al. Improved sea ice shortwave radiation physics in CCSM4: The impact of melt
381 ponds and black carbon. *J. Climate* **25**, 1413-1430, doi:10.1175/JCLI-D-11-00078.1 (2012).

382 44. Hunke, E.C. & Lipscomb, W.H. CICE: The Los Alamos Sea Ice Model. Documentation and Software
383 User's Manual. Version 4.0. T-3 Fluid Dynamics Group, Los Alamos National Laboratory. *Tech. Rep.*
384 **LA-CC-06-012** (2008).

385 45. Gent, P.R., et al. The Community Climate System Model version 4, *J. Climate*, 24, 4973-4991, 2012.

386 46. Oleson, K.W., *et al.* Improvements to the Community Land Model and their impact on the
387 hydrological cycle. *J. Geophys. Res.*, **113**, G01021 (2008).

388 47. Stockli, R., *et al.* Use of FLUXNET in the Community Land Model development. *J. Geophys. Res.* **113**,
389 G01025 (2008).

48. Rosenbloom, N.A., Otto-Bliesner B.L., Brady E.C. & Lawrence, P.J. Simulating the mid-Pliocene climate with the CCSM4 model. *Geosci. Model Dev. Disc*. **5**, 4269-4303 (2012).

49. K-1 Model Developers. *K-1 Coupled Model (MIROC) Description: K-1 Technical Report 1* (eds Hasumi, H. & Emori, S.). (Center for Climate System Research, University of Tokyo, 2004).

50. Oki, T. & Sud, Y. C. Design of Total Runoff Integrating Pathways (TRIP)—A Global River Channel Network. *Earth Interact.* **2**, 1–37 (1998).

51. Chan, W.-L., Abe-Ouchi, A. & Ohgaito, R. Simulating the mid-Pliocene climate with the MIROC general circulation model: experimental design and initial results. *Geosci. Model Dev.* **4**, 1035–1049 (2011).

52. Pope, V.D., Gallani, M.L., Rowntree, P.R. & Stratton, R.A. The impact of new physical parametrizations in the Hadley Centre Climate model: HadAM3. *Clim. Dynam.* **16**, 123-146 (2000).

53. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dynam.* **16**, 147-168 (2000).

54. Cattle, H. & Crossley, J. Modelling arctic climate change. *Phil. Trans. R. Soc. A* **352**, 201-213 (1995).

55. Cox, P. M. *et al.* The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Clim. Dynam.* **15**, 183-203 (1999).

56. Bragg, F.J., Lunt, D.J., & Haywood, A.M. Mid-Pliocene climate modelled using the UK Hadley Centre Model: PlioMIP Experiments 1 and 2. *Geosci. Model Dev. Discuss.* **5**, 837-871 (2012).

57. Schmidt, G.A. et al.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, in-preparation.

58. Hansen, J., *et al.,* Climate simulations for 1880-2003 with GISS modelE. *Clim. Dynam*. **29**, 661-696, doi:10.1007/s00382-007-0255-8 (2007).

59. Liu, J. *et al.* Sensitivity of sea ice to physical parameterizations in the GISS global climate model. *J. Geophys. Res.* **108**, 3053 (2003).

414 60. Kiang, N.Y., et al. A dynamic global terrestrial ecosystem model for coupling with GCMs, American
415 Geophysical Union Fall Meeting. San Francisco, CA, December 10-15 (2006).

416 61. Chandler, M.A., *et al.,* Simulations of the Mid-Pliocene Warm Period using the NASA/GISS
417 ModelE2-R Earth System Model. *Geosci. Model Dev. Discuss.* **5**, 2811-2842 (2012).

418 62. Roeckner, E., *et al.,* The atmospheric general circulation model ECHAM5, PART I: Model
419 description, Report 349, Max-Planck-Institut fuer Meteorologie, Hamburg (2003).

420 63. Marsland, S.J., *et al.,* The Max-Planck-Institute global ocean/sea ice model with orthogonal
421 curvilinear coordinates. *Ocean Modell*. **5**, 91–127 (2003).

422 64. Hibler, W.D.: A Dynamic Thermodynamic Sea Ice Model. *J. Phys. Oceanogr*. **9**, 815–846 (1979).

423 65. Jungclaus, J. H., *et al.,* Ocean Circulation and Tropical Variability in the Coupled Model
424 ECHAM5/MPI-OM. *J. Climate* **19**, 3952–3972 (2006).

425 66. Raddatz, T.J., *et al.,* Will the tropical land biosphere dominate the climate-carbon cycle feedback
426 during the twentyfirst century? *Clim. Dynam*. **29**, 565–574 (2007).

427 67. Hagemann, S. and Duemenil, L.: Documentation for the Hydrological Discharge Model, DKRZ
428 *Technical Report No. 17*, Deutsches Klimarechenzentrum, Hamburg (1998).

429 68. Hagemann, S. and Gates, L. D.: Improving a subgrid runoff parameterization scheme for climate
430 models by the use of high resolution data derived from satellite observations. *Clim. Dynam*. **21**, 349–
431 359 (2003).

432 69. Stepanek, C. and Lohmann, G.: Modelling mid-Pliocene climate with COSMOS. *Geosci. Model Dev*. **5**,
433 1221–1243 (2012).

