**A return to large-scale features of Pliocene climate: the Pliocene Model Intercomparison Project Phase 2**

Haywood1\*, A.M., Tindall1, J.C., Dowsett2, H.J., Dolan1, A.M., Foley2, K.M., Hunter1, S.J., Hill1, D.J., Chan3, W-L., Abe-Ouchi3, A., Stepanek4, C., Lohmann4, G., Chandan5, D., Peltier5, R., Tan6/7, N., Contoux7, C., Ramstein7, G., Li6/8, X., Zhang6/8, Z. Nisancioglu8, K.H., Zhang9, Q., Zheng9, J., Kamae10/11, Y., Yoshida12, K., Ueda13, H., Chandler14, M.A., Sohl14, L.E., Otto-Bliesner15, B., Feng, R16, von der Heydt17, A., Baatsen16, M., Lunt18, D.J.

1. School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds, West Yorkshire, LS29JT, UK

2. Centre for Earth Surface System Dynamics (CESD), Atmosphere and Ocean Research Institute (AORI), University of Tokyo, Japan

3. Florence Bascom Geoscience Center, U.S. Geological Survey, Reston, VA 20192, USA

4. Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

5. Department of Physics, University of Toronto, Toronto, Ontario, Canada

6. Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

7. Laboratoire des Sciences du Climat et de l'Environnement, Saclay, France

8. Bjerknes Centre for Climate Research, Bergen, Norway

9. Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

10. Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

11. Scripps Institution of Oceanography, University of California San Diego, USA

12. Meteorological Research Institute, Tsukuba, 305-0052, Japan

13. Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

14. CCSR/GISS, Columbia University, New York, USA

15. National Center for Atmospheric Research, Boulder, Colorado, USA

16. College of Liberal Arts and Sciences, University of Connecticut, Connecticut, USA

17. Department of Physics, Centre for Complex Systems Science, Utrecht University, Utrecht, The Netherlands

18. School of Geographical Sciences, University of Bristol, Bristol, UK.

\*Corresponding author email: earamh@leeds.ac.uk

**Keywords**: Pliocene; climate; PlioMIP2; Equilibrium Climate Sensitivity, Earth System Sensitivity; Proxy data

**Abstract**

The Pliocene epoch has great potential to improve our understanding of the long-term climatic and environmental consequences of an atmospheric CO2 concentration at ~400 ppmv. Here we present the large-scale features of Pliocene climate as simulated by a brand-new ensemble of climate models of varying complexity and spatial resolution (the Pliocene Model Intercomparison Project Phase 2). As a global annual average, Near Surface Air Temperatures increases by 1.4 to 5.1°C relative to pre-industrial with a multi-model mean value of 2.8°C. Annual mean total precipitation rates increase by 6%. On average, near surface air temperature increases are 1.3°C higher over the land than over the oceans, and there is a clear pattern of polar amplification with warming polewards of 70° a factor of 2.8 greater than the global mean. In the Atlantic and Pacific Ocean, meridional temperature gradients are reduced, whilst tropical zonal gradients remain largely unchanged. Excluding the extreme models, there is a statistically significant relationship between the Equilibrium Climate Sensitivity of the models comprising the ensemble and the simulated Pliocene surface temperature response. The new ensemble supports the conclusions made from PlioMIP1 that Earth System Sensitivity is greater than Climate Sensitivity by a best estimate of 55%.

1. **Introduction**

*1.1 Pliocene climate modelling and the Pliocene Model Intercomparison Project*

Efforts to understand climate dynamics during the mid-Piacenzian Warm Period (MP; 3.264 to 3.025 million years ago), previously referred to as the mid-Pliocene Warm Period, have been ongoing for more than 25 years. Beginning with the initial climate modelling studies of Chandler et al. (1994), Sloan et al. (1996) and Haywood et al. (2000), the complexity and number of climate models used to study the MP has increased substantially (e.g. Haywood and Valdes 2004). This progression culminated in 2008 with the initiation of a co-ordinated international model intercomparison project for the Pliocene (Pliocene Model Intercomparison Project: PlioMIP). PlioMIP Phase 1 proposed a single set of model boundary conditions based on the US Geological Survey PRISM3D data set (Dowsett et al., 2010), and a unified experimental design for atmosphere-only and fully coupled atmosphere-ocean climate models (Haywood et al. 2010, 2011).

PlioMIP1 produced many publications analysing different aspects of MP climate. The large-scale temperature and precipitation response of the model ensemble was presented in Haywood et al. (2013a). The global annual mean temperature increased by 1.8 to 3.6°C, with warming predicted at all latitudes, but with a clear pattern of polar amplification resulting in a reduced equator to pole surface temperature gradient (Haywood et al., 2013a). Modelled sea-ice responses were studied by Howell et al. (2016), who demonstrated a significant decline in Artic sea-ice extent, with specific models simulating a seasonal sea-ice free Arctic Ocean. Mid-latitude westerly winds shifted poleward, a response consistent with a poleward shift in meridional circulation (Li et al., 2015). Corvec and Fletcher (2017) studied the effect of reduced meridional temperature gradients on tropical atmospheric circulation. They demonstrated a weaker tropical circulation during the MP, specifically a weaker Hadley Circulation, and in specific climate models Walker Circulation. This response is akin to model predictions for the future (IPCC, 2013). Tropical cyclones (TC) were analysed by Yan et al. (2016) who demonstrated that average global TC intensity and duration increased during the MP, but this result was sensitive to how much tropical sea surface temperatures (SSTs) increased in each model. Zhang et al. (2013 and 2016) studied the East Asian and West African summer monsoon response in the PlioMIP1 ensemble and found that both were stronger during the MP. The response of the Atlantic Meridional Overturning Circulation (AMOC) was analysed by Zhang et al. (2014). No clear pattern of either weakening or strengthening of the AMOC could be determined from the model ensemble, a result at odds with long-standing interpretations of MP meridional SST gradients being a result of enhanced Ocean Heat Transport (OHT: e.g. Dowsett et al., 1992).

Hill et al. (2014) analysed the dominant components of MP warming across the PlioMIP1 ensemble using an energy balance analysis. In the tropics, increased temperatures were predominantly in response to direct CO2 forcing, whilst at the high-latitudes changes in clear sky albedo became the dominant contributor, with the warming being only partially offset by cooling driven by cloud albedo changes.

Finally, Hargreaves and Annan (2016) carried out an analysis of Equilibrium Climate Sensitivity (ECS) and attempted to constrain PlioMIP1-derived ECS predictions against available MP tropical SST data. This resulted in a proxy data constrained ECS estimate of 1.9 to 3.7°C. In addition, the PlioMIP1 model ensemble indicated a best estimate of the Earth System Sensitivity was a factor of 1.47 higher than Climate Sensitivity (ensemble mean CS = 3.4°C: ensemble mean ESS = 5.0°C: Haywood et al., 2013a).

*1.2 From PlioMIP1 to PlioMIP2*

The ability of the PlioMIP1 models to reproduce patterns of surface temperature change, reconstructed by marine as well as terrestrial proxies, was investigated via data/model comparison (DMC) in Dowsett et al. (2012; 2013) and Salzmann et al. (2013) respectively. Whilst, the PlioMIP1 ensemble was able to reproduce many of the spatial characteristics of SST and surface air temperature (SAT) warming, the models appeared unable to simulate the magnitude of warming reconstructed at the higher latitudes (Dowsett et al. 2012, 2013; Haywood et al., 2013a; Salzmann et al. 2013). This problem has also been reported as an outcome of DMC studies for other time periods including the early Eocene (e.g. Lunt et al., 2012). Haywood et al. (2013a, 2013b), discussed the possible contributing factors to the noted discrepancies in DMC, noting three primary causal groupings: uncertainty in model physics, uncertainty in model boundary conditions and uncertainty in the interpretation of proxy data.

