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

Alan M. Haywood[1], Julia C. Tindall[1*], Harry J. Dowsett[2], Aisling M. Dolan[1], Kevin M. Foley[2], Stephen J. Hunter[1], Daniel J. Hill[1], Wing-Le Chan[3], Ayako Abe-Ouchi[3], Christian Stepanek[4], Gerrit Lohmann[4], Deepak Chandan[5], W. Richard Peltier[5], Ning Tan[6/7], Camille Contoux[7], Gilles Ramstein[7], Xiangyu Li[8/9], Zhongshi Zhang[8/9/10], Chuncheng Guo[9], Kerim H. Nisancioglu[9], Qiong Zhang[11], Qiang Li[11], Youichi Kamae[12], Mark A. Chandler[13], Linda E. Sohl[13], Bette L. Otto-Bliesner[14], Ran Feng[15], Esther C. Brady[14], Anna S. von der Heydt[16,17], Michiel L. J. Baatsen[17] and Daniel J. Lunt[18].

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

[2]Florence Bascom Geoscience Center, U.S. Geological Survey, Reston, VA 20192, USA

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

[4]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

[5]Department of Physics, University of Toronto, Toronto, Ontario, Canada

[6]Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China.

[7]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

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

[9]NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

[10]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan, China

[11]Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

[12]Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

[13]CCSR/GISS, Columbia University, New York, USA

[14]National Center for Atmospheric Research, Boulder, Colorado, USA

[15]Department of Geosciences, College of Liberal Arts and Sciences, University of Connecticut, Connecticut, USA

[16]Centre for Complex Systems Science, Utrecht University, Utrecht, The Netherlands

[17]Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, The Netherlands.

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

Correspondence to: Julia C. Tindall (earjcti@leeds.ac.uk)

Abstract. The Pliocene epoch has great potential to improve our understanding of the long-term climatic and environmental consequences of an atmospheric CO_2 concentration near ~400 parts per million by volume. Here we present the large-scale features of Pliocene climate as simulated by a new ensemble of climate models of varying complexity and spatial resolution and based on new reconstructions of boundary conditions (the Pliocene Model Intercomparison Project Phase 2; PlioMIP2). As a global annual average, modelled surface air temperatures increase by between 1.4 and 4.7°C relative to pre-industrial with a multi-model mean value of 2.8°C. Annual mean total precipitation rates increase by 6% (range: 2%-13%). On average, surface air temperature (SAT) increases are 1.3°C greater over the land than over the oceans, and there is a clear pattern of polar amplification with warming polewards of 60°N and 60°S exceeding the global mean warming by a factor of 2.4. In the Atlantic and Pacific Oceans, meridional temperature gradients are reduced, while tropical zonal gradients remain largely unchanged. Although there are some modelling constraints, there is a statistically significant relationship between a model's climate response associated with a doubling in CO_2 (Equilibrium Climate Sensitivity; ECS) and its simulated Pliocene surface temperature response. The mean ensemble earth system response to doubling of CO_2 (including ice sheet feedbacks) is approximately 50% greater than ECS, consistent with results from the PlioMIP1 ensemble. Proxy-derived estimates of Pliocene sea-surface temperatures are used to assess model estimates of ECS and indicate a range in ECS from 2.5 to 4.3°C. This result is in general accord with the range in ECS presented by previous IPCC Assessment Reports.

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 since 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 (PlioMIP1) proposed a single set of model boundary conditions based on the U. S. 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 several publications analysing diverse 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 surface air temperature was found to have increased compared to the preindustrial, with models showing warming of between 1.8 and

3.6°C. The warming was predicted at all latitudes but showed a clear pattern of polar amplification resulting in a reduced equator to pole surface temperature gradient. Modelled sea-ice responses were studied by Howell et al. (2016), who demonstrated a significant decline in Artic sea-ice extent, with some models simulating a seasonally 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 also a weaker 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 a response to direct CO_2 forcing, while at 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.

The PlioMIP1 ensemble was also used to help constrain Equilibrium Climate Sensitivity (ECS; Hargreaves and Annan 2016). ECS is defined as the global temperature response to a doubling of CO_2, once the energy balance has reached equilibrium (this diagnostic is discussed further in section 2.4). Based on the PRISM3 compilation of MP tropical SSTs, Hargreaves and Annan (2016) estimated that ECS is between 1.9 and 3.7°C. In addition, they used the PlioMIP1 model ensemble to estimate Earth System Sensitivity (ESS). ESS is defined as the temperature change associated with a doubling of CO_2 and includes all ECS feedbacks along with long timescale feedbacks such as those involving ice sheets. In PlioMIP1, ESS was estimated to be a factor of 1.47 higher than the ECS (ensemble mean ECS = 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. Although 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, in particular in the high North Atlantic (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 boundary conditions, uncertainty in the interpretation of proxy data and uncertainty in model physics (for example, recent studies have demonstrated that this model-proxy mismatch has been reduced by including explicit aerosol-cloud interactions in the newer generations of models (Sagoo and Storelvmo 2017; Feng et al., 2019)).

These findings substantially influenced the experimental design for the second Phase of PlioMIP (PlioMIP2). Specifically, PlioMIP2 was developed to (a) reduce uncertainty in model boundary conditions and (b) reduce uncertainty in proxy data reconstruction. In order to accomplish (a), state-of-the-art approaches were adopted to generate an entirely new 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, Otto-Bliesner et al. 2017). These include the Bering Strait and Canadian Archipelago becoming sub-aerial and modification of the land/sea mask in the Indonesian/Australian region for the emergence of the Sunda and Sahul Shelf. In order to achieve (b) it was necessary to move away from time-averaged 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 and ice sheet configuration 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, 3.205Ma). The new PRISM4 (Pliocene Research, Interpretation and Synoptic Mapping version 4) global community-sourced data set of high temporal resolution SSTs (Foley and Dowsett, 2019) targets the same interval in order to produce 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 *midPliocene-eoi400)* 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 and 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 *midPliocene-eoi400* experiment (for wider accessibility this experiment will hereafter be 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 were required to be initialised with pre-industrial vegetation and spun-up until an equilibrium vegetation distribution is reached. The concentration of atmospheric CO_2 for experiment $Plio_{Core}$ was set at 400 parts per million by volume (ppmv), a value almost identical to that chosen for the PlioMIP1 experimental design (405 ppmv), and in line with the very latest high-resolution proxy reconstruction of atmospheric CO_2 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's PI_{Ctrl} experiment. The Greenland ice sheet (GIS) was confined to high elevations in the Eastern Greenland Mountains, covering an area approximately 25% of the present-day GIS. The PlioMIP2 Antarctic ice sheet configuration is the same as PlioMIP1 and has no ice over Western Antarctica. The reconstructed PRISM4 ice sheets have a total volume of 20.1×10^6 km^3, equating to a sea-level increase relative to present day of less than ~24 m (Dowsett et al. 2016; this volume). Integration length was set to be 'as long as possible', or a minimum of 500 simulated years. 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 to comprise the PlioMIP2 ensemble. This ensemble is almost double the size of that presented in the PlioMIP1 large-scale features publication (Haywood et al. 2013a). Summary details of the included models and model physics along with information regarding the implementation of PRISM4 boundary conditions and each model's ECS can be found in Table 1 and Supplementary Table 1. Each group uploaded the final 100 years of each simulation for analysis. These were then regridded onto a regular 1° × 1° grid using a bilinear interpolation, to enable each model to be analysed in the same way. Means and standard deviations for each model were then calculated across the final 50 years.

2.4 Equilibrium Climate Sensitivity (ECS) and Earth System Sensitivity (ESS)

In Section 3.6 we will use the *Plio*$_{Core}$ and *PI*$_{Ctrl}$ simulations to investigate ECS and ESS. The *Plio*$_{Core}$ experiments represent a 400ppmv world that is in quasi-equilibrium with respect to both climate and ice-sheets and hence represents an 'Earth System' response to the 400ppmv CO_2 forcing. The 'Earth System' response to a doubling of CO_2 (ie 560ppmv-280ppmv; ESS) can then be estimated as follows:

$$ESS = \frac{ln\frac{560}{280}}{ln\frac{400}{280}}(Plio_{Core}[SAT] - PI_{Ctrl}[SAT]) \qquad [1]$$

Strictly speaking, there will be small errors in the estimate of ESS from the above equation. These are due to changes between *Plio*$_{Core}$ and *PI*$_{Ctrl}$ which should not be included in estimates of ESS, such as: land-sea mask changes, topographic changes, changes in soil properties and lake changes. However, all these additional changes are minimal compared to the ice sheet and GHG changes and are expected to have only a negligible impact on the globally averaged temperature and therefore the estimate of ESS. For example, Pound et al. 2014 found that the inclusion of Pliocene soils and lakes would have an insignificant effect on global temperature (even though regional changes could be important).

To assess the relationship between the ESS estimated from the ensemble and a model's reported ECS, we will correlate the *Plio*$_{Core}$ - *PI*$_{Ctrl}$ temperature anomaly with each model's reported ECS on a range of temporal and spatial scales. A strong

correlation at a location would suggest that MP proxy data at that location could be used to derive a proxy-data constrained estimate of ECS, while a weak correlation would suggest that proxy-data at that location could not be used in ECS estimates.

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 4.7 °C (Fig. 1a; Table 2), with an ensemble mean ΔT of 2.8°C. Increased SATs are predicted throughout the globe (Fig. 1b), with an ensemble average warming of ~1.7°C for the tropical oceans (20°N-20°S), which amplifies towards the high latitudes (Figs. 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 magnitude, the CCSM4-2deg model has the greatest apparent sensitivity to imposing MP boundary conditions with a simulated ΔT of 4.7°C (Fig 1a). This is higher than CCSM4-UoT (ΔT = 3.8°C) and CCSM4-1deg (ΔT = 2.6°C). A notable difference between these simulations is the response in the Southern Hemisphere (SH) at ~70°S, where the zonal mean warming in the CCSM4-2deg simulation is over 5°C higher than in the CCSM4-UoT and CCSM4-1deg simulation (Fig. 1c). Supplementary table 1 shows that even though the CCSM4 models differ in their response they all appear to be close to equilibrium. In addition, they are all reported to have the same ECS (table 1) and they all have the same physics apart from changes to the standard ocean model in the CCSM4-UoT simulations and the $Plio_{Core}$ CCSM4-2deg simulation. These changes (discussed by Chandan & Peltier, 2017, this volume) are: 1. the vertical profile of background diapycnal mixing has been fixed to a hyperbolic tangent form, and 2. tidal mixing as well as dense water overflow parameterization schemes have been turned off. Although the exact cause of the differences in ΔT between the CCSM4 models remains unclear, the changes in the ocean parameterisations and differences in initialization may contribute to the ΔT differences, in particular the changes in ocean mixing between different versions of the model (Fedorov et al., 2010).

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 Fig. 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). The anomaly is said to be "robust" if 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 are hatched in Fig. 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 surface air temperature, land/sea temperature contrasts and polar amplification

The seasonal cycle of the SAT anomaly for each model is presented in Fig. 3a. Overall, the ensemble mean displays a small enhancement in SAT warming in the Northern Hemisphere (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. CCSM4-2deg, CCSM4-1deg, CCSM4-UoT, CESM1.2, HadCM3, MIROC4), 2) those that show no clear seasonal SAT amplification (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 whether 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.

The ensemble results for land/sea temperature contrasts clearly indicate a greater warming over land than over the oceans (Fig. 3b). This result also holds when only the land/sea temperature contrast in the tropics is considered. The only exception is the EC-Earth3.1 model that predicts that the oceans warm slightly more than the land in the tropics and show no clear divergence between land/sea warming averaged over the globe.