434 70. Hourdin, F., *et al.,* The LMDZ4 general circulation model: climate performance and sensitivity to
435 parametrized physics with emphasis on tropical convection. *Clim. Dynam.* **27**, 787–813 (2006).

436 71. Hourdin F., *et al.,* Climate and sensitivity of the IPSL-CM5A coupled model: impact of the LMDZ
437 atmospheric grid configuration. *Clim. Dynam.* In Press (2013).

438 72. Dufresne *et al.,* Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3
439 to CMIP5. *Clim. Dynam*., In Press (2013).

440 73. Madec, G., *et al.,* OPA version 8.1 Ocean general circulation model reference manual, 3. LODYC,
441 Technical Report, 91 pp. available at: http: //www.nemo-ocean.eu/About-NEMO/Reference-
442 manuals, (1997).

443 74. Fichefet, T. and Morales-Maqueda, M. A.: Sensitivity of a global sea ice model to the treatment of
444 ice thermodynamics and dynamics. *J. Geophys. Res*. **102**, 12609–12646 (1997).

445 75. Fichefet, T. and Morales-Maqueda, M. A.: Modelling the influence of snow accumulation and snow-
446 ice formation on the seasonal cycle of the Antarctic sea-ice cover. *Clim. Dynam*. **15** 251–268 (1999).

447 76. Marti O., *et al.,* Key features of the IPSL ocean atmosphere model and its sensitivity to
448 atmospheric resolution. *Clim. Dynam*. **34**, 1–26 (2010).

449 77. Krinner, G., *et al.,* A dynamic global vegetation model for studies of the coupled atmosphere-
450 biosphere system. *Global Biogeochem. Cycles* **19**, GB1015, doi:10.1029/2003GB002199, (2005).

451 78. Contoux, C., *et al.,* Modelling the mid-Pliocene Warm Period climate with the IPSL coupled model
452 and its atmospheric component LMDZ5A. *Geosci. Model Dev*. **5** (2012).

453 79. Yukimoto, S., *et al.,* Present-day and climate sensitivity in the meteorological research institute
454 coupled GCM version 2.3 (MRI-CGCM2.3). *J. Meteorol. Soc. Japan* **84**, 333–363 (2006).

455 80. Mellor, G. L. and Kantha, L.: An ice-ocean coupled model. J. Geophys. Res. **94**, 10937–10954
456 (1989).

457 81. Sellers, P.J., *et al.,* A simple biosphere model (SiB) for use within general circulation models. *J.*
458 *Atmos. Sci.* **43**, 505–531 (1986).

459 82. Sato, N., *et al.,* Effects of implementing the simple biosphere model in a general circulation model.
460 *J. Atmos. Sci*. **46**, 2757–2782 (1989).

461 83. Kamae, Y. and Ueda, H.: Mid-Pliocene global climate simulation with MRI-CGCM2.3: set-up and
462 initial results of PlioMIP Experiments 1 and 2. *Geosci. Model Dev.* **5**, 793-808 (2012).

463 84. Zhang, Z.S., *et al.,* Pre-industrial and mid-Pliocene simulations with NorESM-L. *Geosci. Model Dev.*
464 **5**, 523-533 (2012).

465 85. Dowsett, H.J. *et al*. Middle Pliocene paleoenvironmental reconstruction: PRISM2. *USGS Open File*
466 *Report*, **99**-535(1999).

467 86. Rybczynski, N., Gosse, J.C., Harington, C.R., Wogelius, R.A., Hidy, A.J. & Buckley, M. Mid-Pliocene
468 warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Commun.*
469 **4**, 1550*, DOI: 10.1038/ncomms2516* (2013)

470 87. Ballantyne, A.P. *et al.* Significantly warmer Arctic surface temperatures during the
471 Pliocene indicated by multiple independent proxies. *Geology* **38**, 603–606, (2010).

472 88. Fradkina, A.F. in *Abstracts of the Joint US/USSR Workshop on Pliocene Paleoclimates*. (eds.
473 Thompson, R.S., Borisova, O.K. & Svetlitskaya, T.V.), pp. 22 (1991).

474 89. Nelson, R.E. & Carter, L.D. Pollen analysis of a late Pliocene & early Pleistocene section from the
475 Gubik Formation of Arctic Alaska. *Quaternary Res.* **24**, 295-306 (1985).

476 90. Ager, T.A., Matthews, J.J.V. & Yeend, W. Pliocene terrace gravels of the ancestral Yukon River near
477 Circle, Alaska: Palynology, paleobotany, paleoenvironmental reconstruction and regional correlation.
478 *Quatern. Int.* **22-23**, 185-206 (1994).

479 91. Popova S., Utescher, T., Gromyko, D., Bruch, A.A. & Mosbrugger, V. Palaeoclimate Evolution in
480 Siberia and the Russian Far East from the Oligocene to Pliocene – Evidence from Fruit and Seed
481 Floras. *Turkish J. Earth Sci.* **21**, 315–334 (2012).

482 92. Volkova, V.S. in *Abstracts of the Joint US/USSR Workshop on Pliocene Paleoclimates*. (eds.
483 Thompson, R.S., Borisova, O.K. & Svetlitskaya, T.V.), pp. 44-45 (1991).