These findings substantially influenced the experimental design for the second Phase of PlioMIP (PlioMIP2). Specifically, PlioMIP2 was developed to (a) better sample sources of uncertainty in MP modelling, (b) reduce uncertainty in model boundary conditions, (c) reduce uncertainty in proxy data reconstruction. In order to accomplish (b), state-of-the-art approaches were adopted to generate an entirely original palaeogeography (compared to PlioMIP1), including accounting for glacial isostatic adjustments and changes in dynamic topography. This led to specific changes, compared to the PlioMIP1 palaeogeography, capable of influencing climate model simulations (Dowsett et al. 2016). This includes the Bering Strait and Canadian Archipelago becoming sub-aerial and modification of the land/sea mask in the Indonesian/Australian region, including the emergence of the Sunda and Sahul Shelf. In order to achieve (c) it was necessary to move away from time-averaged (and therefore aliased) global SST reconstructions, towards the examination of a narrow time slice during the late Pliocene that had almost identical astronomical parameters to the present-day. This made the orbital parameters specified in model experimental design, consistent with the way in which orbital parameters would have influenced the pattern of surface climate preserved in the geological record. Using the astronomical solution of Laskar et al. (2004), Haywood et al. (2013b) identified a suitable interglacial event during the late Pliocene (Marine Isotope Stage KM5c). The new PRISM4 (Pliocene Research, Interpretation and Synoptic Mapping) global community-sourced data set of high temporal resolution SSTs (Foley and Dowsett, 2019) targets the same interval in order to produce a point-based SST data.

Here we briefly present the PlioMIP2 experimental design, details of the climate models included in the ensemble, as well as the boundary conditions used. Following this, we present the large-scale features of the PlioMIP2 ensemble focussed solely on an examination of the control MP simulation designated as a CMIP6 simulation (called *MPEoi400)* and its differences to simulated conditions for the pre-industrial era (PI). PlioMIP2 sensitivity experiments will be presented in subsequent studies. We conclude by presenting the outcomes from a DMC using the PlioMIP2 model ensemble using a newly constructed PRISM4 global compilation of SSTs (Foley and Dowsett, 2019), and assess the significance of the PlioMIP2 ensemble in understanding Equilibrium Climate Sensitivity (ECS) and Earth System Sensitivity (ESS).

**2. Methods**

*2.1 Boundary Conditions*

All model groups participating in PlioMIP2 were required to use standardised boundary condition data sets for the core *MPEoi400* experiment (hereafter referred to as Plio\_Core). These were derived from the U.S. Geological Survey PRISM data set, specifically the latest iteration of the reconstruction known as PRISM4 (Dowsett et al. 2016). This includes spatially complete gridded data sets at 1° × 1° of latitude/longitude resolution for the distribution of land versus sea, topography, bathymetry, as well as vegetation, soils, lakes and land ice cover. Two versions of the PRISM4 boundary conditions were produced known as standard and enhanced. The standard version of the PRISM4 boundary conditions provides the best possible realisation of Pliocene conditions based around a modern land/sea mask. The enhanced boundary conditions include all reconstructed changes to the land/sea mask and ocean bathymetry. For full details of the PRISM4 reconstruction and methods associated with its development, the reader is referred to Dowsett et al. (2016: this volume).

*2.2 Experimental Design*

The experimental design for Plio\_Core and associated PI control experiments (hereafter referred to as (PI\_Ctrl) was presented in Haywood et al. (2016; this volume), and the reader is referred to this paper for full details of the experimental design. In brief, participating model groups had a choice of which version of the PRISM4 boundary conditions to implement (standard or enhanced). This approach was taken in recognition of the technical complexity associated with the modification of the land/sea mask and ocean bathymetry in some of the very latest climate and earth system models. A choice was also included regarding the treatment of vegetation. Model groups could either prescribe vegetation cover from the PRISM4 dataset (vegetation sourced from Salzmann et al. 2008), or simulate the vegetation using a dynamic global vegetation model. If the latter was chosen, all models should be initialised with pre-industrial vegetation and spun-up until an equilibrium condition is reached. The concentration of atmospheric CO2 for experiment Plio\_Core was set at 400 ppmv, a value almost identical to that chosen for the PlioMIP1 experimental design (405ppmv), and in line with the very latest high-resolution proxy reconstruction of atmospheric CO2 of ~400 ppmv for ~3.2 million years ago using Boron isotopes (De La Vega et al. 2018). All other trace gases, orbital parameters and the solar constant were specified to be consistent with each model PI\_Ctrl experiment. Integration length was set to be ‘as long as possible’, or a minimum of 500 simulated years. The final 100 years of the PI\_Ctrl and Plio\_Core experiments were used to compute the required climatological means. In addition, all modelling groups were requested to fully detail their implementation of PRISM4 boundary conditions, along with the initialisation and spin-up of their experiments in separate dedicated papers that also present some of the key science results from each model, or family of models (see the separate papers within this special volume: <https://www.clim-past.net/special_issue642.html>). NetCDF versions of all boundary conditions used for the Plio\_Core experiment, along with guidance notes for modelling groups can be found here: <https://geology.er.usgs.gov/egpsc/prism/7.2_pliomip2_data.html>.

*2.3 Participating Models*

Fifteen climate models, which were developed at different times, having differing levels of complexity and spatial resolution, completed the Plio\_Core experiment. Summary details of the included models and model physics can be found in Table 1, along with information regarding the implementation of PRISM4 boundary conditions and each models Equilibrium Climate Sensitivity (ECS). The ensemble is almost double the size of that presented in the PlioMIP1 large-scale features publication (Haywood et al. 2013a) and represents one of the largest ensembles of models ever created as part of a formal palaeoclimate model intercomparison project. Each group uploaded the final 100 years of each simulation for analysis. These were then regridded onto a regular 1° × 1° grid using a linear interpolation. Means and standard deviations were then calculated across the 100 years.

**3. Climate Results**

*3.1 Surface air temperature (SAT)*

Increases in Plio\_Core global annual mean SATs, compared to each of the contributing models PI\_Ctrl experiment, range from 1.4 to 5.1 °C (Fig 1a), with an ensemble mean ΔT of 2.8. Increased SATs are predicted throughout the globe (Fig 1b), with an ensemble average warming of ~1.7°C in tropical oceans, and with the warming becoming progressively more intense towards the higher latitudes (Fig 1b,c). Multi-model mean SAT warming can exceed 12°C in Baffin Bay and 7°C in the Greenland Sea (Fig 1b). A result potentially influenced by the closure of the Canadian Archipelago and Bering Strait, as well as by the specified loss of most of the Greenland Ice Sheet (GRIS), and the simulated reduction in Northern Hemisphere sea-ice cover (not shown). In the Southern Hemisphere, warming is pronounced in regions of Antarctica that were deglaciated in the MP in both west and east Antarctica (Fig 1b). Warming in the interior of east Antarctica is limited by the prescribed topography of the MP East Antarctic Ice Sheet (EAIS), which in some places exceeds the topography of the EAIS in the models PI\_Ctrl experiments. In terms of the magnitude, the CESM1.0.5 model (Utrecht Group) has the greatest apparent sensitivity to imposing MP boundary conditions with a simulated ΔT of 5.1°C (Fig. 1a). This is higher than CCSM4-UoT (ΔT = 3.8°C) and CCSM4 (ΔT = 2.6°C). CCSM4 has the same physics as CESM1.0.5, as does CCSM4-UoT apart from specified changes by the University of Toronto (UoT) to the parameterization for the overflow of density-driven currents over oceanic ridges and the diapycnal mixing profiles (Chandan & Peltier, 2016 this volume). A notable difference between the UoT and Utrecht simulations is the response in the Southern Hemisphere (SH) at ~70°S, where the zonal mean warming in the Utrecht CESM1.0.5 simulation is over 7°C higher than in the CESM-UoT simulation (Fig. 1c). Examination of the spatial nature of the respective model results indicates that the difference is focussed in the SH sea-ice zone (Supplementary Figure S1). The least sensitive model is EC-Earth3.1 with a global annual mean ΔT of 1.4°C (Fig. 1a). This is despite the model simulating one of the largest temperature changes at the high northern latitudes (Fig. 1c). The temperature response in this model is muted over most of the globe but is particularly weak over the land surface (Supplementary Figure S1).

Analysis of the standard deviation of the model ensemble (Fig 1d) indicates that models are generally consistent in terms of the magnitude of temperature response in the tropics, especially over the oceans. However, they can differ markedly in the higher latitudes, where the inter-model standard deviation reaches more than 4.5°C.