Whether or not the NH warms more than the Southern Hemisphere (SH) 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 CCSM4-2deg has a large warming at ~70°S which contributes to this model simulating a greater degree of warming in the SH versus the NH.

Polar amplification (PA) is defined as the ratio of polar warming (poleward of 60° in each hemisphere) to global mean warming (Smith et al 2019). The PA for each model for the NH and the SH is shown in Fig. 3d. All models show PA > 1, for both hemispheres, although whether there is more PA in the NH or SH is a model dependent feature. The ensemble average PA is 2.4 and is similar in the NH and the SH.

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; Tierney et al., 2019).

The multi-model mean meridional profile of SSTs for the Atlantic Ocean is shown in Fig. 4a. In the tropics and sub-tropics, the SST increase between the $Plio_{Core}$ and PI_{Ctrl} experiments is 1.2-2.0°C. This difference increases to ~4.0°C in the NH at ~55°N, where the range in warming across the models also increases to between 0°C and 9°C. The Pliocene and Pre-industrial meridional SST profiles in the Pacific are similar to that of the Atlantic, with little indication from the multi-model mean for a high latitude enhancement in meridional temperature (Fig. 4b). However, a large range in the ensemble response is noted.

In the tropical Atlantic (20°N -20°S) the multi-model mean zonal mean SST for the Pliocene increases by ~1.4°C (ensemble range from 0.1°C to 2.4°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 that are associated with upwelling (Fig. 4d). As such, a clear east-west temperature gradient is evident in the Pliocene tropical Pacific in the PlioMIP2 ensemble (similar to PlioMIP1) and none of the models show a SST anomaly that would be consistent with a permanent El-Niño (see Supplementary Fig. S2). 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 and found that the proxy gradient was within the range of previous model estimates (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 Fig. S3). Supplementary Fig. S2 shows the difference between the Pliocene ΔSST for each model in the PlioMIP2 ensemble and the Pliocene ΔSST of the multi-model mean. This illustrates that despite the significance of the climate anomaly across the ensemble there are many regions (e.g. Southern

Ocean, North Atlantic Ocean, Arctic Ocean) where there is large inter-model spread of the magnitude of the Pliocene SST anomalies.

3.4 Total precipitation rate

Simulated increases in $Plio_{Core}$ global annual mean precipitation rates, compared to each contributing model's PI_{Ctrl} experiment, (hereafter referred to as ΔPrecip) ranges from 0.07 to 0.37 mm/day (Fig. 5a), which is notably larger than the PlioMIP1 range of 0.09-0.18 mm/day. The PlioMIP2 ensemble mean ΔPrecip is 0.17 mm/day. The spatial pattern (Fig. 5b) shows enhanced precipitation over high latitudes and reduced precipitation over parts of the subtropics. The largest ΔPrecip is found in the tropics, 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 modelling results that have demonstrated a weakening in Hadley Circulation linked to reduced pole to equator temperature gradient (e.g. Corvec and Fletcher 2017). Greenland shows increased $Plio_{Core}$ precipitation in regions that have become deglaciated and are therefore substantially warmer. Latitudes associated with the westerly wind belts also show enhanced $Plio_{Core}$ precipitation, 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 ΔPrecip appears 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 ΔPrecip (CCSM4-2deg, 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 of PlioMIP1 (Haywood et al., 2013a) and is likely because more precipitation falls in the tropics rather than extratropics, and therefore the inter-model differences are larger in the tropics. The methodology of Mba et al. (2018) and Nikulin et al. (2018) (described in section 3.1), was used to determine the robustness of ΔPrecip (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 ensemble precipitation signal is uncertain. Changes in precipitation rates in the subtropics have some intermodel coherence in many places because at least 12 of the 15 models agree on the sign of change. However, these predicted changes are not robust because the magnitude of change is not large compared to the standard deviation seen in the ensemble (Fig. 5c), this is consistent with results from CMIP5, which show predicted precipitation changes have low confidence particularly in the low and medium emissions scenarios (IPCC, 2013). 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 Fig. S4 shows the difference between each model's ΔPrecip and the multi-model mean ΔPrecip (shown in Fig. 5b), highlighting that there is uncertainty in the ensemble with respect to the regional patterns of precipitation change.

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 dependent, although the ensemble average shows a small NH late spring to autumn $Plio_{Core}$ enhancement in precipitation (Fig. 6a). This is evident in the models CCSM4-2deg, CCSM4-1deg, CESM1.2, CCSM4-UoT and IPSLCM6A. Most other models have no clear structure to the monthly distribution of precipitation changes, other than the GISS2.1G model that simulates the NH late spring to autumn ΔPrecip being supressed compared to the rest of the year. An increase in NH summer precipitation is consistent with a general trend of West African, Indian and East Asia Summer monsoon enhancement, and this will be discussed in detail in a forthcoming PlioMIP2 paper.

In terms of the land/sea ΔPrecip contrast the ensemble divides into two groups (Fig. 6b). One in which a clear pattern of precipitation anomaly enhancement over land compared to the oceans is seen (CCSM4-1deg, CCSM4-2deg, 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, COSMOS, GISS2.1G, IPSLCM6A, IPSLCM5A2, IPSLCM5A, MRI-CGCM2.3). Most models suggest that precipitation is more enhanced over land than over ocean. A clear exception to this rule is the COSMOS, where ocean regions receive slightly more precipitation than continents.

3.6 Climate and Earth System Sensitivity

This section will consider the relationship between ECS and ESS across the ensemble. Table 2 shows the ECS for each model (referenced in table 1) and the ESS estimated from the $Plio_{Core} - PI_{Ctrl}$ temperature anomaly (equation 1). The mean ESS / ECS ratio is 1.55, suggesting that the ESS based on the ensemble is 55% larger than the ECS. While most of the models show an ESS / ECS ratio similar to the mean (the ESS / ECS ratio intermodel standard deviation is 0.5), some of the models show very different values.

Of particular note is the EC-Earth3.1 model which has ESS / ECS = 0.83. This means that the ESS (calculated from $Plio_{Core}$ and equation 1) is less than the model's published ECS. Since ESS is expected to be larger than ECS, this suggests that there are some inconsistencies between the way in which ECS and ESS were obtained. Because of these inconsistencies our analysis of the relationship between ECS and ESS will exclude this model.

There also appear to be inconsistencies between the CCSM4 family of models. The ECS for CCSM4-UoT, CCSM4-2deg and CCSM4-1deg are all reported to be the same as that obtained by Bitz et al. 2012 (i.e. ECS = 3.2°), yet the ESS for each

of these models is substantially different (7.3°C, 9.1°C, 5.1°C). We suggest that each of these models provides a different but equally valid realisation of ESS based on the $Plio_{Core}$ experiments, and that the difference is due to the way in which the models have been set up for $Plio_{Core}$ (see supplementary table 1 and section 3.1 for more details). We therefore use the average ESS from CCSM4 models (CCSM4-avg in table 2) in our analysis of the relationship between ECS and ESS in preference to using ESS from individual simulations. The ESS / ECS ratio from CCSM4-avg is 2.24, which is larger than the ESS / ECS value across the ensemble.

The first analysis of how ECS relates to ESS will consider the correlation between ECS and the globally averaged $Plio_{Core}$ - PI_{Ctrl} temperature anomaly (Fig. 7a). There is a significant relationship between the two at the 95% confidence level ($p=0.03$). Note, however, that including the model with ESS < ECS (EC-Earth3.1), and the individual CCSM4 models would not give a significant relationship. This highlights the fact that there must be consistency between model simulations in order to relate quantities between different models.

Next, we investigate how a model's ECS is related to the $Plio_{Core} - PI_{Ctrl}$ SAT anomaly on different temporal and spatial scales. In the analysis that follows we will simply assess whether such a correlation exists and if so, how strong it is, by looking at *p-values* and *R-squared* values, calculated from the models in the ensemble. Fig. 7b shows the relationship (*p-value* – blue, *R-squared* - red) between the annual mean ECS and the Pliocene SAT anomaly for each month. There is a significant correlation ($p < 0.05$) from August through to March. This means that if we want to develop an emergent constraint to use mPWP proxy data to estimate ECS it would be better to use proxies that represent August-March instead of proxies representing April-July.

Further examination looked at which latitudes the $Plio_{Core} - PI_{Ctrl}$ SAT anomaly is most closely related to the global ECS (Fig. 7c). We find a significant relationship ($p < 0.05$) between ECS and the zonal mean Pliocene temperature anomaly throughout most of the tropics. This relationship becomes significant at the 99% confidence level ($p < 0.01$) between 23°N and 12°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 point based $Plio_{Core} - PI_{Ctrl}$ SAT anomaly was assessed. This follows a similar approach to that of Hargreaves and Annan (2016). In Fig. 7d colours show the *R-squared* correlation between a model's global ECS and the point based $Plio_{Core} - PI_{Ctrl}$ SAT anomaly. The regions where the relationship between the two is significant at the 95% confidence level is hatched. The relationship between ECS and the grid box $Plio_{Core} - PI_{Ctrl}$ SAT anomaly is significant over large parts of the tropics (e.g. the tropical Pacific and Northern Africa), and over Greenland and parts of Antarctica. 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.

Of the models used in this section, the multi-model mean ESS is 5.4°C, and the ESS/ECS ratio is 1.53. 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 owing to the comparison between time averaged PRISM3D SST and SAT data, and climate model representations 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 to either side of the peak of MIS KM5c) and the other covers 30,000 years (up to 15,000 years to either side of the peak; this latter dataset will hereafter be referred to as 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 from observed pre-industrial SSTs derived for years 1870-1899 of the NOAA ERSST version 5 data set (Huang et al., 2017; Fig. 8a and Fig. 8b). Fig. 8c shows the proxy data ΔSST minus the multi-model mean ΔSST. Using the multi-model mean results, 17 of the 37 sites show a difference in model/data ΔSST of no greater than +/- 1°C (Fig. 8c). These are located mostly in the tropics, but also includes sites in the North Atlantic, along the coastal regions of California and New Zealand and in the North Pacific. In terms of discrepancies, the clearest and most consistent signal comes from the Benguela upwelling system (off the south west coast of Africa) 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 to 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. However, the Benguela region show greater model-data agreement when considering absolute SSTs than when considering anomalies. Furthermore, a somewhat clearer picture emerges of the model ensemble not producing SSTs that are warm enough in 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 ΔSAT shown by the PlioMIP2 ensemble (from 1.4 to 4.7°C) is akin to the best estimate (and uncertainty bounds) of predicted global temperature change by 2100CE using the RCP2.6 to 8.5 scenarios (RCP2.5 = 1.0 ± 0.4°C and RCP8.5 = 3.7 ± 0.7 °C IPCC, 2013; table 12.2). The ensemble multi-model mean response of 2.8°C is similar to the expected 2100 CE response predicted under RCP6.0 (2.2 ± 0.5°C). Comparing the degree of Pliocene temperature change to predicted changes at 2300 CE, the multi-model mean SAT change is most akin to RCP4.5 (2.5 ± 0.6°C) with the ensemble within the range of RCP2.6 (0.6± 0.3°C) and RCP8.5 (7.8 ± 2.9°C).