484 93. Head, M.J. Pollen and dinoflagellates from the Red Crag at Walton-on-the-Naze, Essex; evidence
485 for a mild climatic phase during the early late Pliocene of eastern England. *Geol. Mag.* **135**, 803-817
486 (1998).

94. Mai, D.H. & Walther, H., 1988. Die pliozänen Floren von Thüringen, Deutsche Demokratische Republik. *Quartärpaläontology* **7**, 55–297.

95. Willis, K.J., Kleczkowski, A. & Crowhurst, S.J. 124,000-year periodicity in terrestrial vegetation change during the late Pliocene epoch. *Nature* **397**, 685-688 (1999).

96. Wolfe, J.A. in *Pliocene Climates: Scenario for Global Warming*. (eds. Gosnell, L.B. & Poore, R.Z.), 39-42 (1990).

97. Fauquette, S., & Bertini, A. Quantification of the northern Italy Pliocene climate from pollen data: evidence for a very peculiar climate pattern. *Historical Biology* **32**, 361-369 (2003).

98. Fauquette, S., *et al.* Climate and biomes in the West Mediterranean area during the Pliocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **152**, 15-36 (1999).

99. Suc, J.P., & Cravatte, J. Etude palynologique du Pliocene de Catalogne (Nord-ouest de l'Espagne). *Paléobiologie continentale* **13**, 1-31 (1982).

100. Mamedov, A.V. in *Abstracts of the Joint US/USSR Workshop on Pliocene Paleoclimates*. (eds. Thompson, R.S., Borisova, O.K. & Svetlitskaya, T.V.), pp. 28 - 31 (1991).

101. Axelrod, D.I. The Sonoma Flora (California). *Publ. Carnegie Inst. Wash.* **553**, 167-206 (1944).

102. Thompson, R.S. Pliocene environments and climates in the western United States. *Quaternary Sci. Rev.* **10**, 115-132 (1991).

103. Iwauchi, A. Late Cenozoic vegetational and climatic changes in Kyushu, Japan. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **108**, 229-280 (1994).

104. Graham, A. & Dilcher, D.L. Studies in Neotropical paleobotany. XII. A palynoflora from the Pliocene Rio Banano Formation of Costa Rica and the Neogene vegetation of Mesoamerica. *Am. J. Bot.*, **85**, 1426-1438 (1998).

105. Martin, H.A. & McMinn, A. Palynology of Sites 815 and 823: the Neogene vegetation history of coastal north-eastern Australia. *P. Ocean Drill. Program, Sci Results* **133**, 115-123 (1993).

106. Macphail, M.K. Late Neogene climate in Australia. *Aust. J. Biol.* **45**, 425-460 (1997).

512 107. Kershaw, A.P. & Sluiter, I.R. Late Cenozoic Pollen Spectra from the Atherton Tableland, North-
513 eastern Australia. *Aust. J. Bot.* **30**, 279-295 (1982).

514 108. Kershaw, P., Martin, H.A., & McEwen Mason, J.R.C. in *History of Australian Vegetation: Cretaceous*
515 *to Recent*. (ed. Hill, R.S.), pp. 299 - 327. (Cambridge Univ. Press, Cambridge, 1994).

516 109. Dodson, J.R. & Macphail, M.K. Palynological evidence for aridity events and vegetation change
517 during the Middle Pliocene, a warm period in Southwestern Australia. *Global Planet. Change* **41**, 285-
518 307 (2004).

519 110. Macphail, M.K., Colhoun, E.A., & Fitzsimons, S.J. Key Periods in the Evolution of the Cenozoic
520 Vegetation and Flora in Western Tasmania: the Late Pliocene. *Aust. J. Bot.* **43**, 505-526 (1995).

521 111. Sohl, L.E., *et al.* PRISM3/GISS topographic reconstruction. *U. S. Geol. Surv. Data Series* **419**, 6 pp.,
522 2009.

523 112. Utescher, T. Mosbrugger, V. Palaeoflora Database. Available at http://www.palaeoflora.de
524 (2010).

525 113. Demske, D., Mohr, B. & Oberhansli, H. Late Pliocene vegetation and climate of the Lake Baikal
526 region, southern East Siberia, reconstructed from palynological data. *Palaeogeogr. Palaeoclimatol.*
527 *Palaeoecol.* **184**, 107-129 (2002).

528 114. Utescher, T., Mosbrugger, V., & Ashraf, A.R. Terrestrial Climate Evolution in Northwest Germany
529 over the Last 25 Million Years. *PALAIOS* **15**, 430-449 (2000).

530 115. Willard, D.A. Palynological record from the North Atlantic region at 3 Ma: vegetational
531 distribution during a period of global warmth. *Rev. Palaeobot. Palyno.* **83**, 275-297 (1994).

532 116. Suc, J.P., *et al.* Zanclean (~Brunssumian) to early Piacenzian (~ early-middle Reuverian) climate
533 from 4 to 54 north latitude (West Africa, West Europe and West Mediterranean areas). *Meded. Rijks*
534 *Geol. Dienst.* **52**, 43-56 (1995).

117. Bonnefille, R., Potts, R., Chalie, F., Jolly, D. & Peyron, O. High-resolution vegetation and climate change associated with Pliocene Australopithecus afarensis. *P. Nat. Acad. Sci.* **101**, 12125-12129 (2004).

