Response to Reviewer 2:

We would like to thank the reviewer for providing helpful comments on this manuscript. The suggestions will help improve the manuscript in a revised version. We provide a response to specific comments below.

The manuscript by Haywood et al. presents results from the Pliocene Model Intercomparison Project Phase 2 and follows on an earlier Pliocene model intercomparison published in 2013. Despite the title, the manuscript focuses on two issues: Pliocene large-scale climate features and earth system/climate sensitivity. The study has the potential to be an important contribution. As presented, the study has several shortcomings, described below. My recommendation is to revise the manuscript, eliminating the analysis of earth system sensitivity and expanding the analysis of large-scale features and the comparison with proxy data.

The manuscript is intended as an introduction to the main results from PlioMIP2 and includes a large range of features in temperature and precipitation across different timescales and spatial scales. We also present a comparison with proxy data and consider the relationship between climate sensitivity and the Pliocene climate anomaly.

We agree that it would be possible to move the analysis of Earth System Sensitivity and Climate sensitivity to another paper. However, we do not agree that this is necessary or indeed desirable. The inclusion of this analysis provides a useful link between PlioMIP2 output, Pliocene proxy data and future climate change, which broadens the relevance of the manuscript. Removing this analysis would be to the detriment of the manuscript.

However, we acknowledge that the ESS/ECS section can be improved. We will simplify this section in the revised version of the paper. In the previous version complications arose from

1. *The EC-Earth3.1 model, which had ESS < ECS – which was unreasonable. This model has been now withdrawn from PlioMIP2 by the EC-Earth modelling group due to recently determined problems in sea ice sensitivity.*
2. *The CCSM4 models not being internally consistent. We had three very different representations of the Pliocene climate from CCSM4 due to different groups using different model parameterisations and different ways in which the Pliocene boundary conditions were implemented. We therefore took decision to average these models together to investigate the relationship between ECS (which was the same in all the CCSM4 models) and ESS (which was different in the CCSM4 models).*

In the revised version of the paper we will include all models in our ECS/ESS analysis, including treating all versions of CCSM4 as separate and distinct models. This will make the ECS/ESS analysis consistent with the rest of the paper, where all versions of CCSM4 were treated as separate models.

In addition, since the previous version of the paper we have received two further contributions to PlioMIP2. These are EC-Earth3.3 and CESM2. In the revised version we will include these two additional models, and this strengthen the results seen in the initial submission. In particular, the inclusion of these additional models strengthens the link between ECS and ESS and makes this section more robust.

We agree with the reviewer that the title did not accurately reflect the contents of the paper. In the revised version we will change the title to show that the paper deals with both large-scale features of the Pliocene climate and climate sensitivity.

General comments

Discussion paper

As an analysis of a model intercomparison project, the manuscript is fine. It reports the ensemble average and range across a spectrum of large-scale features including global, zonal and seasonal temperature; polar amplification; SST gradients, and precipitation rate, and compares them to pre-industrial conditions. These results are interesting and relevant, but the analysis is rather cursory. It would have also been interesting for the authors to include some direct comparisons with PlioMIP1 in the Results (for example, include PlioMIP1 ensemble means in the figures and description of results). In the Discussion, there is some speculation about why the PlioMIP2 results differ from PlioMIP1, but it's just speculation. In addition, the analysis of PlioMIP2 models lacks investigation of why large-scale climate features differ among PlioMIP2 models. Both of these are major missed opportunities.

The manuscript represents the first results from the PlioMIP2 ensemble and therefore we included many comparisons/analysis that subsequent studies will expand on. We acknowledge that the reviewer thought the analysis was too brief, however, the manuscript is an overview for a broad audience and already runs to a number of pages, hence we would not like to add a large amount of additional detail.

In a number of places, the text compares results with PlioMIP1, however we note from the reviewer's comment that a more quantitative comparison would be desirable. In the revised version of the manuscript we will add the PlioMIP1 results to appropriate figures (fig 1a, fig 1c, fig3, fig4a, fig6).