Studies have suggested that the Arctic temperature response to a doubling of atmospheric CO_2 concentration change may be 1-3 times that of the global annual mean temperature response (Hind et al., 2016). Fourteen out of the fifteen models within the PlioMIP2 ensemble simulate a polar amplification factor (PA; averaged over the NH and SH) of ≤ 3 (meaning that the high latitude temperature increase is less than 3 times the global mean temperature increase), only EC-Earth3.1 shows a PA > 3. However, 2 further models (GISS2.1G and NorESM-L) show PA > 3 in the SH. An important caveat 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 (Byrne and O'Gorman, 2013) 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, in reality the rise is non-monotonic and is regulated by latitudinal and regional variations in the availability of soil moisture that influences 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, 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 whether we are considering a transient (RCP-like) 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 extratropics. 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 as ocean heat transport mechanisms (Russell et al., 2006). Within the PlioMIP2 ensemble, 9 out of 15 models show a larger temperature change in the NH extratropics than the SH extratropics (Fig. 3c). This can in part be explained by the area of land in the NH being larger than in the SH 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 has 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 RCP-like 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 2.8°C increase in multi-model mean temperature is associated with a 6% increase in global annual mean precipitation. According to the Clausius-Clapeyron equation, the water holding capacity of the atmosphere increases by about 7% for each 1°C of temperature increase. The increase in precipitation is therefore less than would be expected if it were assumed that all aspects of the hydrological cycle remained the same as pre-industrial. This is in 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, but sluggish moist convection (Held and Soden, 2006).

A particularly robust feature of precipitation change across the ensemble is over the modern Sahara Desert and over the Asian monsoon region (figure 3d). These regions also experience enhanced precipitation under the RCP8.5 scenario for 2100 (IPCC, 2013; figure SPM.7). However, in other tropical and subtropical regions the $Plio_{Core}$ model response is small compared to current interannual variations.

Corvec and Fletcher (2017) showed that in PlioMIP1 studies the tropical overturning circulations in the mPWP were weaker than preindustrial simulations, while Sun et al. (2013) showed that both Hadley Cells expanded polewards, a result consistent with (but weaker than) the RCP4.5 scenario. These changes in circulation are consistent with the expansion of the subtropical highs and the corresponding reduction in subtropical oceanic precipitation seen in figure 5 for the $Plio_{Core}$ ensemble and in IPCC, 2013; figure SPM.7 for RCP scenarios at year 2100.

Although there are many similarities in tropical atmospheric circulation response between Pliocene experiments and the RCP future climate change experiments, there are specific differences mainly relating to a) the ice sheet changes and their effects on the equator-to-pole temperature gradient changes during the Pliocene vis-à-vis the future, and b) the fixed vs transient GHG changes. Nonetheless the similarities between the general features of the Pliocene experiments and future experiments continues to support the use of the Pliocene as one of the best geological analogues for the near future (Burke et al. 2018).

Across the ensemble, model sensitivity to Pliocene boundary conditions does not appear to correlate with the release date of the model (figure S6; i.e. older models are not demonstrably less sensitive than newer models). An example of this is the GISS2.1G model, which was released in 2019, yet has one of the smallest $Plio_{Core} - PI_{Crtl}$ anomalies for both temperature and precipitation. Within model families, however, some hint of a correlation can be seen. For example, IPSLCM6A (2018) is more sensitive than IPSLCM5A2 (2017) and IPSLCM5A (2010). The CESM1.2 model (release date 2013) is more sensitive than CCSM4-1deg (release date 2011), when both are run with the same resolution, boundary and initial conditions. However, other models from the CCSM4 family (CCSM4-UoT and CCSM4-2deg; both with release date 2011) are the most

sensitive of this family, suggesting that model sensitivity is more strongly related to parameterisation choices and initial condition choices than the release date of the model.

Across the ensemble there is no clear correlation between sensitivity and model resolution (Supplementary Fig. S6), however this is largely due to the most sensitive model (CCSM4-2deg) having relatively low resolution and the least sensitive model (EC-Earth3.1) having relatively high resolution. Again, it is hypothesised that the model parameterisations and nature of the spin up (even when all models have reached quasi-equilibrium) may have a larger effect on sensitivity than the resolution.

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 a 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 the PlioMIP2 boundary conditions have changed compared to PlioMIP1 (particularly the Land-Sea Mask, and the topography). 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 PlioMIP1 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 temperatures were recalculated using the BAYSPLINE calibration. Therefore, the choice of proxy and inter-proxy calibration alone can be enough to alter the interpretation of the extent to which the model and data agree. However, the choice of the observed SST data set used to create the Pliocene minus pre-industrial SST anomaly can also be important. Supplementary Fig. 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, while using HadISST data led

to an absence of SST warming at these locations. The difference between using NOAA ERSST V5 or HadISST is 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 transport of less saline waters 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 is 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, across 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 (Leduc et al. 2014).

5.3 Equilibrium Climate Sensitivity, Earth System Sensitivity and Pliocene climate

From the analysis shown in section 3.6 a strong relationship between ECS and the ensemble-simulated Pliocene temperature anomaly is discernible. This point is true for the globally averaged 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. 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 strongly related to modelled ECS, suggesting that SST data from the Pliocene tropics has the potential to constrain model estimates of ECS, highlighting the benefits for deriving estimates of ECS from a concentrated effort to reconstruct tropical SST response using the geological record.

For PlioMIP1, Hargreaves and Annan (2016) also found that modelled $Plio_{Core} - PI_{Crtl}$ SST anomalies over the tropics (30°N-30°S) were strongly correlated with a models ECS, according to:

$$ECS = \alpha\, \Delta T(30°N - 30°S) + C + \varepsilon \qquad (2)$$

Where ε represents all errors in the regression equation. They then used equation (2) along with tropical SST data from PRISM3D (an interpolated dataset of Pliocene proxy SST) to provide a Pliocene data constrained estimate of ECS of 1. 9°C -3.7°C. In order to constrain ECS from the data and modelling used in PlioMIP2, we slightly amend the Hargreaves and Annan (2016) methodology because PlioMIP2 data is more sparsely distributed than PlioMIP1 data and we cannot obtain a reliable estimate of tropical average SST from the data available. To estimate ECS for PlioMIP2 we instead rely on point-based observations (Fig 8a) and local regressions between $Plio_{Core} - PI_{Crtl}$ SST and modelled ECS (figure 7d). Hence equation (2) will now use ΔSST at individual data sites, and α and C will now be location dependent. Using this altered methodology, a different estimate of ECS is obtained for each datapoint, these estimates are shown in fig 9., and have a range of 2.5°C-4.3°C with a mean ECS of 3.3°C. Fig. 9 does not imply that ECS is different for each location, instead each value in Fig. 9 is an estimate of ECS and incorporates the true Pliocene constrained ECS along with a number of errors. For a data-point to be included in Fig. 9. we required that two conditions were met. Firstly, we required that the relationship between $Plio_{Core} - PI_{Crtl}$ and modelled ECS was significant at the 95% confidence level ($p < 0.05$; these regions are hatched

in figure 7d). Secondly, we required that the model and data were sufficiently consistent that the use of such regressions was appropriate; this second condition meant that we excluded two sites off the Eastern United States, two sites from the Mediterranean, and one site from Benguela – despite these sites showing a theoretical relationship between $Plio_{Core} - PI_{Crtl}$ and ECS. Although only 7 datapoints fulfilled both these conditions, the range of estimates of ECS from PlioMIP2 are in good agreement with IPCC and are similar to (but slightly larger than) PlioMIP1. It is noted, however, that all the extreme estimates within the range are geographically clustered near Central America while the average value of ECS for this region is 3.4°C. The other 3 sites are distributed across the globe and provide ECS estimates of 3.1°C, 3.2°C, 3.4°C respectively. We therefore hypothesise that the most likely estimate of Pliocene data constrained ECS (based on currently available PlioMIP2 ΔSST and models) is in the range 3.1°C-3.4°C, however we acknowledge that this estimate is based on data from a very small number of locations and that there are uncertainties in both model and data.

The emergence of the concept of longer-term sensitivity, ESS, 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 between ECS, $Plio_{Core} - PI_{Crtl}$, 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 CO_2 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 2015 and De Schepper et al., 2015), and which may have a bearing on how well the Pliocene is seen to approximate an equilibrated modern Earth system 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 similar (1.1 to 2.2 and 1.5 respectively). Thus, PlioMIP2 quantitively supports the relationship between ESS and ECS that was obtained from PlioMIP1, however new constraints on the data means that the ECS estimated from PlioMIP2 ΔSST data has increased.

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 which underpins the new data/model comparison. The major findings of the work include:

- Global annual mean surface air temperatures increases by 1.4 to 4.7°C compared to the pre-industrial, with a multi-model average increase of 2.8°C.
- The multi-model mean annual total precipitation rate increases by 6% compared to the pre-industrial, with the modelled range of precipitation increases between 2% and 13%.
- The predicted anomaly between Pliocene and pre-industrial is statistically robust for surface air temperature and sea surface temperature over most of the globe. The modelled precipitation anomaly is significant at mid-high latitudes and in monsoon regions but is not significant in many parts of the tropics and subtropics.
- The degree of polar amplification of surface air temperature change is generally consistent with RCP transient climate modelling experiments used to predict future climate, implying that CO_2 changes dominate the ice sheet changes in the $Plio_{Core}$ experiments.
- 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 2100 CE climate, the difference in the average warming between the hemispheres is subdued. This is likely due to the substantial changes to the albedo feedback mechanism in the Southern Hemisphere following the removal of large areas of the Antarctic ice sheet in the mid-Pliocene.
- There is a statistically significant relationship between ECS and Pliocene global annual average temperature change. The PlioMIP2 ensemble finds that ESS is greater than ECS by a best estimate of 53%.
- Model estimates of the relationship between ECS and $Plio_{Core} - PI_{Crtl,}$ combined with the PlioMIP2 ΔSST, provides a data constrained estimate of ECS with a range of 2.5°C-4.3°C, but with a best estimate of 3.1°C-3.4°C.
- Across the ensemble, there is no clear relationship between the simulated temperature and precipitation anomalies and the year of model release. However newer models may be more sensitive than older models within the same 'family'.
- The PlioMIP2 model ensemble shows broad agreement on polar amplification of the global warming signal and tropical enhancement of rainfall anomalies. Inter-model differences in simulated temperature are mostly found in polar regions and where land-sea-mask and orography of Pliocene paleogeography differ from today.
- 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. JCT was also supported through the Centre for Environmental Modelling And Computation (CEMAC), University of Leeds. 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. BLO-B and RF acknowledge that material for their participation is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation (NSF) under Cooperative Agreement No. 1852977 and NSF OPP grant 1418411. The CESM project is supported primarily by the National Science Foundation. Computing and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX), were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. NCAR is sponsored by the National Science Foundation. NT, CC and GR were granted access to the HPC resources of TGCC under the allocations 2016-A0030107732, 2017-R0040110492 and 2018-R0040110492 (gencmip6) and 2019-A0050102212 (gen2212) provided by GENCI. The IPSL-CM6 team of the IPSL Climate Modelling Centre (https://cmc.ipsl.fr) is acknowledged for having developed, tested, evaluated, tuned the IPSL climate model, as well as performed and published the CMIP6 experiments. CS acknowledges funding by the Helmholtz Climate Initiative REKLIM. CS and GL acknowledge funding via the Alfred Wegener Institute's research programme Marine, Coastal and Polar Systems. QZ acknowledge support from the Swedish Research Council (2013-06476 and 2017-04232). Simulations with EC-Earth were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC). WLC and AAO acknowledge funding from JSPS KAKENHI grant 17H06104 and MEXT KAKENHI grant 17H06323. Their simulations with MIROC4m were performed on the Earth Simulator at JAMSTEC, Yokohama. The work by AvdH and MLJB was carried out under under the program of the Netherlands Earth System Science Centre (NESSC), financially supported by the Ministry of Education, Culture and Science (OCW, grantnr. 024.002.001). Simulations with CCSM4-2deg were performed at the SURFsara dutch national computing facilities and were sponsored by NWO-EW (Netherlands Organisation for Scientific Research, Exact Sciences) under the project 17189. WRP and DC were supported by Canadian NSERC Discovery Grant A9627 and they wish to acknowledge the support of SciNet HPC Consortium for providing computing facilities. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, the Ontario Research Fund – Research Excellence, and the University of Toronto. The NorESM simulations benefitted from resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway.