To evaluate whether the multi-model mean Plio\_Core – PI\_Ctrl anomaly at a gridbox is robust we follow the methodology of Mba et al. (2018) and Nikulin et al. (2018). This requires that two conditions are fulfilled: (1) at least 12 of the 15 models agree on the sign of the anomaly, and (2) the signal to noise ratio (i.e. the ratio of the size of the mean anomaly to the inter-model standard deviation [Fig 1b / Fig 1d]) is greater than or equal to one. Regions where the SAT anomaly is considered robust according to these criteria is shown in figure 2, it is seen that for SAT the Plio\_Core – PI\_Ctrl anomaly is considered robust across the ensemble over nearly all the globe.

*3.2 Seasonal cycle of Near Surface Air Temperature, land/sea temperature contrasts and polar amplification*

The seasonal cycle of the SAT anomaly is presented in Fig. 3a. Overall, the ensemble mean displays a small enhancement in SAT warming in the NH summer and early autumn. However, models within the ensemble have very different characteristics in terms of the monthly and seasonal distribution of the warming. Members within the ensemble fall into three groups in terms of their seasonal response. 1) models which show a clear NH spring/summer to early autumn amplification of SAT warming (e.g. CESM1.0.5, CCSM4-UoT, HadCM3, MIROC4), 2) those that show no clear signal of a seasonal bias (e. g. IPSLCM6A, COSMOS, GISS2.1G) and 3) those that indicate the warming during the NH spring/summer to early autumn is weaker than during other periods of the year (e. g. MRI2.3, EC-Earth3.1). The lack of consistency in the seasonal signal of warming has interesting implications in terms of what degree PlioMIP2 outputs could be used to examine the potential for seasonal bias in proxy data sets. To do this meaningfully would require clear consistency in model seasonal responses, which is absent in the PlioMIP2 ensemble.

In contrast to the lack of consistency regarding the seasonal distribution of SAT change, the ensemble results for land/sea temperature contrasts clearly indicate a greater warming over land than over the oceans (Fig. 3b). The only exception to that is the EC-Earth3.1 model that predicts that the oceans warm slightly more than the land. This result is consistent when changes in the land versus oceans only in the tropics is considered (Fig. 3b – lower panel).

Whether or not the NH warms more than the SH (or vice versa) is another area of model dependency (Fig. 3c), however 9 of the 15 models indicate that 45°N-90°N warms more than 45°S-90°S. The models that tend to indicate greater SH versus NH warming are those that have weaker differences between land and ocean warming (e.g. GISS2.1G, NorESM-L), or show oceans warming more than the land in the case of EC-Earth3.1. As noted in section 3.1 CESM1.0.5 has a large warming ~70°S which contributes to this model simulating a greater degree of warming in the SH versus the NH.

The NH and SH polar amplification (PA) factor for each model is shown in figure 3d. This was obtained by dividing the Plio\_Core – PI\_Ctl SAT anomaly polewards of 70° by the globally averaged SAT anomaly. All models simulate a PA of the warming, although whether there is more PA in the NH or SH is a model dependent feature. Although the average PA across the models is 2.9 for the NH and 2.7 for the SH, the most extreme model (EC-Earth3.1) shows NH polar amplification of 6.1.

*3.3 Meridional/zonal SST gradients in the Pacific and Atlantic*

There has been great interest in the reconstruction of Pliocene SST gradients in the Atlantic and Pacific to provide first order assessments of Pliocene climate change, and to assess possible mechanisms of Pliocene temperature enhancement and ocean/atmospheric dynamic responses (Rind and Chandler, 1991). For example, the meridional gradient in the Atlantic has been discussed in terms of the potential for enhanced Ocean Heat Transport in the Pliocene (e.g. Dowsett et al., 1992). In addition, the SST gradient across the tropical Pacific has been used to examine the potential for change in Walker Circulation and, through this, ENSO dynamics and teleconnection patterns during the Pliocene (Fedorov et al., 2013; Burls & Fedorov, 2014).

The multi-model mean meridional profile of SSTs for the Atlantic Ocean is shown in figure 4a. It is seen that the main difference between the Plio\_Core and PI\_Ctrl experiments in the tropics and sub-tropics is an SST increase of 1.5-2.5°C. The mean difference increases to ~4.5°C in the NH at ~55°N, where the range in warming across the models also increases to between 2°C and 9°C. The Pliocene and Pre-industrial meridional SST profiles in the Pacific is similar to that of the Atlantic, but there is no indication from the multi-model mean for a higher latitude enhancement in meridional temperature (Fig. 4b). However, a large range in the ensemble response is noted.

In the tropical Atlantic (20°N to 20°S) the multi-model mean zonal mean SST for the Pliocene increases by ~1.6°C (ensemble range 0.4°C to 2.8°C), with a flat zonal temperature gradient across the tropical Atlantic (Fig. 4c). In the tropical Pacific both Pliocene and pre-industrial ensembles clearly show the signature of a western Pacific Warm Pool, and the relatively cool waters in the eastern Pacific associated with upwelling (Fig. 4d). As such, a clear temperature gradient is evident in the Pliocene tropical Pacific in the PlioMIP2 ensemble (similar to PlioMIP1) and permanent El-Niño condition are not seen in any of the models (see Supplementary Figure S3). The PlioMIP2 ensemble supports a recent proxy derived reconstruction for the Pacific that, through careful assessments of proxy uncertainty, allows the western tropical Pacific to warm as well as the eastern tropical Pacific (e.g. Tierney et al., 2019).

Using the methodology of Mba et al. (2018) and Nikulin et al. (2018), the signal of SST change seen in the multi-model mean is significant in most ocean grid cells (Supplementary figure S2). Supplementary Figure S3 show the difference between the Pliocene ΔSST for each model in the PlioMIP2 ensemble and the Pliocene ΔSST of the multi-model mean.

*3.4 Total precipitation rate (mm/day)*

Increases in Plio\_Core global annual mean precipitation rates, compared to each contributing models PI\_Ctrl experiment, range from 0.07 to 0.40 mm/day (Fig. 5a). In PlioMIP1 the range was 0.09-0.18mm/day. The ensemble mean ΔPrecip. is 0.17 mm/day. The spatial pattern (Fig. 5b) is generally indicative of an enhanced hydrological cycle, with the largest increases in precipitation rate found in the tropics and in regions of the world that are dominated by the monsoons (west Africa, India, East Asia). The enhancement in precipitation over North Africa is consistent with previous Pliocene model results that have demonstrated a weakening in Hadley Circulation linked to reduced pole to equator temperature gradient (e.g. Corvec and Fletcher 2017). Precipitation rates in Greenland are increased in regions that have become deglaciated and are therefore substantially warmer. Precipitation rates at latitudes associated with the westerly wind belts are enhanced, with an indication of a poleward shift in higher latitude precipitation. This result is consistent with findings from PlioMIP1 (Li et al. 2015). Other more locally defined increases in precipitation rate appear closely linked to localised variations in Pliocene topography and land/sea mask changes for example the Sahul and Sunda Shelf that become subaerial in the Plio\_Core experiment. In general, the models that display the largest SAT sensitivity (i.e. greatest ΔSAT) to the prescription of Pliocene boundary conditions also display the largest enhancements in total precipitation rate (CESM1.0.5, CESM1.2 and CCSM4\_UoT). This is consistent with a warmer atmosphere leading to a greater moisture carrying capacity and therefore greater evaporation and precipitation. The model showing the least sensitivity in terms of precipitation response is GISS2.1G.