10. Supplementary Tables

Table S1: Details of models used for the terrestrial data/model comparison, details of boundary conditions as well as the climate sensitivity values (°C) for each model[39].

Model ID, Vintage	Sponsor(s), Country	Atmosphere Top Resolution References	Ocean Resolution Z Coord., Top BC References	Sea Ice Dynamics, Leads, References	Coupling Flux adjust-ments, references	Land Soils, Plants, Routing, References	PlioMIP Experi-ment 2[7] Preferred/ Alternate	Climate Sensi-tivity (°C)
CCSM4, 2010	National Center for Atmospheric Research, USA	Top = 2.2 hPa 0.9x1.25°, L26 (40)	1° x 1°, L60 Depth, free surface (41, 42)	Rheology, melt ponds (43, 44)	No adjustments (45)	Layers, canopy, routing (46, 47)	Alternate (48)	3.2
MIROC4m, 2004	Center for Climate System Research (Uni. Tokyo, National Inst. for Env. Studies, Frontier Research Center for Global Change, JAMSTEC), Japan	Top = 30 km T42 (~ 2.8° x 2.8°) L20 (49)	0.5° -1.4° x 1.4°, L43 Sigma/depth free surface (49)	Rheology, leads (49)	No adjustments (49)	Layers, canopy , routing (49, 50)	Preferred (51)	4.05
HadCM3, 1997	Hadley Centre for Climate Prediction and Research/Met Office UK	Top = 5 hPa 2.5° x 3.75°, L19 (52)	1.25° x 1.25°, L20 Depth, rigid lid (53)	Free drift, leads (54)	No adjustments (53)	Layers, canopy , routing (55)	Alternate (56)	3.1
GISS-E2-R, 2010	NASA/GISS, USA	Top = 0.1 hPa 2˚ x 2.5˚, L40 (57)	1˚ x 1.25˚, L32 Mass/area, free surface (58)	Rheology, leads (57, 59)	(57)	Layers, canopy, routing (60)	Preferred (61)	2.7 to 2.9
COSMOS COSMOS-landveg r 2413, 2009	Alfred Wegener Institute, Germany	Top = 10 hPa T31 (3.75˚x 3.75˚), L19 (62)	Bipolar orthogonal curvilinear GR30, L40 (formal 3.0˚x 1.8˚) Depth, free surface (63)	Rheology, leads (63), following (64)	No adjustments (65)	Layers, canopy, routing (66, 67, 68)	Preferred (69)	4.1
IPSLCM5A, 2010	Laboratoire des Sciences du Climat et de l'Environnement (LSCE), France	Top = 70 km 3.75° x 1.9°, L39 (70, 71)	0.5°-2° x 2°, L31 Free surface, Z-coordinates (72, 73)	Thermodyna mics, Rheology, Leads (74, 75)	No adjustment (72, 75)	Layers, canopy, routing, phenology (72, 76, 77)	Alternate (78)	3.4
MRI-CGCM 2.3, 2006	Meteorological Research Institute and University of Tsukuba, Japan	Top = 0.4 hPa T42 (~2.8° x 2.8°) L30 (79)	0.5°-2.0° x 2.5°, L23 Depth, rigid lid (79)	Free drift, leads (80)	Heat, fresh water and momentum (12°S-12°N) (79)	Layers, canopy, routing (81, 82)	Alternate (83)	3.2
NorESM-L (CAM4), 2011	Bjerknes Centre for Climate Research, Bergen, Norway	Top = 3.5 hPa T31 (~3.75° × 3.75°), L26 (CAM4)	G37 (~3° x 3°), L30 isopycnal layers	Same as CCSM4	Same as CCSM4	Same as CCSM4	Alternate (84)	3.1

547 **Table S2**: Details of the additional orbital sensitivity experiments performed with HadCM3 and

548 MIROC4m.

Experiment	Description	PRISM2 (Boundary Conditions[50])	PRISM3 (Boundary Conditions[7])	Orbit (kyr)	CO_2 (ppmv)	Eccentricity	Precession	Obliquity (°)
$Plio^{PRISM}$	Pliocene control simulation using PRISM boundary conditions[7].	HadCM3 (CO_2 set at 400)	HadCM3 & MIROC4m	Modern	405	Modern	Modern	Modern
$Plio^{NHMax}{}_{400}$	Pliocene simulation with maximum Pliocene incoming insolation at 65°N in July.	HadCM3	HadCM3 & MIROC4m	3037	400	0.051086	-0.04239	23.642
$Plio^{NHMax}{}_{450}$	Pliocene simulation with maximum Pliocene incoming insolation at 65°N in July and high CO_2 levels.	HadCM3	HadCM3 & MIROC4m	3037	450	0.051086	-0.04239	23.642
$Plio^{SHMax}{}_{400}$	Pliocene simulation with maximum Pliocene incoming insolation at 65°S in January.	HadCM3	HadCM3 & MIROC4m	3049	400	0.054523	0.05204	23.143
$Plio^{SHMax}{}_{450}$	Pliocene simulation with maximum Pliocene incoming insolation at 65°S in January and high CO_2 levels.	HadCM3	HadCM3 & MIROC4m	3049	450	0.054523	0.05204	23.143