References

Balsamo, G., Viterbo, P., Beljaars, A., van den Hurk, B., Hirschi, M., Betts, A.K., Scipal, K. 2009. A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. Journal of Hydrometeorology. 10(3), 623-643.

Balsamo G., Pappenberger F., Dutra E., Viterbo P., van der Hurk B. 2011. A revised land hydrology in the ECMWF model: a step towards daily water flux prediction in a fully-closed water cycle, Hydrological Processes, 25 (7), 1046-1054, 10.1002/hyp.7808.

Bitz, C.M., Shell, K.M., Gent, P.R., Bailey, D.A., Danabasoglu, G., Armour, K.C., Holland, M.M., Kiehl, J.T. 2012. Climate Sensitivity of the Community Climate System Model, Version 4. Journal of Climate, 25(9) 3053-3070.

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.

Cattle, H., Crossley, J., Dewry, D.J., Wadhams, P. Dowdeswell, J. A. and Schofield, A.N. 1995. Modelling Arctic climate change. Philosophical Transactions of the Royal Society of Lonson. Series A: Physical and Engineering Sciences, 352. 10.1098/rsta.1995.0064

Chan, W.-L. and Abe-Ouchi, A., in prep. PlioMIP2 simulations using the MIROC4m climate model. Climate of the Past.

Chandan, D. and Peltier, W. R. 2017. Regional and global climate for the mid-Pliocene using the University of Toronto version of CCSM4 and PlioMIP2 boundary conditions. Climate of the Past, 13, 919–942, https://doi.org/10.5194/cp-13-919-2017.

Chandan, D. and Peltier, W. R. 2018. On the mechanisms of warming the mid-Pliocene and the inference of a hierarchy of climate sensitivities with relevance to the understanding of climate futures. Climate of the Past, 14, 825–856, https://doi.org/10.5194/cp-14-825-2018.

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.

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.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rown-tree, P. R., and Smith, J. 1999: The impact of new land surface physicson the GCM simulation of climate and climate sensitivity, Clim.Dynam., 15, 183–203, 10.1007/s003820050276.

Craig, A.P., Vertenstein, M., Jacob, R. 2012. A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. International Journal Of High Performance Computing Applications. 26(1), 31-42.

Danabasoglu, G., Bates, S.C., Briegleb, B.P., Jayne, S.R., Jochum, M., Large, W.G., Peacock, S., Yeager, S.G. 2012. The CCSM4 Ocean Component. Journal of Climate. 25(5), 1361-1389.

De La Vega, E., Chalk, T.B., Foster, G.L., Bysani, R., Wilson, P.A. 2018. Warming in a warm world: Orbital CO_2 forcing variations in the warm Pliocene. Abstract PP21C-1434 presented at the 2018 AGU Fall Meeting, San Francisco, Calif., 11 Dec.

De Schepper, S., Schreck, M., Beck, K.M., Matthiessen, J., Fahl, K., Mangerud, G. 2015. Early Pliocene onset of modern Nordic Seas circulation related to ocean gateway changes. Nature Communications. 6. Article Number 8659.

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.

Dufresne, J.L., Foujols, M.A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., Noblet, N., Duvel, J.P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N. 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, vol. 40, https://doi.org/10.1007/s00382-012-1636-1, http: //link.springer.com/10.1007/s00382-012-1636-1, 2013.

Feng, R., Otto-Bliesner, B.L., Xu, Y.Y., Brad, E., Fletcher, T., Ballantyne, A., 2019. Contributions of aerosol-cloud interactions to mid-Piacenzian seasonally sea ice-free Arctic Ocean. Geophysical Research Letters, 46(16), 9920-9929.

Fedorov, A.V., Brierley, C.M., Emanuel, K., 2010. Tropical cyclones and permanent El Nino in the early Pliocene epoch. Nature, 463 (7284), 1866-U84.

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.

Fichefet, T., Maqueda, M.A.M. 1997. Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics. Journal of Geophysical Research-Oceans. 102(C6), 12609-12646.

Fichefet, T., Maqueda, M.A.M. 1999. Modelling the influence of snow accumulation and snow-ice formation on the seasonal cycle of the Antarctic sea-ice cover. Climate Dynamics. 15(4),251-268.

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.

Gent, P.R., Danabasoglu, G., Donner, L.J., Holland, M.M., Hunke, E.C., Jayne, S.R., Lawrence, D.M., Neale, R.B., Rasch, P.J., Vertenstein, M. and Worley, P.H., 2011. The community climate system model version 4. Journal of Climate, 24(19), pp.4973-4991.

Gettelman, A., Kay, J.E., Shell, K.M. 2012. The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model. Journal of Climate. 25(5), 1453-1469.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. Climate. Dynamics., 16, 147–168, 10.1007/s003820050010.

Guo, C.C. Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., Ottera, O.H. 2019. Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM). Geoscientific Model Development, 12(1), 343-362.

Hagemann, S., Gates, L.D. 2003. Improving a subgrid runoff parameterization scheme for climate models by the use of high-resolution data derived from satellite observations. Climate Dynamics. 21(3-4), 349-359.

Hagemann, S., Dumenil, L. 1998. A parametrization of the lateral waterflow for the global scale. Climate Dynamics. 14(1), 17-31.

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.

Hazeleger, W., Wang, X., Severijns, C., Stefanescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., van den Hurk, B., van Noije, T., van der Linden, E., van der Wiel., K. 2012. EC-Earth V2.2: description and validation of a new seamless earth system prediction model. Climate Dynamics. 39(11), 2611-2629.

Held, I.M., Soden, B.J. 2006. Robust responses of the hydrological cycle to global warming. Journal of Climate, 19, (21), 5686-5699.

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.

Holland, M.M., Bailey, D.A., Briegleb, B.P., Light, B., Hunke, E. 2012. Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice. Journal of Climate. 25(5), 1413-1430.

Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J.L., Fairhead, L., Filiberti, M.A., Friedlingstein, P., Grandpeix, J.Y., Krinner, G., Levan, P., Li, Z.X., Lott, F. 2006. The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. Climate Dynamics. 27(7-8), 787-813.

Hourdin, F., Grandpeix, J.Y., Rio, C., Bony, S., Jam, A., Cheruy, F., Rochetin, N., Fairhead, L., Idelkadi, A., Musat, I., Dufresne, J.L., Lahellec, A., Lefebvre, M.P., Roehrig, R. 2013. LMDZ5B: the atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection. Climate Dynamics. 40(9-10), 2193-2222.

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.

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.

Hunke, E.C., Lipscomb, W.H., Turner, A.K. 2010. Sea-ice models for climate study: retrospective and new directions. Journal of Glaciology. 56 (200), 1162-1172.

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. 10.5194/cp-15-1691-2019.

Hurrell, J.W., Holland, M.M., Gent, P.R., Ghan, S., Kay, J.E., Kushner, P.J., Lamarque, J.F., Large, W.G., Lawrence, D., Lindsay, K. and Lipscomb, W.H., 2013. The community earth system model: A framework for collaborative research, B. Am. Meteorol. Soc., 94, 1339–1360.

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., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*,

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: 10.1155/2015/510713.

Jungclaus, J.H., Keenlyside, N., Botzet, M., Haak, H., Luo, J.J., Latif, M., Marotzke, J., Mikolajewicz, U., Roeckner, E. 2006. Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. Journal of Climate. 19(16), 3952-3972.

K-1 model developers, 2004: K-1 coupled model (MIROC) description. K-1 technical report, edited by: Hasumi, H. and Emori, S., Center for Climate System Research, The University of Tokyo,Japan, 34 pp.

Kamae, Y., Yoshida, K., Ueda, H. 2016. Sensitivity of Pliocene climate simulations in MRI-CGCM2.3 to respective boundary conditions. Climate of the Past. 12(8), 1619-1634.

Krinner, G., Viovy, N., de Noblet-Ducoudre, N., Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C. 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Global Biogeochemical Cycles. 19 (1), GB1015.

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.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Thornton, P.E., Swenson, S.C., Lawrence, P.J., Zeng, X.B., Yang, Z.L., Levis, S., Sakaguchi, K., Bonan, G.B., Slater, A.G. 2011. Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. Journal of Advances in Modeling Earth Systems. 3. Article Number: M03001.

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Fletcher, C.G., Lawrence, P.J., Levis, S., Swenson, S.C., Bonan, G.B. 2012. The CCSM4 Land Simulation, 1850-2005: Assessment of Surface Climate and New Capabilities. Journal of Climate. 25(7). 2240-2260.

Leduc, G., Garbe-Schonberg, D., Regenberg, M., Contoux, C. 2014. The late Pliocene Benguela upwelling status revisited by means of multiple temperature proxies. Geochemistry Geophysics Geosystems, 15(2), 475-491.

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.

Madec, G. 2017. NEMO ocean engine. Technical note, IPSL, available at: https://doi.org/10.5281/zenodo.1464816, NEMO book.pdf (last access: 24 November 2019).

Madec, G. 2008. NEMO Ocean Engine, Note du Pole de Modlisation. Institut Pierre-Simon Laplace (IPSL), Paris, France. No. 27, ISSN No. 1288e1618.

Madec, G., Imbard, M. 1996. A global ocean mesh to overcome the North Pole singularity, Climate Dynamics, 12, 381–388, https://doi.org/10.1007/BF00211684.

Marsland, S.J., Haak, H., Jungclaus, J.H., Latif, M., Roske, F. 2003. The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. Ocean Modelling. 5(2), 91-127.

Marti, O., Braconnot, P., Dufresne, J.L., Bellier, J., Benshila, R., Bony, S., Brockmann, P., Cadule, P., Caubel, A., Codron, F., de Noblet, N., Denvil, S., Fairhead, L., Fichefet, T., Foujols, MA., Friedlingstein, P., Goosse, H., Grandpeix, J.Y., Guilyardi, E., Hourdin, F., Idelkadi, A., Kageyama, M., Krinner, G., Levy, C., Madec, G., Mignot, J., Musat, I., Swingedouw, D., Talandier, C. 2010. Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. Climate Dynamics. 34(1), 1-26.

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.

Mellor, G.L., Kantha, L. 1989. An Ice-Ocean Coupled Model. Journal of Geophysical Research-Oceans. 94(C8) 10937-10954.

Neale, R.B., Richter, J., Park, S., Lauritzen, P.H., Vavrus, S.J., Rasch, P.J., Zhang, M.H. 2013. The Mean Climate of the Community Atmosphere Model (CAM4) in Forced SST and Fully Coupled Experiments. Journal of Climate, 26(14), 5150-5168.

Neale, R.B., Chen, C.C., Gettelman, A., Lauritzen, P.H., Park, S., Williamson, D.L., Conley, A.J., Garcia, R., Kinnison, D., Lamarque, J.F. and Marsh, D., 2010b. Description of the NCAR community atmosphere model (CAM 5.0). *NCAR Tech. Note NCAR/TN-486+ STR*, *1*(1), pp.1-12.

Neale, R.B., Richter, J.H., Conley, A.J., Park, S., Lauritzen, P.H., Gettelman, A., Williamson, D.L., Vavrus, S.J., Taylor, M.A., Collins, W.D. and Zhang, M., 2010a. Description of the NCAR Community Atmosphere Model (CAM 4.0).

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.

Oleson, K.W., Niu, G.Y., Yang, Z.L., Lawrence, D.M., Thornton, P.E., Lawrence, P.J., Stockli, R., Dickinson, R.E., Bonan, G.B., Levis, S., Dai, A., Qian, T. 2008. Improvements to the Community Land Model and their impact on the hydrological cycle. Journal of Geophysical Research-Biogeosciences. 113(G1), G01021.