Analysis of the standard deviation within the ensemble demonstrates that, in contrast to SAT, models are most consistent regarding ΔPrecip in the extratropics (Fig. 5c). This is similar to the findings from PlioMIP1 (Haywood et al., 2013a) and likely because more precipitation falls in the tropics rather than extratropics. The methodology of Mba et al. (2018) and Nikulin et al. (2018) (described in section 3.1), was used to determine the robustness of the precipitation anomalies (Fig. 5d). Unlike the temperature signal, which was robust throughout most of the globe, there are large regions in the tropics and subtropics where the precipitation signal is uncertain. Changes in precipitation rates in the subtropics are partially robust in many places because at least 12 of the 15 models agree on the sign of change. However, these predicted changes are not fully robust because the magnitude of change is not large compared to the standard deviation seen in the ensemble (Fig. 5c). The signal of precipitation change is determined to be robust in the high latitudes and in the mid-latitudes in regions influenced by the westerlies. This is also the case in regions influenced by the West African, Indian Summer and East Asian Summer Monsoons (Fig. 5d). Supplementary Figure S4 shows the difference between each model’s precipitation anomaly and the multi-model mean anomaly (shown in Fig. 5b)

*3.5 Seasonal cycle of total precipitation and land/sea precipitation contrasts*

As was the case for SAT, the monthly and seasonal distribution of precipitation anomalies are highly model dependant. As an ensemble average there is an indication of a small NH late spring to autumn enhancement in the precipitation anomaly signal (Fig. 6a). This is evident in the models CESM1.0.5, CESM1.2, CCSM4-UoT and IPSLCM6A. Most other models have no clear structure to the monthly distribution of precipitation rate anomalies, other than the GISS2.1G model that simulates the NH late spring to autumn precipitation anomaly as being supressed compared to the rest of the year. An increase in summer precipitation is consistent with a general trend of West African, Indian and East Asia Summer monsoon enhancement, and this will be studied in detail in a forthcoming PlioMIP2 paper.

In terms of the land/sea precipitation anomaly contrast the ensemble, with the exception of the COSMOS model, divides into two groups (Fig. 6b). One in which a clear pattern of precipitation anomaly enhancement over land compared to the oceans is seen (CESM1.0.5, CCSM4, EC-Earth3.1, HadCM3, MIROC4m, NorESM1-F, NorESM-L and CCSM4-UoT), and the other where there is no clear enhancement in the land versus oceans (CESM1.2, GISS2.1G, IPSLCM6A, IPSLCM5A2, IPSLCM5A, MRI-CGCM2.3).

*3.6 Climate and Earth System Sensitivity*

Equilibrium Climate Sensitivity (ECS) is defined as the equilibrium temperature change associated with a doubling of CO2 but ignoring slow feedbacks such as ice sheet changes. The ECS value for each model, used in this paper, is shown (and the source referenced) in Table 1. Here the relationship between the Plio\_Core – PI\_Ctrl SAT anomaly and the ECS will be explored to help provide a link between MP temperature anomalies and possible future anomalies.

Firstly, we consider ECS versus the globally averaged Plio\_Core minus PI\_Ctrl temperature anomaly (Fig. 7a). We see there is no significant relationship between the two (p=0.18). However, if we exclude the most extreme models from our analysis (those models with the largest and smallest Plio\_Core – PI\_Ctl temperature anomalies; CESM1.0.5 and EC-Earth3.1) the relationship between ECS and the Pliocene temperature anomaly becomes significant (Fig. 7b) with p=0.03. This means that we can be more than 95% confident that there is a relationship between a model’s climate sensitivity and the globally averaged MP warming that the model shows.

Next, we investigate how a model’s ECS is related to the Plio-Core – PI\_CTL SAT anomaly on different temporal and spatial scales. In the analysis that follows we will simply assess whether such a correlation exists and its strength by looking at p-values and R squared values. Results from individual models are not presented. Fig. 7c shows the relationship (p-value – blue, Rsquared - red) between ECS and the Pliocene SAT anomaly for each month; with a correlation defined as significant if p < 0.05. The lowest p-value observed is in January and is ≈ 0.12, meaning that ECS is not signiﬁcantly related to the Pliocene temperature anomaly in any individual month. However, if we calculate the statistics after excluding the models with the most warming (CESM1.0.5) and the least warming (EC-Earth3.1) a significant correlation emerges (Fig. 7d) from August through to March.

Further examination looked at which latitudes the Pliocene temperature anomaly is most closely related to ECS. Using the Pliocene temperature anomaly at each latitude for each of the models, the relationship (Rsq and p-value) between the anomaly and ECS was calculated (Fig. 7e). Since a p-value of < 0.05 is needed for the relationship to be signiﬁcant at the 95% conﬁdence level, there is only a signiﬁcant relationship between ECS and the Pliocene temperature anomaly in the NH tropics. The greatest correlation occurs at at 17°N, with Rsq ≈ 0.49 (meaning that 49% of the variance in the Pliocene temperature anomaly at this latitude could be explained by changes in ECS between the models within the ensemble). Again, if we exclude the models with the most warming (CESM1.0.5) and least warming (EC-Earth3.1) (Fig. 7f) a stronger correlation emerges. In this case there is a signiﬁcant relationship between ECS and the Pliocene temperature anomaly at the latitude of interest for the whole of the tropics. This relationship becomes significant at the 99% confidence level between 24°N and 24°S, where a high proportion of the inter-model variability in ECS can be related to the inter-model variability in the latitudinal average Pliocene SAT anomaly, reaching a maximum of 65% at 10°N.

Next the relationship between global ECS and the Pliocene temperature anomaly at a grid point scale was assessed. This follows a similar approach to that of Hargreaves and Annan (2016). In Fig. 7g colours show the Rsq correlation between the grid point Pliocene temperature anomaly and the ECS across the models. The regions where the relationship between the grid point temperature anomaly and the published climate sensitivity is signiﬁcant at the 95% conﬁdence level is hatched. Using data from all models in the ensemble, the relationship between ECS and the grid box Pliocene temperature anomaly is signiﬁcant over parts of the tropics (e.g. the tropical Paciﬁc and Northern Africa.) However, if we exclude EC-Earth3.1 and CESM1.0.5 models from the ensemble (Fig. 7i) there is a much stronger relationship between the grid box temperature anomaly and ECS. In many cases, the tropical oceans show a temperature anomaly more strongly related to ECS than the land although this is not always the case.

Following the methodology presented in Haywood et al. (2013a) we consider the elevated CO2 concentration in the atmosphere to be the ultimate forcing of Pliocene warmth (an assumption consistent with the study by Tierney et al., 2019), and so PlioMIP2 simulations represent the equilibrium state of a world at 400 ppmv of CO2. To convert this to Earth System Sensitivity (i.e. a CO2 doubling from 280 to 560ppmv), the Pliocene warming is multiplied by ln(560.0/280.0) / ln(400.0/280.0) = 1.94. For the multi-model mean ESS is 5.5°C, and the ESS/ECS ratio is 1.55, values for each individual model are provided in Table 2.

**4. Data/Model Comparison**

Haywood et al (2013a/b) proposed that the proxy data/climate model comparison in PlioMIP1 could include discrepancies related to the time averaging process of the PRISM3D SST and SAT data, being compared with a climate model representation of a single timeslice. In order to improve the integrity of the data/model comparisons in PlioMIP2 Foley and Dowsett (2019) synthesised alkenone SST data that can be confidently attributed to the MIS KM5c time slice that experiment Plio\_Core is designed to represent. Foley and Dowsett (2019) provide two different SST data sets. One data set includes all SST data for an interval 10,000 years around the time slice (5,000 years either side of the peak of MIS KM5c) and the other 30,000 years (up to 15,000 years either side of the peak; F&D19\_30). Prescott et al. (2014) demonstrated that due to the specific nature of orbital forcing 20,000 years before and after the peak of MIS KM5c, age and site correlation uncertainty within that interval would be unlikely to introduce significant errors into SST-based DMC. Given this, and in order to maximise the number of ocean sites where SST can be derived, we carry out a point-based SST data/model comparison using the F&D19\_30 data set.