549

550

551

552

553

554

555

556

557

558

559

560

561

Table S3a: Late Pliocene mean annual temperature estimates (SATs) from vegetation reconstructions. Temperature estimates for sites with modern altitude > 1000 m a.s.l. are not included.

location	continent	latitude	longitude	altitude[a]	age (Ma)	method[b]	SAT	bioclim. range	temporal variability	confi-dence[c]	reference
Beaver Pond/Ellesmere Isl.	North America	78.40	-82.00	350	3.8-3.4	Multi-Proxies	-1.4	± 4.0	n/a	1	86, 87
Lena River	Asia	72.20	125.97	5	3.2-2.6	QualEst	1.5	± 1.0	n/a	4	88
Ocean Point	North America	70.00	-153.00	308	2.7-2.6	QualEst	1.5	n/a	n/a	3	89
Circle, Alaska	North America	65.50	-144.08	325	3.6-3	QualEst	3.0	n/a	n/a	3	90
Blizkiy	Asia	64.00	162.00	400	3.6-1.8	CA	5.3	± 5.8	n/a	3	91
Nenana Valley, Alaska	North America	64.53	-149.08	295	3.6-2.8	QualEst	3.0	n/a	n/a	3	90
Lost Chicken Mine	North America	64.06	-141.95	325	3.3-2.5	QualEst	2.5	n/a	n/a	3	90
Delyankir	Asia	63.00	133.00	600[a]	3.3-1.8	CA	7.4	± 0.5	n/a	3	91
Magadan District	Asia	59.98	150.65	97	3.2-2.6	QualEst	2.0	n/a	n/a	4	88
West Siberia	Asia	56.03	70.32	25	3.2-2.6	QualEst	13.5	± 1.5	n/a	4	92
Merkutlinskiy	Asia	56.00	72.00	50	3.3-1.8	CA	11.8	± 4.5	n/a	3	91
Kabinet / 42km	Asia	55.00	80.00	50	3.3-1.8	CA	8.9	± 2.3	n/a	3	91
Mirny	Asia	55.00	82.00	50	3.3-1.8	CA	11.2	± 1.3	n/a	3	91
Maly-shik / Logovskoy	Asia	54.00	81.00	50	3.3-1.8	CA	8.5	± 4.1	n/a	3	91
Walton-on-the-Naze**	Europe	51.84	1.27	25	3-2.6	CA*	12.8	± 1.3	n/a	2	93
Willershausen	Europe	51.77	10.10	212	3.2-2.6	CLAMP, CA	13.9	± 2.7	n/a	2	18
Berga/Thuringia	Europe	51.53	11.02	212	2.65-2.6	CA	13.5	± 0.5	n/a	3	18, 94
Pula Maar	Europe	47.05	17.38	200	3-2.98	CA*	12.8	± 1.2	n/a	1	95
Oak Grove Forest	North America	45.80	-121.60	212	3.05-2.95	CLAMP	11.9	± 1.0	n/a	3	96,102
Stirone	Europe	44.60	10.15	779	2.8-2.6	CAM	15.0	± 2.0	± 3.0	1	97
Pavlovskaya Depression	Asia	44.09	132.09	200	3.2-2.6	CA	5.9	± 1.5	n/a	3	16
Garraf, Catalonia	Europe	41.17	2.02	62	3.6-3.2	QualEst	19.0	n/a	n/a	2	98,99
Kura Depression	Europe	40.53	49.69	226	3.2-2.6	QualEst	21.0	n/a	n/a	4	100
California/Sonoma-Napa	North America	38.30	-122.45	313	3.45-3.35	CLAMP	17.6	± 2.0	n/a	3	101, 102
Central Kyushu*	Asia	33.10	131.50	149	2.9-2.8	QuantEst	18.0	n/a	± 1.0	2	103
Rio Banano, Zent	North America	10.03	-83.28	12	3.6-2.6	QualEst	27.0	n/a	n/a	3	104
ODP 823, Leg 133 (b)**	Australia	-16.69	145.21	190	3.2-2.6	QualEst	21.5	± 1.5	n/a	4	105, 106
West Butcher Creek	Australia	-17.35	145.70	190	3.6-2.6	QualEst	21.2	n/a	n/a	3	107
Lake George	Australia	-35.15	149.42	720	2.8-2.6	QualEst	14.0	± 2.0	n/a	3	106, 108
Yallalie, Perth	Australia	-30.43	115.76	77	3.2-2.95	QualEst	21.0	± 5.0	n/a	2	109
Linda Valley, Tasmania	Australia	-42.83	145.67	300[a]	3.6-2.5	QualEst	14.0	± 2.0	n/a	3	110

* corrected SAT at sea level after Iwauchi [103]

** land surface SAT of potential nearest terrestrial source area

n/a - no range or climate variability not identified

Altitude[a]: Palaeoaltitude in masl., after Sohl *et al.*[111] (modern altitude has been used for Chara Basin and Linda Valley)

Key to confidence assessment[c]: 1 – very high, 2 – high, 3 - medium, 4 - low

Key to "Methods"[b]

QualEst	Qualitative estimates using modern analogues
CLAMP	Climate Leaf Analysis Multivariate Programe
CAM	Climate Amplitude Method
BA	Best Analogue/Plant Functional Type Method
QuantEst	Quantitative estimates using pollen indices
CA	Coexistence Approach
CA*	estimated from Palaeoflora Database [112]

Table S3b: Late Pliocene mean annual temperature estimates (SATs) from vegetation reconstructions for sites providing information on temporal variability during the PRISM time slab (~3.3-3.0 Ma). Temperature estimates for marine sites more than 250 km away from mainland are not included.