Oki, T., Sud, Y.C. 1998. Design of Total Runoff Integrating Pathways (TRIP)A31 Global River Channel Network, Earth Interactions, 2(1), 1–36, available online at32 https://www.ametsoc.org/index.cfm/ams/publications/journals/earth-interactions/.

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.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A.:The impact of new physical parametrizations in the HadleyCentre climate model: HadAM3, Clim. Dynam., 16, 123–146,https://doi.org/10.1007/s003820050009, 2000.

Pound, M.J., Tindall, J., Pickering, S.J., Haywood, A.M., Dowsett, H.J., Salzmann, U . 2014. Late Pliocene lakes and soils: a global data set for the analysis of climate feedbacks in a warmer world. Climate of the Past, 10(1), 167-180.

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.

Raddatz, T.J., Reick, C.H.,Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.G., Wetzel, P. Jungclaus, J. 2007. Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century? Climate Dynamics. 29(6), 565-574.

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

Roeckner, E., Bäuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Rhodin, A., Schlese, U., Schulzweida, U., and Tompkins, A. 2003. The atmospheric general circulation model ECHAM5. PART I: Model description, Report 349, Max-Planck-Institut für Meteorologie, Hamburg.

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.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthelemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., Vivier, F. 2015. The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities. Geoscientific Model Development. 8(10) 2991-3005.

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.

Sagoo, N., Storelvmo, T., 2017. Testing the sensitivity of past climates to the indirect effects of dust. Geophysical Research Letters, 44(11), 5807-5817.

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.

Sato, N., Sellers, P. J., Randall, D. A., Schneider, E. K., Shukla, J., Kinter, J. L., Hou, Y.-Y., and Albertazzi, E. 1989. Effects of implementing the simple biosphere model in a general circulation model, J. Atmos. Sci., 46, 2757–2782, 1989.

Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A. 1986. A simple biosphere model (SiB) for use within general circulation models, J. Atmos. Sci., 43, 505–531, 1986.

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.

Smith, R., Jones, P., Briegleb, B.; et al. 2010. The Parallel Ocean Program (POP) Reference Manual: Ocean Component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM). Tech. rep., available at: last access: 23 November 2019 , Publisher: Los Alamos National Laboratory, URL: http://www.cesm.ucar.edu/models/ccsm4.0/pop/doc/sci/POPRefManual.pdf

Smith, D. M., Screen, J. A., Deser, C., Cohen, J., Fyfe, J. C., García-Serrano, J., Jung, T., Kattsov, V., Matei, D., Msadek, R., Peings, Y., Sigmond, M., Ukita, J., Yoon, J.-H., and Zhang, X.: The Polar Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: investigating the causes and consequences of polar amplification, Geosci. Model Dev., 12, 1139–1164.

Stouffer, R.J., S. Manabe, and K. Bryan. 1989. Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO_2. 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.

Tan, N., Contoux, C., Ramstein, G., Sun, Y., Dumas, C., Sepulchre, P. (2019). Modelling a Modern-like-*p*CO2 Warm Period (MIS KM5c) with Two Versions of IPSL AOGCM, Clim. Past Discuss., https://doi.org/10.5194/cp-2019-83.

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.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., Maqueda, M.A.M. 2009. Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Modelling . 27(1-2), 33-53.

Yukimoto, S., Noda, A., Kitoh, A., Hosaka, M., Yoshimura, H., Uchiyama, T., Shibata, K., Arakawa, O., Kusunoki, S. 2006. Present-day climate and climate sensitivity in the Meteorological Research Institute coupled GCM version 2.3 (MRI-CGCM2.3). Journal of The Meteorological Society of Japan. 84(2) 333-363.

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.

.

(a) **Model ID, Vintage**	(b) **Sponsor(s), Country**	(c) **Atmosphere** **Top Resolution & Model References**	(d) **Ocean*** **Resolution Z Coord., Top BC, & Model References**	(e) **Sea Ice*** **Dynamics, Leads & Model References**	(f) **Coupling*** **Flux adjustments & Model References**	(g) **Land** **Soils, Plants, Routing & Model References**	(h) **PlioMIP2 Experiment Eoi400** **(Boundary Conditions & Experiment Citation)**	(i) **Vegetation (Static - Salzmann et al. 2008 or Dynamic)**	(j) **Climate Sensitivity (ECS) °C** **(incl. source)**
CCSM4_1deg (CESM 1.0.5) 2011	National Center for Atmospheric Research	Top = 3.6 hPa FV0.9x1.25 (~1°), L26 (CAM4) (Neale et al. 2010a)	G16 (~1°), L60 depth, rigid lid	Rheology, melt ponds Holland et al. (2012); Hunke and Lipscomb (2010)	No adjustments Gent et al. (2011)	Layers, prescribed vegetation type with prognostic phenology, carbon cycle, routing Oleson et al. (2008)	Enhanced Feng et al. (in prep)	Salzmann et al. (2008)	3.2 (Bitz et al. 2012)
CCSM4_2deg (CESM 1.0.5) 2011	IMAU, Utrecht University, the Netherlands	CAM4 Top = 2hPa FV (2.5° x 1.9°) L26 Neale et al. (2013)	POP2 Bipolar Curvilinear 320 x 384 (formal 1°) L60 Smith et al. (2010)	CICE4 Hunke and Lipscomb (2008)	CPL7 Craig et al. (2012)	CLM4 Oleson et al (2010); Lawrence et al. (2011)	Enhanced TBA?	Salzmann et al. (2008)	3.2 (Bitz et al. 2012)
CCSM4-UoT 2011	University of Toronto, Canada	Top = 2.2 hPa 1.25° x 0.9°, L26 Neale et al. (2013)	0.27-0.54° x 1.1°, L60 Depth, free surface Smith et al. (2010), Danabasoglu et al. (2012), Chandan and Peltier (2017)	Rheology, melt ponds Holland et al. (2012, Hunke and Lipscomb, (2010)	No adjustment Craig et al., (2012)	Layers, canopy, routing Lawrence et al. (2011)	Enhanced Chandan and Peltier (2017, 2018)	Salzmann et al. (2008)	3.2 (Bitz et al. 2012)
CESM1.2 2013	National Center for Atmospheric Research	Top = 3.6 hPa FV0.9x1.25 (~1°), L30 (CAM5) (Neale et al. 2010b)	G16 (~1°), L60 depth, rigid lid	Rheology, melt ponds Holland et al. (2012); Hunke and Lipscomb (2010)	No adjustments Hurrell et al. (2013)	Layers, prescribed vegetation type with prognostic phenology, carbon cycle, routing Oleson et al. (2008)	Enhanced Feng et al. (in prep)	Salzmann et al. (2008)	4.1 (Gettelman et al. 2012)
COSMOS COSMOS-landveg r2413 2009	Alfred Wegener Institute, Germany	Top = 10 hPa T31 (3.75° x 3.75°), L19 Roeckner et al. (2003)	Bipolar orthogonal curvilinear GR30, L40 (formal 3.0° x 1.8°) Depth, free surface Marsland et al. (2003)	Rheology, leads Marsland et al. (2003),	No adjustments Jungclaus et al. (2006)	Layers, canopy, routing Raddatz et al. (2007), Hagemann and Dümenil (1998), Hagemann and Gates (2003)	Enhanced Stepanek et al. (in prep.)	Dynamic	4.7 (Uploaded 2 x CO_2 minus PI experiment)

EC-Earth 3.1 2013	Stockholm University, Sweden	IFS cycle 36r4 Top = 5 hPa 1.125° x 1.125°, L62 Hazeleger et al. (2012)	NEMO3.6, ORAC1 1.0° x 1.0°, L46 Madec (2008)	LIM3 Vancoppenolle et al. (2009)	No adjustments Hazeleger et al. (2012)	Layers, canopy, routing Balsamo et al. (2009), Balsamo et al. (2011)	Enhanced Zheng et al. (2019)	Salzmann et al. (2008)	3.2 (value calculated from a 4 x CO_2 experiment)
GISS2.1G 2019	Goddard Institute for Space Studies, United States of America	Top = 0.1 mb 2.0˚ x 2.5˚, L40 Kelley et al. (in prep)	1.0˚ x 1.25˚, L40 P*, free surface Kelley et al. (in prep)	Visco-plastic rheology, leads, melt ponds Kelley et al. (in prep)	No adjustments Kelley et al. (in prep)	Layers, canopy, routing Kelley et al. (in prep)	Enhanced Chandler et al. (in prep)	Salzmann et al. (2008)	3.3 (Kelley et al. in prep)
HadCM3 1997	University of Leeds, United Kingdom	Top = 5 hPa 2.5° x 3.75°, L19 Pope et al. (2000)	1.25° x 1.25°, L20 Depth, rigid lid Gordon et al. (2000)	Free drift, leads Cattle and Crossley, (1995)	No adjustments Gordon et al. (2000)	Layers, canopy, routing Cox et al. (1999)	Enhanced Hunter et al. (2019)	Salzmann et al. (2008)	3.5 (Hunter et al. 2019)
IPSLCM6A-LR 2018	Laboratoire des Sciences du Climat et de l'Environnement (LSCE), France	Top = 1 hPa 2.5° x 1.26°, L79 Hourdin et al. (in prep)	1° x 1°, refined at 1/3° in the tropics, L75 Free surface, Z-coordinates Madec et al. (2017)	Thermodynamics, Rheology, Leads Vancoppenolle et al. (2009), Rousset et al. (2015)	No adjustments Marti et al. (2010), Mignot et al. (in prep)	Layers, canopy, routing, phenology Peylin et al. (in prep)	Enhanced Contoux et al. (in-prep)	Salzmann et al. (2008)	4.8 (Mignot et al. in prep)
IPSLCM5A2.1 2017	Laboratoire des Sciences du Climat et de l'Environnement (LSCE), France	Top = 70 km 3.75° x 1.9°, L39 Hourdin et al. (2006, 2013), Sepulchre et al. (in prep)	0.5°-2° x 2°, L31 Free surface, Z-coordinates Dufresne et al. (2013), Madec et al. (1996), Sepulchre et al. (in prep)	Thermodynamics, Rheology, Leads Fichefet and Morales-Maqueda, (1997, 1999), Sepulchre et al. (in prep)	No adjustment Marti et al. (2010), Sepulchre et al. (in prep)	Layers, canopy, routing, phenology Krinner et al., (2005), Marti et al. (2010), Dufresne et al. (2013)	Enhanced Tan et al. (submitted)	Salzmann et al. (2008)	3.6 (Sepulchre Pierre pers. Comm.)
IPSLCM5A 2010	Laboratoire des Sciences du Climat et de l'Environnement (LSCE), France	Top = 70 km 3.75° x 1.9°, L39 Hourdin et al. (2006, 2013)	0.5°-2° x 2°, L31 Free surface, Z-coordinates Dufresne et al. (2013), Madec et al. (1996)	Thermodynamics, Rheology, Leads Fichefet and Morales-Maqueda, (1997, 1999)	No adjustment Marti et al. (2010), Dufresne et al. (2013)	Layers, canopy, routing, phenology Krinner et al. (2005), Marti et al. (2010), Dufresne et al. (2013)	Enhanced Tan et al. (submitted)	Salzmann et al. (2008)	4.1 (Dufresne et al. 2013)