We agree with the reviewer that it would be very interesting to understand why large-scale climate features differ among different models. However, this is a very difficult question to answer and it is beyond the scope of this study to do this thoroughly for all diagnostics. In general, the models that have the greatest Pliocene warming also have the greatest published climate sensitivity, and this will be emphasised more in the conclusions. Where all models have been presented individually (i.e. fig 1a, fig4a etc) we will reorder the models in terms of highest ECS to lowest so that the reader can more clearly see how the diagnostics relate to each other.

The estimate of earth system sensitivity (ESS) (equation 1) is not explained or justified. There is no a priori reason to think that the ESS will scale as the ratio of ln (560/280)/ln(400/280). This scaling would be appropriate if CO2 were the only factor changing between simulations. That this scaling is inappropriate is illustrated by the differences in ESS for CCSM4 (Table 2, compare CCSM4-2deg and CCSM4-UofT). In these two simulations with the same model, the Eo400-E280 SAT differs by 0.9 C due to difference in treatment of Pliocene boundary conditions. (No surprise.) The ESS estimate (equation 1) grows the difference to 1.8 C. Why would the same model respond so differently to an increase in CO2 of 160 ppm? If the authors believe this result is justified, they must demonstrate it by running these two CCSM4 Pliocene simulations with 560 ppm CO2. Given the shortcoming in the ESS estimate, all discussion of earth system sensitivity should be removed from the manuscript.

The reviewer makes two points here:

1. *That equation 1 is not explained or justified*
2. *That the CCSM4-2deg and CCSM4-UoT provide a very different response to the mPWP boundary conditions, and hence ESS (calculated from these different version of the same model) is very different when it should not be.*

To respond to point 1:

The ESS is simply

*DeltaR*ESS=DeltaT*

*Where DeltaR is 5.35 * ln(400/280)*

As far as the determination of ESS is concerned we are not doing anything unusual. As elaborated in Chandan and Peltier 2018 (this issue), the usage of this equation is justifiable on ESS timescale if we consider the various ice-sheet and GIA related changes as long-timescale earth system responses arising from CO2 changes in the first place. This is illustrated schematically as Figure 11d in that paper. To the extent this argument is correct, i.e. to the extent that non GIA related orography changes are small, and which appears to be the case from Dowsett et al. (2016), the above formula should indeed give an estimate of ESS even though on the LHS the forcing is just that from CO2 and on the RHS the total temperature difference includes contributions from various other factors. A minor caveat is that the non-GIA related changes are most pronounced in the infilling of the straits in Northern Canada and the Hudson Bay. But this would only significantly affect local calculations and is not expected to affect the results from the above equation which is based on a global mean.

To respond to point 2:

The reviewer notes that "In these two simulations with the same model, the Eo400-E280 SAT differs by 0.9 °C due to difference in treatment of Pliocene boundary conditions. (No surprise.)" Actually, this difference of 0.9°C between two very similar models was a surprise to us, and we spent a great deal of time trying to understand the exact differences between how the boundary conditions had been implemented in CCSM4-UoT and CCSM4-2deg. This is discussed in lines 205-215 of the manuscript. The reviewer's suggestion to run these two CCSM4 simulations with 560ppm CO2 would provide clarity to this issue, however these simulations are very expensive to run and were not a requirement for a modelling group to contribute to PlioMIP2 (Haywood et al. 2016 – this issue).

While we agree that some uncertainties remain about the CCSM4 models, we do not agree that this translates into uncertainties of our ESS methodology. The disagreement between CCSM4 models is apparent throughout the manuscript – not just in the ESS section. Even with this disagreement the strong correlation between ESS and ECS is sufficient that the PlioMIP2 ensemble *and proxy data can be used to help constrain ECS. This will become more apparent in the revised version due to improvements in the ESS/ECS analysis.*

Specific comments:

Introduction.