We compare the multi-model mean SST anomaly to a proxy SST anomaly created by differencing the F&D19\_30 data set to observed pre-industrial SSTs derived from years 1870-1899 of the NOAA ERSST version 5 data set (Huang et al., 2017; Fig. 8a and Fig. 8b). Figure 8c shows the proxy data ΔSST minus the multi-model mean ΔSST. This enables an assessment of how consistent the sensitivity of the model ensemble is with the sensitivity of the proxy recorder. Using the multi-model mean results 17 sites show a difference in model/data sensitivity of no greater than +/- 1°C. These are located mostly in the tropics, but also includes sites in North Atlantic, California margin, New Zealand and the North Pacific. In terms of discrepancies, the clearest and most consistent signal comes from the Benguela upwelling system where the model ensemble does not predict the scale of warming seen in the proxy reconstruction. The ensemble is insufficiently sensitive in the two Mediterranean Sea sites, along the east coast of North America (Yorktown Formation), and at one location west of Svalbard close the sea-ice margin. The multi-model mean predicts too great a warming at one location off the Florida and Norwegian coasts. No discernible spatial pattern or structure is seen (outside of the Benguela region) for sites where the ensemble under or overestimates the magnitude of SST change.

Comparing model predicted and proxy based absolute SST estimate for the timeslice (Fig. 8d) yields a similar outcome to the comparison of SST anomalies. Furthermore, a somewhat clearer picture emerges of the model ensemble not producing SSTs that are warm enough as you move to the higher latitudes of the North Atlantic and especially Nordic Sea. Although this appears site dependant as the ensemble overestimates absolute SSTs near Scandinavia.

**5. Discussion**

*5.1 Large-scale features of a warmer climate (palaeo vs future, older vs younger models)*

The range in the global annual mean ΔSST shown by the PlioMIP2 ensemble (1.4 to 5.1°C) is akin to the best estimate (and uncertainty bounds) of predicted global temperature change by 2100AD using the RCP2.6 to 8.5 scenarios (IPCC, 2013). The ensemble multi-model mean response of 2.8°C is akin to the expected 2100AD response predicted under RCP6.0. Comparing the degree of Pliocene temperature change to predicted changes at 2300AD, the multi-model mean SAT change is most akin to RCP4.5 with the ensemble within the range of RCP4.5 and 8.5.

Studies have suggested that polar temperature change may be amplified between 1 to 3 times the global annual mean temperature response due to a doubling of atmospheric CO2 concentration (Hind et al., 2016). Eleven out of the fifteen models within the PlioMIP2 ensemble simulate a polar amplification factor (averaged over the NH and SH) of ≤ 3. Of the 4 models that simulate a larger polar amplification factor (EC-Earth3.1, GISS2.1G, NorESM-L, MRI2.3), GISS2.1G and NorESM-L only show polar amplification > 3 in Southern Hemisphere. An important difference to note in the comparison between Pliocene and future predicted polar amplification factors is the major changes in the size of the ice sheets, which in terms of area of ice difference affect the Southern Hemisphere far more than the Northern Hemisphere.

Both model simulations and observations show that as temperatures rise, the land warms more than the oceans. This is due to differential lapse rates linked to moisture availability on land. From a theoretical standpoint the difference in land/sea warming is expected to be monotonic with increases in temperature. However, the rise is non-monotonic and is regulated by latitudinal and regional variations in the availability of soil moisture that influence lapse rates (Byrne and O’Gorman, 2013). This is evident in the PlioMIP2 ensemble with land/sea amplification of warming noted more strongly in the global mean than in the tropics where precipitation is most abundant (Fig. 3b). For the PI case, modelling and observational studies have shown that land warms 30 to 70% more than the oceans (Lambert & Webb, 2011). The PlioMIP2 ensemble broadly supports this conclusion and previous work. It also supports studies that have indicated that the land/sea warming contrast is not dependent upon either a transient or an equilibrium-type climate change scenario (e.g. Lambert & Webb, 2011).

In predictions of future climate change, a consistent result from models is that the warming signal is amplified in the Northern compared to Southern Hemisphere in the extra tropics. There have been several studies which have proposed mechanisms to explain this including heat uptake by the Southern Ocean (Stouffer et al., 1989) as well ocean heat transport mechanisms (Russell et al., 2006). Within the PlioMIP2 ensemble, 9 out of 15 models show a larger temperature change in in the Northern Hemisphere extratropics than the Southern Hemisphere extratropics (Fig. 3c). This can in part be explained by the area of land in the Northern Hemisphere being larger than the South and the already discussed amplification of warming over the land versus the oceans. However, the degree of difference is highly model dependent and not as large as been reported for simulation of future climate change by the IPCC (IPCC, 2013). This may be linked to the intrinsic difference in response between a transient and equilibrium climate experiment, and in the Pliocene substantially reduced ice sheets on Antarctica, which are not specified in future climate change simulations. Hence, the noted hemispheric difference in warming for the future may simply be a transient feature that would not be sustained as the ice sheets on Antarctica responded to the warming over centennial to millennial timescales.

The hydrological cycle is generally enhanced in the Plio\_Core simulations versus the PI. The 2.8°C increase in multi-model mean temperature is associated with a 6% increase in global annual mean precipitation. The water holding capacity of the atmosphere increases by about 7% for each 1°C of temperature increase. The increase in precipitation is less than would be expected if it were assumed that all aspects of the hydrological cycle remained the same as preindustrial. This is line with model simulations of future climate change linked to greater temperatures enhancing evaporation from the surface and the atmosphere having a greater moisture carrying capacity. Whilst the Pliocene results support the wet get wetter and the dry get dryer paradigm to a degree this is not seen in regions experiencing substantial precipitation change in the Pliocene. Most notable is the ensemble response over the modern Sahara Desert and the general prediction of precipitation in the sub tropics in the Pliocene scenario, which is at odds with simulations of climate change for the end of this century. This phenomenon has been discussed in detail previously within the atmospheric circulation and dynamics studies of Corvec and Fletcher (2017) and Sun et al. (2013). Whilst there are many similarities in tropical atmospheric circulation response between Pliocene experiments and future climate change experiment using the RCP4.5 scenario for example, there are specific differences ultimately relating to the nature of equator to pole temperature gradient changes during the Pliocene vis-à-vis the future. Sun et al. (2013) show a dampening in Pliocene Hadley cell intensity in the northern tropics and an increase in both subtropics. Both, northern and southern Hadley cells expand poleward, but the response of Hadley cell is stronger for RCP4.5 scenario than for the Pliocene scenario.

Model sensitivity to Pliocene boundary conditions does not appear to correlate with the release date of the model (i.e. older models are not demonstrably less sensitive than newer models). A correlation between sensitivity and model resolution is present when the model showing the greatest Pliocene warming and the model showing the least Pliocene warming are excluded. However, this is result is strongly influenced by the NCAR family of models that have high resolution and show relatively large precipitation and temperature anomalies. This result would therefore need to be verified by additional studies. The correlation between SAT anomalies and precipitation anomalies and model horizontal resolution is show in supplementary figure S5. The correlation is stronger between model resolution and precipitation (p≈0.01) than it is between model resolution and SAT (p≈0.05).

*5.2 Model representations of Pliocene climate vis-a-vis proxy data*

One of the most fundamental changes in experimental design between PlioMIP2/PRISM4 and PlioMIP1/PRISM3D was the approach towards geological data synthesis for data/model comparison. In particular, moving from SST and vegetation estimates for a broad time slab to short SST time series encompassing the MIS KM5c timeslice. This was necessary in order to assess to what degree climate variability within the Pliocene could affect the outcomes of data/model comparison and, fundamentally, to derive greater confidence in the outcomes which could be derived from Pliocene data/model comparison (Haywood et al., 2013a/b). In addition, PlioMIP2 contains many new models not used in PlioMIP1, and PlioMIP2 boundary conditions have changed almost completely compared to PlioMIP1. Nevertheless, what emerges from the comparison of the PlioMIP2 SST ensemble to the F&D19\_30 SST data set is a nuanced picture of widespread model/data agreement with specific areas of concern.