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 K-1 Developers (2004)	0.5° -1.4° x 1.4°, L43 Sigma/depth free surface K-1 Developers (2004)	Rheology, leads K-1 Developers (2004)	No adjustments K-1 Developers (2004)	Layers, canopy , routing K-1 Developers (2004); Oki and Sud (1998)	Enhanced Chan et al. (in prep)	Salzmann et al. (2008)	3.9 (Uploaded 2 x CO_2 minus PI experiment)
MRI-CGCM 2.3 2006	Meteorological Research Institute and University of Tsukuba, Japan	Top = 0.4 hPa T42 (~2.8° x 2.8°) L30 Yukimoto et al. (2006)	0.5°-2.0° x 2.5°, L23 Depth, rigid lid Yukimoto et al. (2006)	Free drift, leads Mellor and Kantha (1989)	Heat, fresh water and momentum (12°S-12°N) Yukimoto et al. (2006)	Layers, canopy, routing Sellers et al. (1986); Sato et al. (1989)	Standard Kamae et al. (2016)	Salzmann et al. (2008)	2.8 (Uploaded 2 x CO_2 minus PI experiment)
NorESM-F 2017	NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway	Top = 3.5 hPa 1.9° × 2.5°, L26 (CAM4)	~1° x 1°, L53 isopycnal layers	Rheology, melt ponds Holland et al., (2012); Hunke and Lipscomb (2010)	No adjustments Gent et al. (2011)	Layers, canopy, routing Lawrence et al. (2012)	Enhanced (modern soils) Li et al. (in prep)	Salzmann et al. (2008)	2.3 (Guo et al. 2019)
NorESM-L (CAM4) 2011	NORCE Norwegian Research Centre, 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	Rheology, melt ponds Holland et al., (2012); Hunke and Lipscomb (2010)	No adjustments Gent et al. (2011)	Layers, canopy, routing Lawrence et al. (2012)	Enhanced (modern soils) Li et al. (in prep)	Salzmann et al. (2008)	3.1 (Haywood et al. 2013a)

Table 1: Details of climate models used with the MPEoi400 (Plio_Core) experiment (a to g), plus details of boundary conditions (h), treatment of vegetation (i) and Equilibrium Climate Sensitivity values (j) (°C).

975

Model Name	ECS	Eoi400 SAT	E280 SAT	Eoi400-E280 SAT	ESS (eqn 1) ESS	ESS/CS ratio
CCSM4-2deg	3.2	18.5	13.8	4.7	9.1	2.85
CCSM4-1deg	3.2	16.0	13.4	2.6	5.1	1.59
CCSM4-UoT	3.2	16.8	13.0	3.8	7.3	2.29
CESM1.2	4.1	17.3	13.3	4.0	7.7	1.89
COSMOS	4.7	16.8	13.5	3.3	6.5	1.38
EC-Earth3.1	3.2	15.4	14.1	1.4	2.7	0.83
GISS2.1G	3.3	15.9	13.8	2.1	4.0	1.22
HadCM3	3.5	17.0	14.1	2.9	5.6	1.60
IPSLCM6A	4.8	16.0	12.5	3.4	6.7	1.39
IPSLCM5A2	3.6	15.3	13.2	2.2	4.2	1.17
IPSLCM5A	4.1	14.4	12.1	2.3	4.5	1.11
MIROC4m	3.9	15.9	12.8	3.1	6.0	1.54
MRI-CGCM2.3	2.8	15.1	12.7	2.4	4.7	1.66
NorESM-L	3.1	14.6	12.5	2.1	4.1	1.33
NorESM1-F	2.3	16.2	14.5	1.7	3.3	1.45
MMM	**3.5**	**16.1**	**13.3**	**2.8**	**5.4**	**1.55**
CCSM4-avg	3.2	17.1	13.4	3.7	7.2	2.24
MMM (alt)**	**3.5**	**16.0**	**13.2**	**2.8**	**5.4**	**1.53**

Table 2: Details of the relationship between the equilibrium climate sensitivity (ECS) and the Earth System Sensitivity (ESS) for each model. MMM denotes the multimodel mean. Also shown is an alternative multimodel mean (MMM (alt)*), this excludes the EC-Earth3.1 model and uses an average of the CCSM4 models in preference to their individual value.

980

Figures

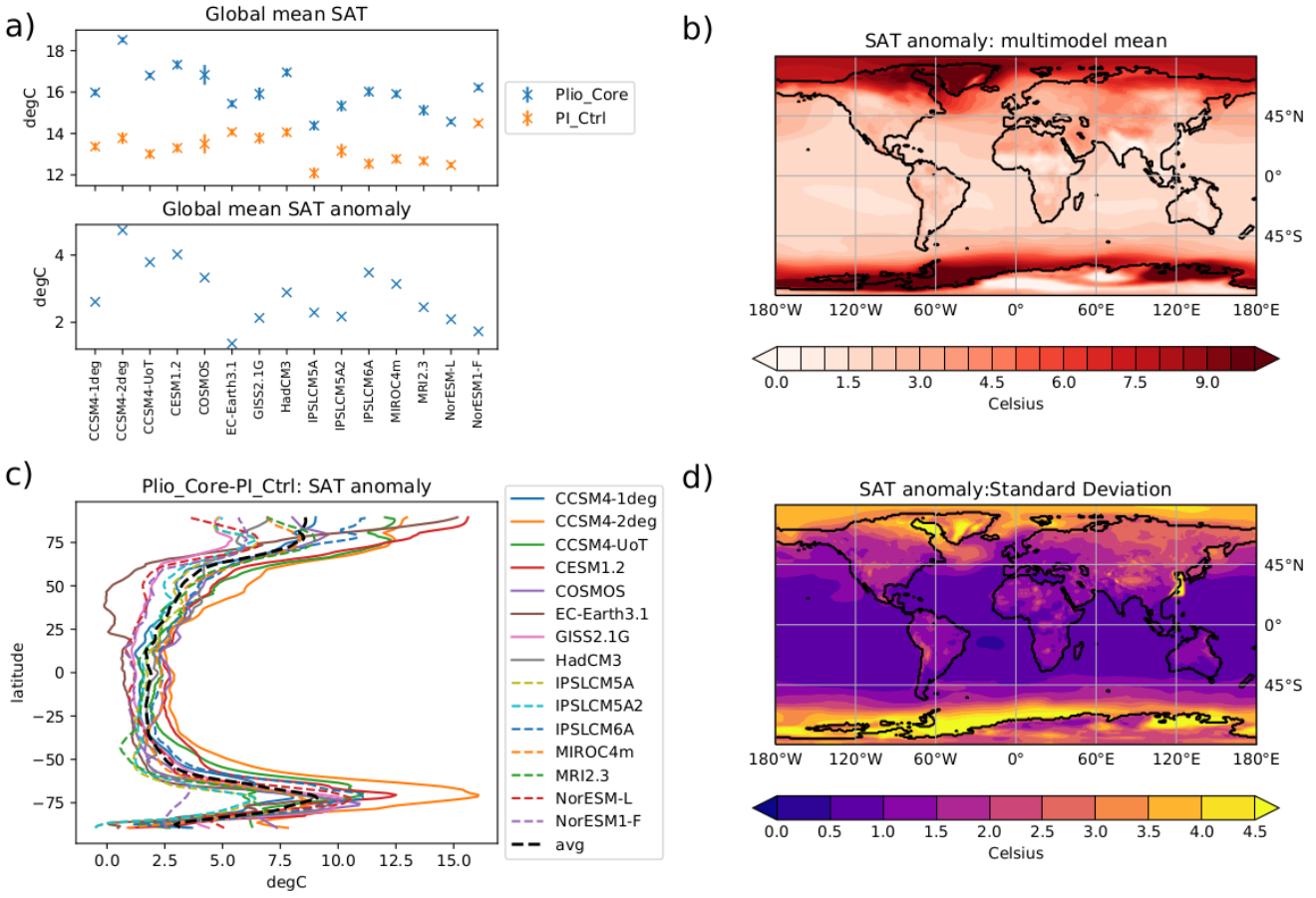

Figure 1: a) Global mean Near Surface Air Temperature [SAT] for the $Plio_{Core}$ and PI_{Ctrl} experiments from each model [upper panel] and the difference between them ($Plio_{Core}$ - PI_{Ctrl}) [lower panel]. Crosses show the mean value while the vertical bars show the interannual standard deviation. b) Multimodel mean SAT $Plio_{Core}$ - PI_{Ctrl} anomaly. c) Latitudinal mean SAT $Plio_{Core}$ - PI_{Ctrl} anomaly. d) Intermodel standard deviation for the $Plio_{Core}$ - PI_{Ctrl} anomaly.

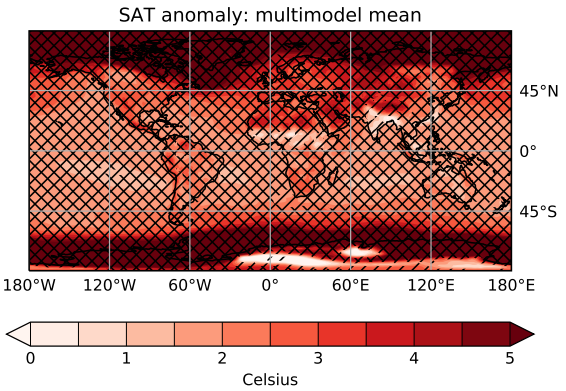

Figure 2: $Plio_{Core}$ - PI_{Ctrl} SAT multimodel mean anomaly. Gridboxes where at least 12 of the 15 models agree on the sign of the change are marked '/'. Gridboxes where the ratio of the multimodel mean SAT change to the PI_{Ctrl} intermodel standard devition is greater than 1 are marked '\'. Gridboxes where both these conditions are satisfied show a robust signal.