													Bioclim.Range + Temp. Variab.*		
location	continent	latitude	longitude	altitude [a]	age (Ma)	method [b]	SAT	bioclim range	temporal variability	No of samples	confi-dence [c]	refer-ence	Method 1	Method 2 (95%)	Method 2 (90%)
Chara Basin, Siberia	Asia	56.97	118.31	700[a]	3.3-3.0	QualEst	12.8	n/a	± 1.8	17	2	17	n/a	n/a	n/a
Lake Baikal	Asia	55.69	108.37	450	3.3-3.0	CA*	7.0	± 2.5	± 3.0	12	1	113	3.91	4.28	3.77
James Bay Lowland	North America	52.83	-83.88	50	3.3-3.0	QualEst	6.0	± 2.0	± 4.0	>40	1	14	4.47	4.74	4.21
Lower Rhine Basin	Europe	51.03	6.53	135	3.6-2.6	CLAMP, CA	14.1	± 0.2	± 0.3	3	1	114	0.36	0.39	0.35
Sessenheim-Auenheim	Europe	48.82	8.01	297	3.6-2.6	CA	14.6	± 0.7	± 0.5	2	2	20	0.86	0.94	0.83
Alpes-Maritimes	Europe	43.82	7.19	193	3.3-3.2	CAM	17.5	± 2.0	± 0.5	18	2	99	2.06	2.05	1.87
Tarragona	Europe	40.83	1.13	401	3.3-3.0	CAM	20.0	± 2.5	± 2.5	3	2	99, 101	3.54	3.88	3.42
Rio Maior	Europe	39.35	-8.93	42	3.6-3.0	CAM	16.0	± 2.0	± 2.0	17	1	99	2.83	3.11	2.74
Yorktown, Virginia	North America	36.59	-76.38	57	3.5-2.9	QualEst	17.5	n/a	± 0.3	20	2	115	n/a	n/a	n/a
Andalucia G1	Europe	36.38	-4.75	305	3.6-2.6	CAM	21.0	± 2.0	± 3.5	12	1	99	4.03	4.32	3.83
Habibas	Africa	35.73	-1.12	325	3.6-3.2	CAM	21.0	± 1.0	± 3.0	>15	2	99, 116	3.16	3.23	2.90
Nador	Africa	35.18	-2.93	206	3.6-2.6	CAM	21.5	± 1.0	± 3.0	7	1	99	3.16	3.23	2.90
Pinecrest, Florida	North America	27.36	-82.44	10	3.5-2.6	QualEst	23.1	n/a	± 0.0	16	2	115	n/a	n/a	n/a
Hadar	Africa	11.29	40.63	849	3.4-2.9	BA	20.5	± 1.0	± 3.5	26	1	117	3.64	3.66	3.32

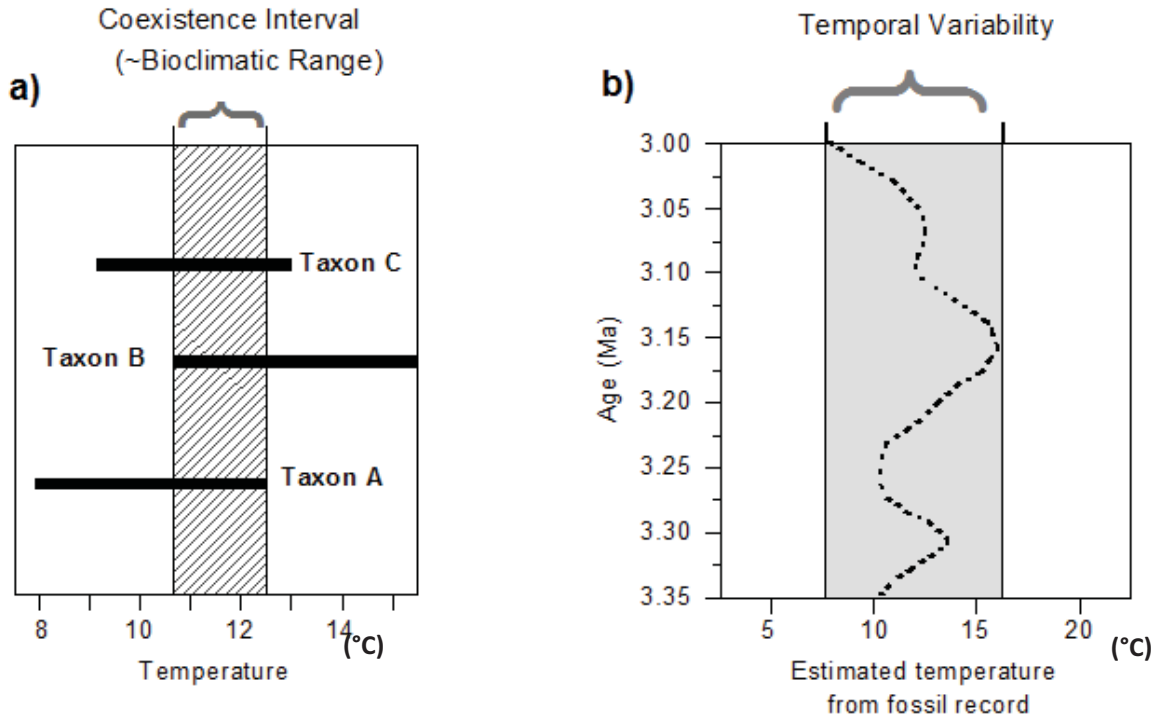
* - see Supplmentary Section 6 for description of Methods
n/a - no range or climate variability not identified