Previous DMCs for the Pliocene have indicated that the PlioMIP ensemble overestimated the amount of SST change as a zonal mean in the tropics (Dowsett et al., 2012; 2013; Fedorov et al., 2013). In PlioMIP2 point-based comparisons, there is little indication of a systematic mismatch between the data and the models. Models and proxy data appear to be broadly consistent in the tropics. The F&D19\_30 data set is comprised of alkenone-based SSTs for a narrow time slice. In contrast, the PRISM3D data set used for DMC in PlioMIP1 was time averaged and composed of estimates from a combination of faunal analysis, Mg/Ca and alkenone-based SSTs. Tierney et al. (2019) demonstrated that the PlioMIP1 ensemble compared well to alkenone-based SST estimates in the tropical Pacific for the whole mid-Pliocene Warm Period, not just the PlioMIP2 time slice, when the alkenone-based temperature were recalculated using the BAYSPLINE calibration. Therefore, the choice of proxy and inter-proxy calibration alone can be enough to alter the picture of data/model agreement (in the tropics), and in cases where the intrinsic difference between data and models is small. However, the choice of the observed SST data set used to create the Pliocene minus Pre-Industrial SST anomaly can also be important. Supplementary Figure S5 shows the proxy data reconstructed SST change using the F&D19\_30 data set but using two different observed data sets for pre-industrial SSTs to create the required proxy data SST anomaly. Using recently released NOAA ERSST V5 data set (Huang et al., 2017) to create the anomaly instead of the older HadISST data (Rayner et al., 2003) leads to three sites in the North Atlantic showing a much-reduced Pliocene warming. It also means that a number of sites in the tropics now show a small (2 to 3°C) warming during the Pliocene, when using HadISST data led to an absence of SST warming at these locations. The difference between using NOAA ERSST V5 or HadISST are sufficiently large that it can determine whether the PlioMIP2 ensemble is able to largely match (or mismatch) the proxy-reconstructed temperatures.

Another region of data/model mismatch noted in PlioMIP1 was the North Atlantic Ocean (NA). Haywood et al. (2013a) noted a difference in the model-predicted (multi-model mean) versus proxy reconstructed (PRISM3D) warming signal of between 2 to 7°C in the NA. The PlioMIP2 multi-model mean SST change appears to be broadly consistent with the F&D19\_30 data set, with a SST anomaly at two sites matching to within 1°C and the other to within 3°C (Fig 8). There are a number of possible ways to account for this apparent improvement. Firstly, the total number of sites in the NA in F&D19\_30 is reduced compared to the PRISM3D SST data set (Dowsett et al. 2010). The site that led to the 7°C difference noted in Haywood et al. (2013a) is not present in the F&D19\_30 data set. Secondly, the PlioMIP2 experimental design specified both the Canadian Archipelago and Bering Strait as closed. Otto-Bliesner et al. (2017) performed a series of sensitivity tests based on the NCAR CCSM4 PlioMIP1 experiment and found the closure of these Arctic gateways strengthened the AMOC by inhibiting freshwater transport from the Pacific to the Arctic Ocean and from the Arctic Ocean to the Labrador Sea, warming NA SSTs. Dowsett et al. (2019) also demonstrated an improved consistency between the proxy-based SST changes and model-predicted SST changes after closing these Arctic gateways in models. It is therefore likely that the multi-model mean SST change in the NA in PlioMIP2 has been influenced by the specified change in Arctic gateways leading to a regionally enhanced fit with proxy data. However, the question regarding the veracity of the specified changes in Arctic gateways in the PRISM4 reconstruction, given the uncertain and lack of geological evidence either way remains open and requires further study.

One of the clearest data/model inconsistencies occurs in the Benguela upwelling system, where proxy data indicates a larger degree of SST warming than the multi-model mean. The simulation of upwelling systems is particularly challenging for global numerical climate models due to the spatial scale of the physical processes involved, and the capability of models to represent changes in the structure of the water column (thermocline depth) as well as cloud/surface temperature feedbacks. Dowsett et al. (2013) noted SST discrepancies between the PRISM3D SST reconstruction and the PlioMIP1 ensemble. Their analysis of the seasonal vertical temperature profiles from PlioMIP1 for the Peru Upwelling region indicated that models produced a simple temperature offset between PI and the Pliocene, but did not simulate any change to thermocline depth.

An assumption that proxy-data truly reflect mean annual SSTs in upwelling regions is also worthy of consideration. In upwelling zones, nutrients (and relatively cold waters) are brought to the surface increasing productivity. The upwelling of nutrient rich waters if often seasonally modulated, which could conceivably bias alkenone-based SSTs to the seasonal maximum for nutrient supply and therefore coccolithophore productivity and/or alkenone flux. In the modern ocean during the most intense region of Benguela upwelling, the productivity seems to be year-round, whereas the southern Benguela has highest productivity during the summer (Rosell-Melé and Prahl, 2013). Ismail et al. (2015), based on observational data, demonstrated that it was surface heating, not vertical mixing related to upwelling, which controls the upper ocean temperature gradient in the region today. This lends some credence to the idea that the observed mismatch between PlioMIP2 ΔSST and the F&D19\_30 proxy-based anomaly could arise from the complexities/uncertainties associated with interpreting alkenone-based SSTs in the region as simply an indication of mean annual SST.

*5.3 Equilibrium Climate Sensitivity, Earth System Sensitivity and Pliocene climate change*

From the analysis shown in section 3.6 a strong relationship between ECS and the ensemble-simulated Pliocene temperature anomaly is discernible if the models with the largest and smallest temperature anomalies are omitted. That is not meant to convey criticism of those simulations, rather to highlight the fact that whether there is a clear relationship or not is simply dependent upon the individual model responses that comprise the overall model ensemble. This point is true for the globally average temperature anomaly, monthly averaged temperature anomaly, latitudinal average temperature anomaly and the gridbox based temperature anomaly. It highlights the benefit of multi-model ensembles over the analysis of singular model responses. Without the warmest or coldest model the relationship between ECS and the globally averaged temperature anomaly is signiﬁcant at the 95% level. The globally averaged winter Pliocene temperature anomaly is more strongly related to a models ECS than other seasons, and the tropical Pliocene temperature anomaly is more strongly related to a models ECS than other latitudes. On a gridpoint by gridpoint basis, the tropical oceans are most strongly related to ECS, highlighting the benefits for deriving estimates of ECS from a concentrated effort to reconstruct tropical SST response using the geological record.

The emergence of the concept of longer-term or Earth System Sensitivity can be at least partly attributed to the study of the Pliocene epoch (Lunt et al. 2010; Haywood et al., 2013a). However, as Hunter et al. (2019) state clearly, the comparison of ECS and ESS can only be robust if an assumption is made that the PlioMIP2 model boundary conditions are a good approximation to the equilibrated Earth system under a contemporary doubling of atmospheric CO2 concentration. Whilst this may appear to be a reasonable position now, since the changes in non-glacial elements of the PRISM4 palaeogeography are limited, there has been a clear move towards more radical thinking in terms of Pliocene palaeogeography from PlioMIP1 to PlioMIP2. Yet, within the bounds of plausible uncertainty, a larger number of additional palaeogeographic modifications remain possible for the Late Pliocene than were incorporated into the PRISM4 reconstruction (see Hill D.J. 2015 for further details), and which may have a bearing how well the Pliocene is seen to approximate an equilibrated modern Earth system in in the years ahead.

PlioMIP1 determined a range in the ESS/CS ratio of between 1.1 and 2, with a best estimate of 1.5. In PlioMIP2, which has benefited from the access to a larger array of models and new boundary conditions, the range in and best estimate for the ESS/CS ratio is 0.85 to 3.10 and 1.55 respectively. Thus, PlioMIP2 broadly supports the finding of PlioMIP1; that ESS is greater than ECS (47% PlioMIP1: 55% PlioMIP2).