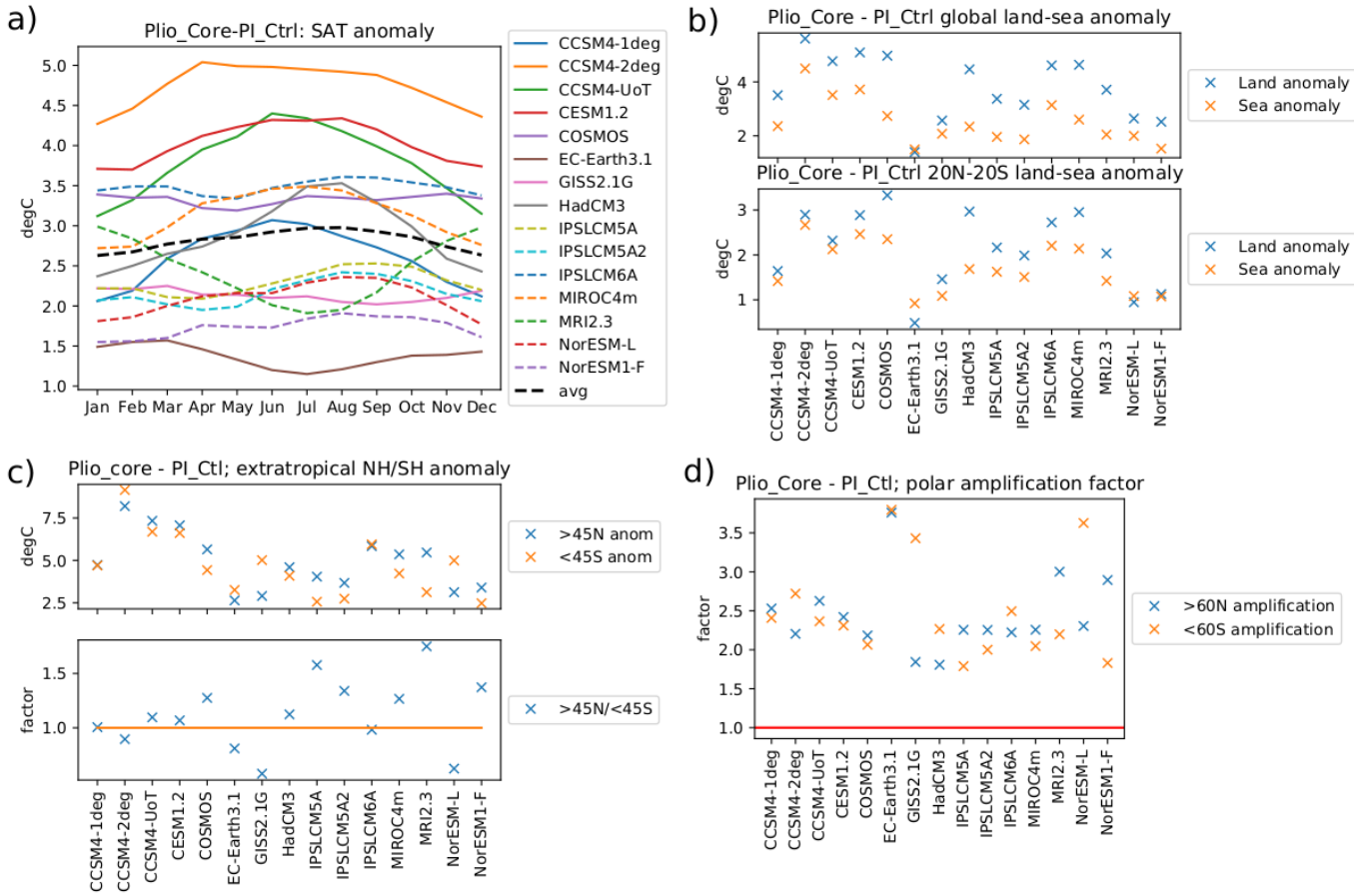

Figure 3: a) Monthly mean $Plio_{Core}$ - PI_{Ctrl} SAT anomaly for each model. b) SAT anomaly for land (blue) and sea (orange) from each model averaged over the globe [top panel] and the 20°N-20°S region [lower panel]. c) SAT anomaly for the northern extratropics (blue) and southern extratropics (orange) [top panel] and the ratio between them [lower panel]. d) SAT anomaly poleward of 60° divided by globally averaged SAT anomaly for the NH (blue) and the SH (orange). The red line highlights a ratio of 1 (i.e. no polar amplification).

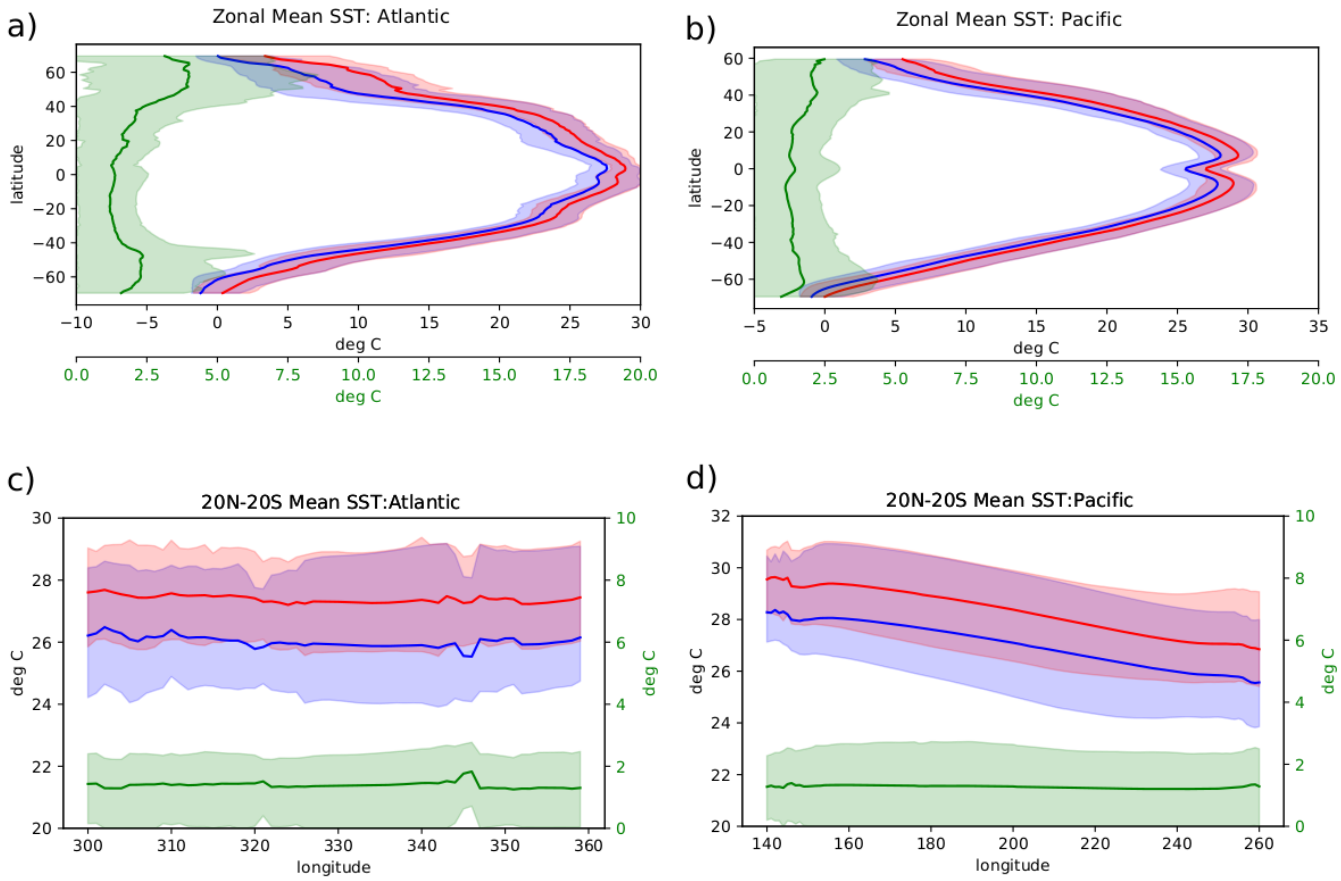

Figure 4: a) and b) show the zonally averaged SST over the Atlantic region (70°W-0°E) and the Pacific region 150°E-100°W respectively. c) and d) show the SST averaged between 20°N and 20°S for the Atlantic and Pacific respectively. In all figures blue shows PI_{Ctrl}, red shows $Plio_{core}$ and green shows the anomaly between them. The solid line shows the multimodel mean, while the shaded area shows the range of modelled values.

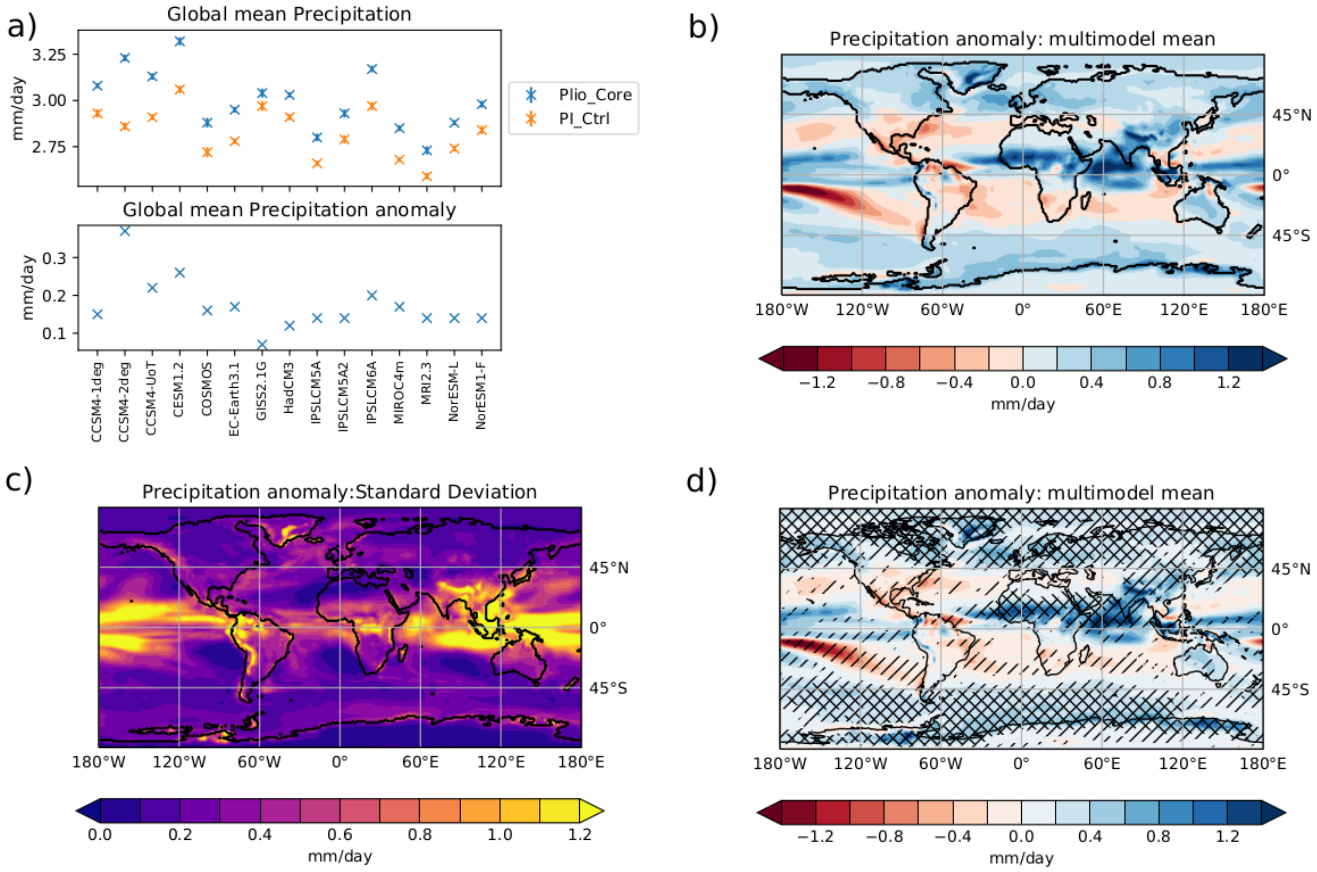

Figure 5: a) Globally averaged precipitation for $Plio_{Core}$ and PI_{Ctrl} from each model [upper panel] and the anomaly between them [lower panel]. b) Multimodel mean $Plio_{Core}$ - PI_{Ctrl} precipitation anomaly. c) Standard deviation across the models for the $Plio_{Core}$ - PI_{Ctrl} Precipitation anomaly. d) $Plio_{Core}$ - PI_{Ctrl} precipitation anomaly, regions which have at least 12 of the 15 models agreeing on the sign of the change are marked with '/'. Regions which have the ratio of the multimodel mean SAT change to the PI_{Ctrl} intermodel standard devition is greater than 1 are marked with '\'.