Table S4: Data-model comparison for BIOME4 HadCM3 and MIROC. Table shows difference between model and proxy-based mega-biome reconstructions using a warm (k-WB) and cold (k-CB) biased dataset. Kappa values [31] (κ) for global, polar, temperate tropical zones are ranked using a subjective assessment scale, whereby "0" means that the agreement is no better than would be expected by chance and "1" stands for a perfect match.

Model	R^2 (SAT)	Global	Polar	Temperate	Tropical
		k-WB/k-CB	k-WB/k-CB	k-WB/k-CB	k-WB/k-CB
Pliocene HadCM3 PRISM2	**0.85**	**0.31/0.25**	0.01/0.00	0.29/0.21	0.03/0.01
NH Max Orbit	0.87	**0.28/0.27**	0.30/0.14	0.17/0.17	0.19/0.16
SH Max Orbit	0.85	**0.30/0.26**	0.00/0.00	0.26/0.21	0.11/0.09
NH Max Orbit CO_2 450 ppm	0.87	**0.31/0.26**	0.42/0.21	0.21/0.16	0.15/0.12
SH Max Orbit CO_2 450 ppm	0.85	**0.33/0.30**	0.03/0.00	0.26/0.23	0.16/0.14
Pliocene HadCM3 PRISM3	**0.83**	**0.30/0.30**	0.41/0.26	0.23/0.24	0.00/0.00
NH Max Orbit	0.84	**0.25/0.25**	0.24/0.14	0.17/0.18	0.14/0.12
SH Max Orbit	0.83	**0.30/0.28**	0.23/0.13	0.25/0.25	0.05/0.03
NH Max Orbit CO_2 450 ppm	0.84	**0.22/0.19**	0.20/0.12	0.14/0.13	0.09/0.06
SH Max Orbit CO_2 450 ppm	0.83	**0.30/0.27**	0.35/0.21	0.24/0.20	0.03/0.01
Pliocene MIROC PRISM3	**0.83**	**0.30/0.27**	0.36/0.22	0.21/0.18	0.06/0.05
NH Max Orbit	0.84	**0.26/0.28**	0.17/0.10	0.21/0.26	0.12/0.10
SH Max Orbit	0.83	**0.29/0.27**	0.15/0.00	0.21/0.20	0.06/0.04
NH Max Orbit CO_2 450 ppm	0.84	**0.25/0.26**	0.12/0.06	0.22/0.25	0.11/0.08
SH Max Orbit CO_2 450 ppm	0.83	**0.29/0.27**	0.30/0.18	0.20/0.18	0.06/0.04