**6. Conclusions**

The Pliocene Model Intercomparison Project Phase 2 represents one of the largest ensembles of climate models of different complexities and spatial resolution ever assembled to study a specific interval in Earth history. PlioMIP2 builds on the findings of PlioMIP1 and incorporates state-of-the-art reconstructions of Pliocene boundary conditions and new temporally consistent sea-surface temperature proxy data underpinning new data/model comparison. The major findings of the work include:

* Global annual mean Near Surface Air Temperatures increase by 1.4 to 5.1°C, with a multi-model average of 2.8°C.
* The multi-model mean annual total precipitation rate increases by 6%.
* The predicted anomaly between Pliocene and pre-industrial is statistically significant for Near Surface Air Temperature and Sea Surface Temperature in most gridboxes. The modelled precipitation anomaly is significant only in specific regions.
* The degree of polar amplification of surface temperature change is generally consistent with transient climate modelling experiments used to predict future climate.
* The land warms more than the oceans in a manner akin to future climate change simulations.
* As an ensemble, average warming is slightly biased towards the Northern Hemisphere summer/autumn although the annual cycle of temperature change is highly model dependent.
* Unlike simulations of 2100AD climate, the difference in the average warming between the hemispheres is subdued.
* Excluding specific models that represent extreme cases within the ensemble, there is a statistically significant relationship between Equilibrium Climate Sensitivity and Pliocene global annual average temperature change. The PlioMIP2 ensemble finds that Earth System Sensitivity is greater than Equilibrium Climate Sensitivity by a best estimate of 55%.
* There is no clear relationship between the simulated temperature and precipitation anomaly and the year of model release but there may be a significant relationship between model sensitivity and horizontal resolution.
* The PlioMIP2 ensemble appears to be broadly reconcilable with new temporally specific records of sea surface temperatures. Significant agreement between simulated and reconstructed temperature change is seen, with notable local signals of data/model disagreement. Differences between observed pre-industrial sea surface temperature data sets are enough to have a significant impact on how well models reproduce proxy-reconstructed ocean temperature changes.
* The closure of the Bering Strait and Canadian Archipelago gateways to the Arctic has led to an improvement in model predicted ocean temperature change in the North Atlantic.

**Acknowledgments**

We acknowledge the use of NOAA\_ERSST\_V5 data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/. AMH, JCT, AMD, SJH and DJH, acknowledge the FP7 Ideas: European Research Council (grant no. PLIO-ESS, 278636), the Past Earth Network (EPSRC grant no. EP/M008.363/1) and the University of Leeds Advanced Research Computing service. HJD and KMF acknowledge support from the USGS Climate Research and Development Program. This research used samples and/or data provided by the Ocean Drilling Program (ODP) and International Ocean Discovery Program (IODP). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

**References**

Burls, N.J., Fedorov, A.V. 2014. Simulating Pliocene warmth and a permanent El Nino-like state: The role of cloud albedo. Paleoceanography, 29, (10), 893-910.

Byrne, M.P., O'Gorman, P.A. 2013. Land-Ocean Warming Contrast over a Wide Range of Climates: Convective Quasi-Equilibrium Theory and Idealized Simulations. Journal of Climate, 26, (12), 4000-4016.

Corvec, S., Fletcher, C.G. 2017. Changes to the tropical circulation in the mid-Pliocene and their implications for future climate. Climate of the Past, 13, (2), 135-147.

Chandler, M., Rind, D., and Thompson, R. 1994. Joint investigations of the middle Pliocene climate II: GISS GCM Northern Hemisphere results. Global and Planetary Change, 9, (3-4), 197-219. 10.1016/0921-8181(94)90016-7.

De La Vega, E., Chalk, T.B., Foster, G.L., Bysani, R., Wilson, P.A. 2018. Warming in a warm world: Orbital CO2 forcing variations in the warm Pliocene. Abstract PP21C-1434 presented at the 2018 AGU Fall Meeting, San Francisco, Calif., 11 Dec.

Dowsett, H.J., Cronin, T.M., Poore, R.Z., Thompson, R.S., Whatley, R.C., and Wood, A.M. 1992. Micropaleontological Evidence for Increased Meridional Heat Transport in the North Atlantic Ocean during the Pliocene. Science, 258, (5085), 1133-1135. 10.1126/science.258.5085.1133.

Dowsett, H., Robinson, M., Haywood, A., Salzmann, U., Hill, D., Sohl, L., Chandler, M., Williams, M., Foley, K., Stoll, D. 2010. The PRISM3D paleoenvironmental reconstruction. Stratigraphy, 7, (2-3), 123-139.

Dowsett, H.J., Robinson, M.M., Haywood, A.M., Hill, D.J., Dolan, A.M., Stoll, D.K., Chan, W.L., Abe-Ouchi, A., Chandler, M.A., Rosenbloom, N.A., Otto-Bliesner, B.L., Bragg, F.J., Lunt, D.J., Foley, K.M., Riesselman, C.R. 2012. Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. Nature Climate Change, 2, (5), 365-371.

Dowsett, H.J., Foley, K.M., Stoll, D.K., Chandler, M.A., Sohl, L.E., Bentsen, M., et al. (2013). Sea Surface Temperature of the mid-Piacenzian Ocean: A Data-Model Comparison. Scientific Reports, 3, [2013]. <https://doi.org/10.1038/srep02013>.

Dowsett, H.J., Dolan, A.M., Rowley, D., Moucha, R., Forte, A.M., Mitrovica, J.X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K., Haywood, A.M. 2016. The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. Climate of the Past, 12, (7), 1519-1538.

Dowsett, H.J., Robinson, M.M., Foley, K.M., Herbert, T.D., Otto-Bliesner, B.L., and Spivey, W., 2019. The mid-Piacenzian of the North Atlantic Ocean. Stratigraphy, 16, 3, 119-144. 10.29041/strat.16.3.119-144.

Fedorov, A.V., Brierley, C.M., Lawrence, K.T., Liu, Z., Dekens, P.S., and Ravelo, A.C., 2013. Patterns and mechanisms of early Pliocene warmth: Nature, 496, (7443), 43-49.

Foley, K.M. and Dowsett, H.J. (2019). Community sourced mid-Piacenzian sea surface temperature (SST) data: US Geological Survey data release: <https://doi.org/10.5066/P9YP3DTV>.

Hargreaves, J.C., and Annan, J.D., 2016. Could the Pliocene constrain the equilibrium climate sensitivity? Climate of the Past, 12, (8), 1591-1599. 10.5194/cp-12-1591-2016.

Haywood, A.M., Valdes, P.J., Sellwood, B.W. 2000. Global scale palaeoclimate reconstruction of the middle Pliocene climate using the UKMO GCM: initial results. Global and Planetary Change, 25, (3-4), 239-256.

Haywood, A.M., Valdes, P.J. 2004. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. Earth and Planetary Science Letters, 218, (3-4), 363-377.

Haywood, A.M., Dowsett, H.J., Otto-Bliesner, B., Chandler, M.A., Dolan, A.M., Hill, D.J., Lunt, D.J., Robinson, M.M., Rosenbloom, N., Salzmann, U., Sohl, L.E. 2010. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1). Geosci. Mod. Dev. 3, (1), 227-242.

Haywood, A.M., Dowsett, H.J., Robinson, M.M., Stoll, D.K., Dolan, A.M., Lunt, D.J., Otto-Bliesner, B., Chandler, M.A. 2011. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 2). Geosci. Mod. Dev. 4, 571-577.

Haywood, A.M., Hill, D.J., Dolan, A.M., Otto-Bliesner, B.L., Bragg, F., Chan, W.L., Chandler, M.A., Contoux, C., Dowsett, H.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Abe-Ouchi, A., Pickering, S.J., Ramstein, G., Rosenbloom, N.A., Salzmann, U., Sohl, L., Stepanek, C., Ueda, H., Yan, Q., Zhang, Z. 2013a. Large-scale features of Pliocene climate: Results from the Pliocene Model Intercomparison Project. Climate of the Past, 9, (1), 191-209.

Haywood, A.M., Dolan, A,M., Pickering, S.J., Dowsett, H.J., McClymont, E.L., Prescott, C.L., Salzmann, U., Hill, D.J., Hunter, S.J., Lunt, D.J., Pope, J.O., Valdes, P.J. 2013b. On the identification of a Pliocene time slice for data-model comparison. Philos Trans A Math Phys Eng Sci. 371, (2001), 20120515.

Haywood, A.M., Dowsett, H.J., Dolan, A.M., Rowley, D., Abe-Ouchi, A., Otto-Bliesner, B., Chandler, M.A., Hunter, S.J., Lunt, D.J., Pound, M., Salzmann, U. 2016. The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: Scientific Objectives and Experimental Design. Climate of the Past, 12, (3), 663-675.

Hill, D.J., Haywood, A.M., Lunt, D.J., Hunter, S.J., Bragg, F.J., Contoux, C., Stepanek, C., Sohl, L., Rosenbloom, N.A., Chan, W.L., Kamae, Y., Zhang, Z., Abe-Ouchi, A., Chandler, M.A., Jost, A., Lohmann, G., Otto-Bliesner, B.L., Ramstein, G., Ueda, H. 2014. Evaluating the dominant components of warming in Pliocene climate simulations. Climate of the Past, 10, (1), 79-90.