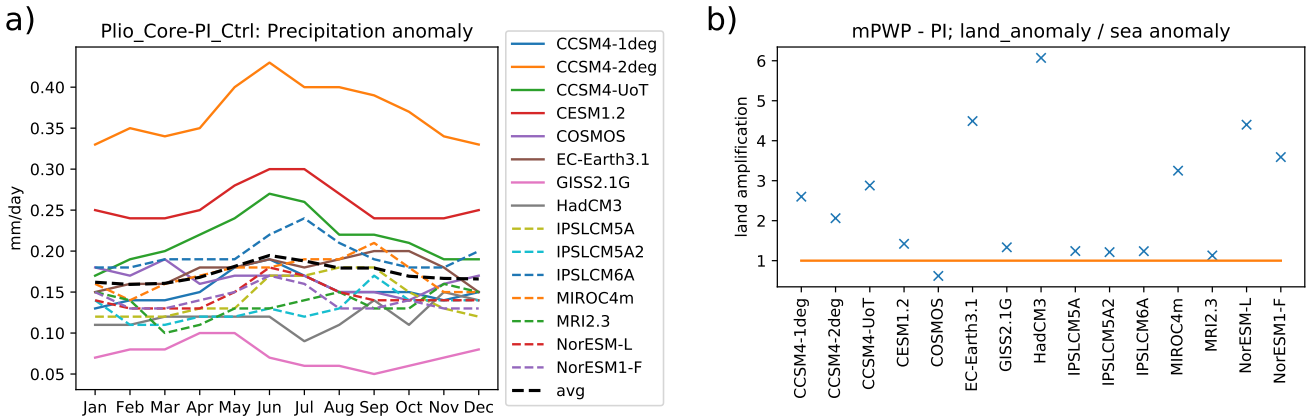

Figure 6: a) shows the globally averaged precipitation anomaly for each model and for each month. b) shows the ratio of the precipitation anomaly over land to the precipitation anomaly over sea (a ratio of 1.0 - where the land precipitation anomaly is the same as the sea precipitation anomaly is shown in orange).

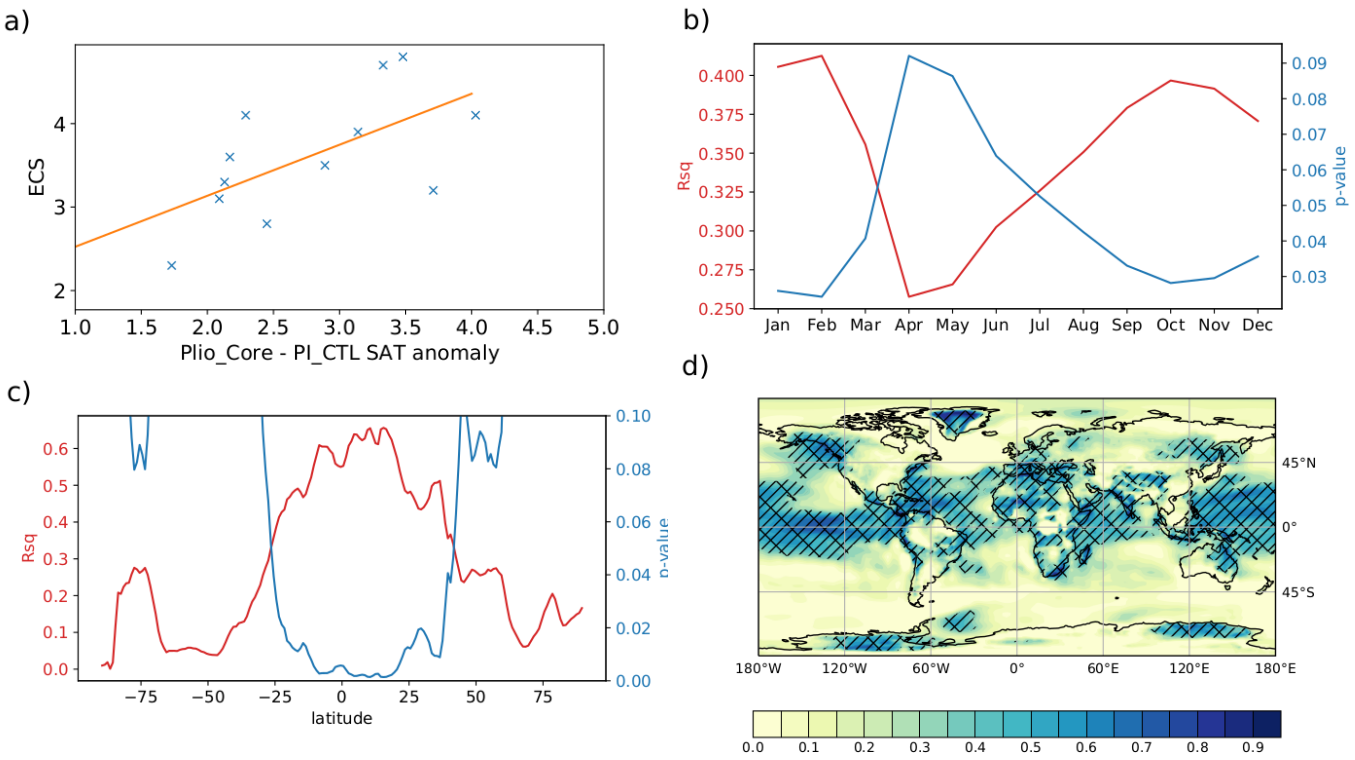

Figure 7: a) the global SAT anomaly for each model vs the published climate sensitivity. b) statistical relationships between the published climate sensitivity and the monthly averaged SAT anomaly across the models. The proportion of climate sensitivity that can be explained by the SAT anomaly in each month (Rsq) is shown in red, while the probability that there is no correlation between the climate sensitivity and the SAT anomaly (p) is shown in blue. c) as b) however climate sensitivity has been correlated with the zonally averaged SAT anomaly. d) colors show the proportion of climate sensitivity across the models that can be explained by the SAT anomaly at any given gridsquare (R_{sq}). Hatching shows a significant relationship (at the 5% confidence level) between SAT anomaly at that gridsquare and climate sensitivity.

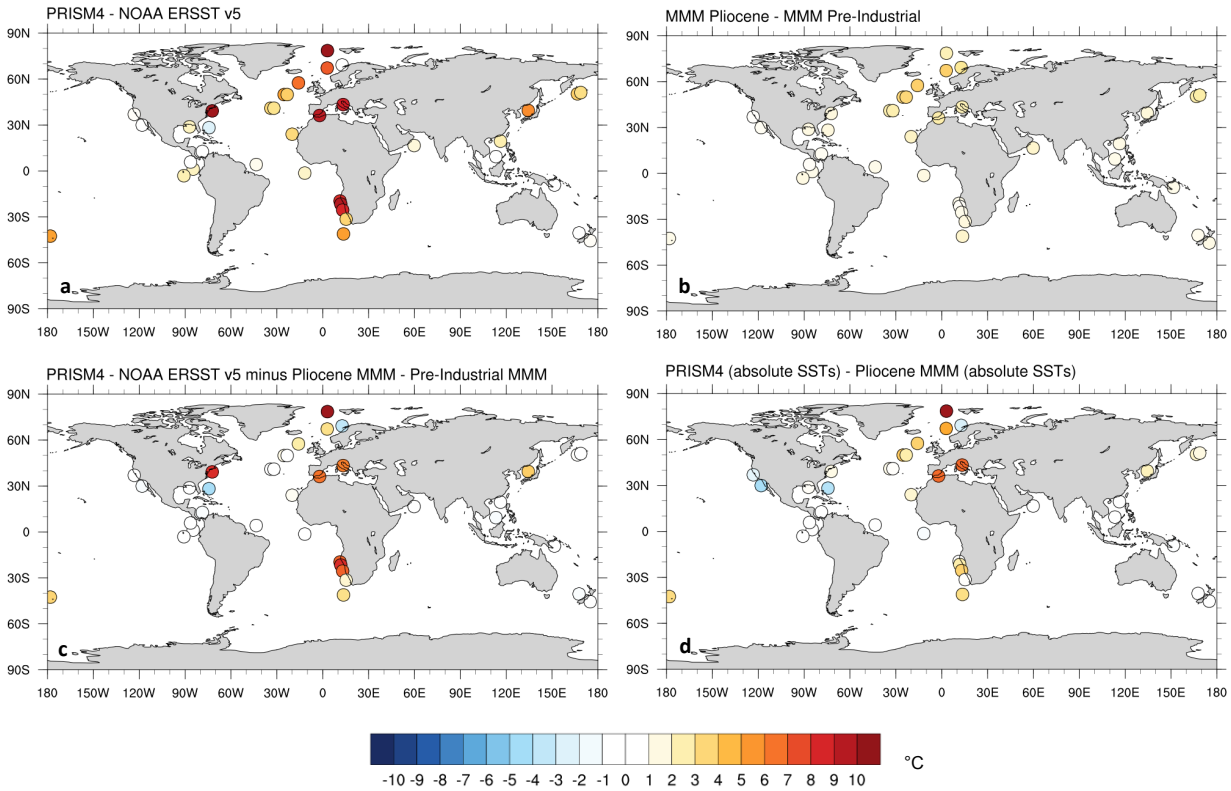

Figure 8: a) PRISM4 - NOAA ERSSTv5 SST anomaly for the datapoints described in section 4. b) multimodel mean $Plio_{Core}$ - PI_{Ctrl} SST anomaly at the points where data are available. c) The difference between the SST anomaly derived from the models (figure8b) and that obtained from the data (figure 8a). d) the PRISM4 SST data minus the $Plio_{Core}$ multimodelmean

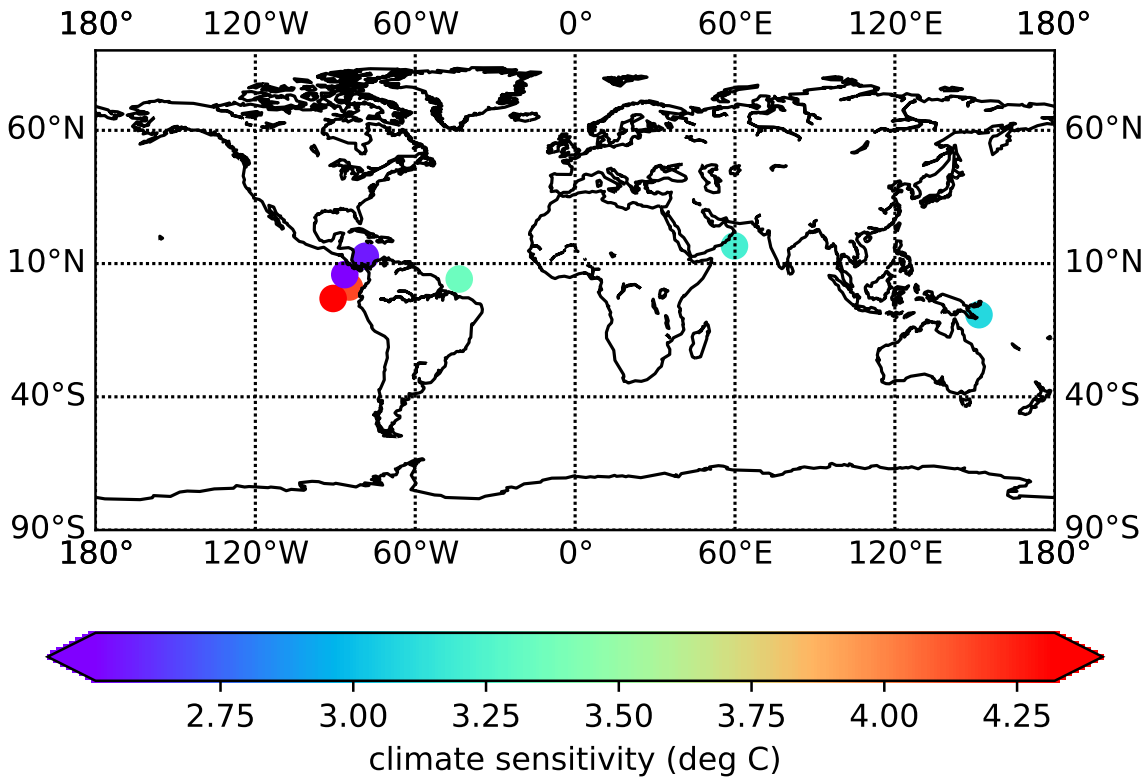

Figure 9: Climate sensitivity calculated from the data shown in figure 8 and using the gridpoint based regressions (discussed in section 3.6 and section 5.3).