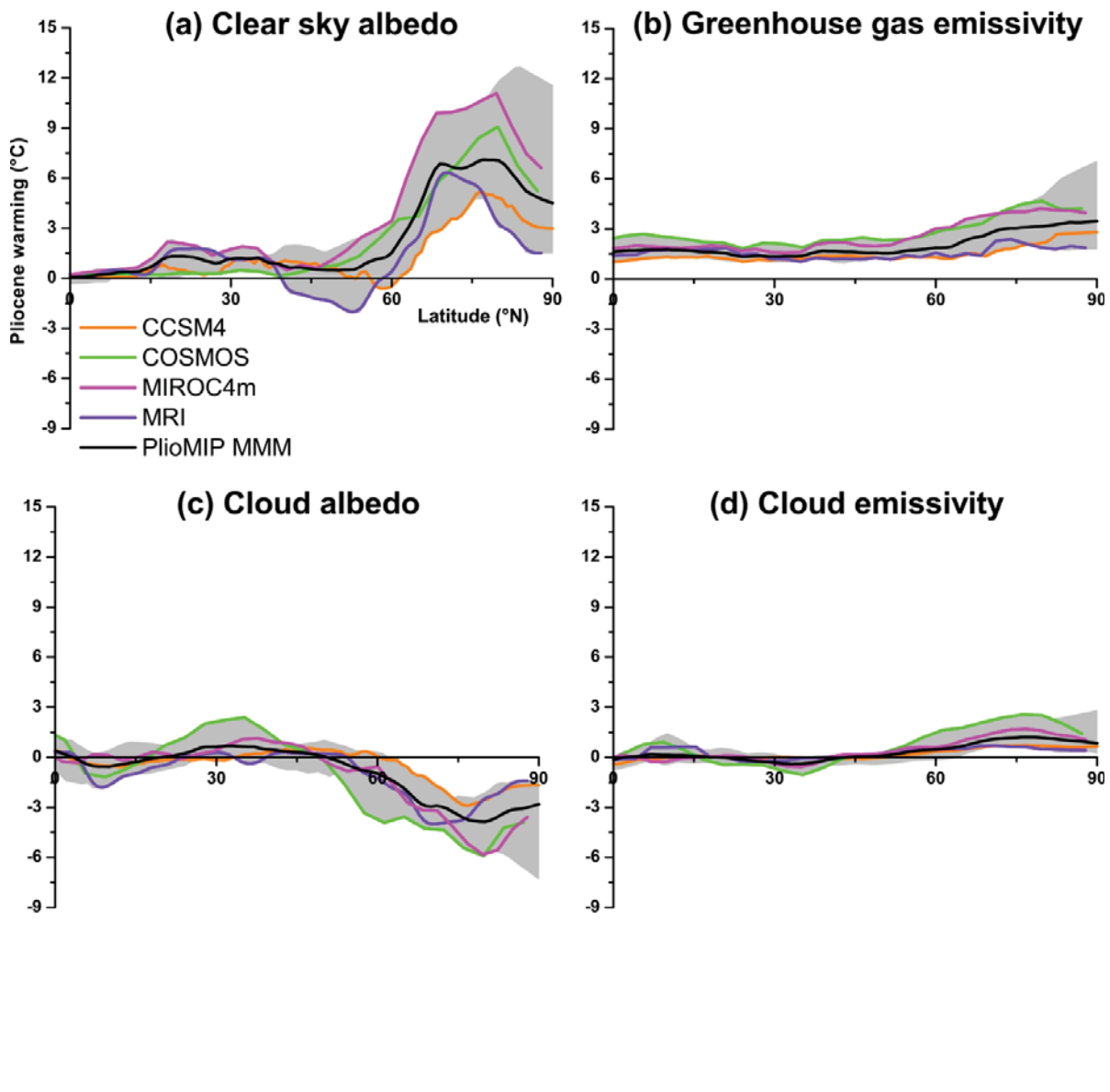
Supplementary Figure S1

"Bioclimatic range" and "temporal variability" used for DMC. Fig S1 a) provides an example for a bioclimatic range produced by the Coexistence Approach, showing the mean annual temperature ranges tolerated by the nearest living relatives taxa A, B and C. The resulting temperature interval in which all taxa can coexist lies between 10.8 and 12.5°C (modified after [24]). b) shows the temporal variability caused by the variation of the reconstructed temperature over the geological period represented in the fossil record.



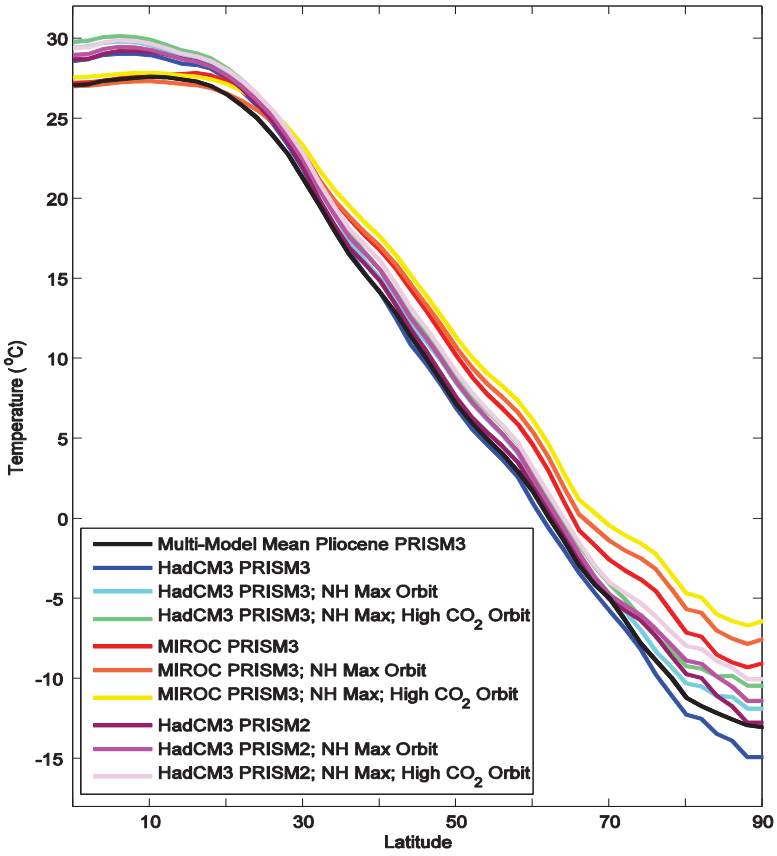
Supplementary Figure S2

Energy balance analysis for PlioMIP Experiment 2. Northern Hemisphere temperature change due to (a) clear sky albedo, (b) greenhouse gas emissivity, (c) cloud albedo and (d) cloud emissivity. PlioMIP Experiment 2 multi-model mean (MMM) is plotted in black, with the grey shading showing the range of values. Two warmest models (COSMOS and MIROC4m), as well as the two least warm (CCSM4 and MRI) are also shown.



615 **Supplementary Figure S3**

616 Zonally averaged mean annual surface air temperatures (SAT) in °C derived from MIROC and HadCM3
617 simulations with varying boundary conditions from the PRISM3/2 dataset, changing orbital parameters
618 and atmospheric CO_2-concentration. The Pliocene MMM is also displayed for reference.



619