Hill, D.J. 2015. The non-analogue nature of Pliocene temperature gradients. Earth and Planetary Science Letters. 425, 232-241.

Hind, A., Zhang, Q., Brattstrom, G. 2016. Problems encountered when defining Arctic amplification as a ratio. Scientific Reports, 6, article 30469.

Huang, B.Y., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S, Zhang, H.M. 2017. Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. Journal of Climate, 30, (20), 8179-8205.

Hunter, S.J., Haywood, A.M., Dolan, A.M., Tindall, J.C. 2019. The HadCM3 contribution to PlioMIP phase 2. Climate of the Past, 15, (5), 1691-1713.

Howell, F.W., Haywood, A.M., Otto-Bliesner, B.L., Bragg, F.J., Chan, W-L, Chandler, M.A., Contoux, C., Kamae, Y., Abe-Ouchi, A., Rosenbloom, N.A., Stepanek, C., Zhang, Z. 2016. Arctic sea ice simulation in the PlioMIP ensemble. Climate of the Past, 12 (3), 749-76.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

Ismail, H.E., Agenbag, J.J., de Villiers, S., Ximba, B.J. 2015. Relation between upwelling intensity and the variability of physical and chemical parameters in the southern Benguela upwelling system. International Journal of Oceanography 2015: 510713. doi: [10.1155/2015/51071310.1155/2015/510713](https://doi.org/10.1155/2015/510713)10.1155/2015/510713.

Lambert, F.H., Webb, M.J., Joshi, M.M. 2011. The Relationship between Land-Ocean Surface Temperature Contrast and Radiative Forcing. Journal of Climate, 24, (13), 3239-325624.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M., and Levrard, B. 2004. A long-term numerical solution for the insolation quantities of the Earth: Astronomy and Astrophysics, 428, 261-285.

Li, X.Y., Jiang, D.B., Zhang, Z.S., Zhang, R., Tian, Z.P., Yan, Q. 2015. Mid-Pliocene westerlies from PlioMIP simulations. Adv. Atmos. Sci., 32, (7), 909-923.

Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., Dowsett, H.J. 2010. Earth system sensitivity inferred from Pliocene modelling and data. Nature Geoscience, 3, (1), 60-64.

Lunt, D.J., Dunkley Jones, T., Heinemann, M., Huber, M., LeGrande, A., Winguth, A., Loptson, C., Marotzke, J., Roberts, C.D., Tindall, J., Valdes, P., Winguth, C. 2012. A model-data comparison for a multi-model ensemble of early Eocene atmosphere-ocean simulations: EoMIP. Climate of the Past, 8, (5), 1717-1736.

Mba, W.P., Longandjo, G.N.T, Moufouma-Okia, W., Bell, J.P., James, R., Vondou, D.A., Haensler, A., Fotso-Nguemo, T.C., Guenang, G.M., Tchotchou, A.L.D., Kamsu-Tamo, P.H., Takong, R.R., Nikulin, G., Lennard, C.J., Dosio, A. 2018. Consequences of 1.5 degrees celsius and 2 degrees celsius global warming levels for temperature and precipitation changes over Central Africa. Environmental Research Letters, 13, (5), 055011, 10.1088/1748-9326/aab048.

Nikulin G, Lennard C, Dosio A, Kjellstrom E, Chen YM, Hansler A, Kupiainen M, Laprise R, Mariotti L, Maule CF, van Meijgaard E, Panitz HJ, Scinocca JF, Somot S. 2018. The effects of 1.5 and 2 degrees of global warming on Africa in the CORDEX ensemble. Environmental Research Letters. 13(6) 065003.

Otto-Bliesner, B., Jahn, A., Feng, R., Brady, E., Hu, A., and Löfverström, M.. 2017. Changes in Arctic Gateways Amplify North Atlantic Warming in the Late Pliocene: Arctic Gateways and Pliocene Climate: Geophysical Research Letters, 44, 957-964. 10.1002/2016GL071805.

Prescott, C.L., Haywood, A.M., Dolan, A.M., Hunter, S.J., Pope, J.O., Pickering, S.J. 2014. Assessing orbitally-forced interglacial climate variability during the mid-Pliocene Warm Period. Earth and Planetary Science Letters, 400, 261-271.

Rind, D., and Chandler, M.A. 1991. Increased ocean heat transports and warmer climate. Journal of Geophysical Research, 96, 7437-7461. 10.1029/91JD00009.

Rosell-Mele, A., Prahl, F.G. 2013. Seasonality of U-37(K)' temperature estimates as inferred from sediment trap data. Quaternary Science Reviews, 72, 128-136.

Russell, J.L., K.W. Dixon, A. Gnanadesikan, R.J. Stouffer, and J.R. Toggweiler, 2006. The Southern Hemisphere Westerlies in a Warming World: Propping Open the Door to the Deep Ocean. J. Climate, 19, (24), 6382-6390.

Salzmann, U., Dolan, A.M., Haywood, A.M., Chan, W-L, Voss, J., Hill, D.J., Abe-Ouchi, A., Otto-Bliesner, B., Bragg, F.J., Chandler, M.A., Contoux, C., Dowsett, H.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Pickering, S.J., Pound, M.J., Ramstein, G., Rosenbloom, N.A., Sohl, L., Stepanek, C., Ueda, H., Zhang, Z. 2013. Challenges in reconstructing terrestrial warming of the Pliocene revealed by data-model discord. Nature Climate Change, 3, 969-974.

Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J., Hill, D.J. 2008. A new global biome reconstruction and data-model comparison for the Middle Pliocene. GLOBAL ECOL BIOGEOGR, 17, (3), 432-447.

Sloan, L.C., Crowley, T.J., and Pollard, D. 1996, Modeling of middle Pliocene climate with the NCAR GENESIS general circulation model: Marine Micropaleontology, 27, (1-4), 51-61. 10.1016/0377-8398(95)00063-1.

Stouffer, R.J., S. Manabe, and K. Bryan. 1989. Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO2. Nature, 342, 660-662.

Sun, Y., Ramstein, G., Contoux, C., Zhou, T. (2013). A comparative study of large-scale atmospheric circulation in the context of future scenario (RCP4.5) and past warmth (Mid Pliocene). Climate of the Past. 9. 10.5194/cpd-9-1449-2013.

Tierney, J.E., Haywood, A.M., Feng, R., Bhattacharya, T., Otto‐Bliesner, B.L. 2019. Pliocene Warmth Consistent with Greenhouse Gas Forcing. Geophysical Research Letters, 46, (15), 9136-9144.

Zhang, R., Yan, Q., Zhang, Z.S., Jiang, D., Otto-Bliesner, B.L., Haywood, A.M., Hill, D.J., Dolan, A.M., Stepanek, C., Lohmann, G., Contoux, C., Bragg, F., Chan, C.L., Chandler, M.A., Jost, A., Kamae, Y., Abe-Ouchi, A., Ramstein, G., Rosenbloom, N.A., Sohl, L., Ueda, H. 2013. Mid-Pliocene East Asian monsoon climate simulated in the PlioMIP. Climate of the Past, 9, (5), 2085-2099.

Zhang, R., Zhang, Z.S., Jiang, D.B., Yan, Q., Zhou, X., Cheng, Z.G. 2016. Strengthened African summer monsoon in the mid-Piacenzian. Advances in Atmospheric Sciences, 33, (9), 1061-1070.

Zhang, Z.S., Nisancioglu, K.H., Chandler, M.A., Haywood, A.M., Otto-Bliesner, B.L., Ramstein, G., Stepanek, C., Abe-Ouchi, A., Chan, W.L., Bragg, F.J., Contoux, C., Dolan, A.M., Hill, D.J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D.J., Rosenbloom, N.A., Sohl, L.E., Ueda, H. 2013. Mid-Pliocene Atlantic Meridional Overturning Circulation not unlike modern. Climate of the Past, 9, (4), 1495-1504.

END.