The Introduction could be improved. It does a poor job of justifying the rationale for conducting PlioMIP2. A strong case could be made that the PlioMIP2 offers an opportunity to evaluate climate models that have been strongly tuned for the present day, and to showcase advances in modeling since PlioMIP1. Instead, the Introduction (paragraph 2 specifically) is a clumsily written laundry list of all the publications that resulted from PlioMIP1. What's the point?

The reviewer makes some good points here. In the revised version of the manuscript we will rewrite the introduction to incorporate the reviewer's suggestions.

L. 156. The minimum integration length was specified as 500 simulated years. For many models, this is not sufficient time to reach equilibrium. I appreciate that the authors report the spin up time and net TOA radiation in Supplementary Figures and suggest that they add this important information to the manuscript.

The manuscript has been written to appeal to a broad audience, only some of whom will be interested in the spin up time and the net TOA radiation. We think it is better to keep this information in the supplement, where it can be found by interested parties without distracting other readers.

L. 164. Here and elsewhere (e.g. L. 473) the manuscript mentions the release date of the model, and even makes statements like "the model sensitivity is more strongly related to parameterization choices and initial conditions than the release date of the model". As the authors are well aware, model performance is related to the accurate representation of the dynamics and physics and has nothing to do with the date of release. Please remove these confusing and unnecessary comments.

The purpose of this analysis was to determine if developments in model physics lead to altered responses to Pliocene boundary conditions. In particular, whether newer (more recently released) models might show a different sensitivity than older models. We will clarify our meaning in the text.

L. 237. "Lack of consistency in the seasonal signal of warming..." An analysis of the global average seasonality is not very useful (or at least ambiguous) here since the seasons are out-of-phase between hemispheres. Please show hemispheric averages, or just the NH average. Also, this is a place where additional analysis would be appreciated to understand the reason for the model differences.

This is a good point. In the revised version of the paper we will just show NH average. We will also remove the seasonal correlation for ESS in figure 7.

L. 339. "...suggests that there are some inconsistencies between the way in which ECS and ESS were obtained." See my comments above about the estimate of ESS used in the manuscript.

The EC-Earth3.1 model (about which this statement was written) has been withdrawn from the PlioMIP2 intercomparison due to problems in sea ice sensitivity. This sentence will not be required in an updated version of the manuscript.

L. 344. "each of these models provides a different but equally valid realization of ESS..." I don't understand this statement. Please elaborate or delete.

This paragraph will be removed in the revised version of the manuscript. This is because we will be treating all the CCSM4 models as separate models in the ESS section (in the same way we have done in the rest of the manuscript).

Data/Model Comparison. This section (lines 389-400) focuses on comparing the ensemble mean to the proxy data. How consistent are the models? It would be valuable to add a figure showing model agreement, the number of models that were within the criteria for a good fit. In addition, it would be valuable to estimate and report the goodness of fit for each model, as well as the ensemble mean.

We agree that additional information about how individual models compare to the proxy data would be useful. In the revised version, a figure showing how each model compares to the proxy data will be added to the supplementary information.

L. 410. Please calculate and report the global mean DSAT/ DSAT estimate from proxy data for comparison. There are a number of ways to do this that have been reported in the literature.

This is an interesting suggestion, but the spatial distribution of Pliocene SST data is too limited, in our opinion, to make such a process reliable.

Section 5.1. This section is quite interesting. In line 460-464, it is stated that there are differences between the Pliocene and RCP predictions. Please elaborate on these differences, in the same way that the similarities have been described.

Here we were discussing changes in boundary conditions between the PlioMIP2 simulations and RCP simulations. We will make this clear.

L. 489. "Previous DMCs for the Pliocene..." I don't understand this sentence. Please clarify.

This sentence will be rewritten in a revised version.

L. 545. "...suggest that SST data from the Pliocene tropics has the potential to constrain model estimates of ECS..." The discussion that follows (L. 550-577) about equilibrium climate sensitivity is not robust. ECS is calculated from a handful of local points from the same regions and is justified because it agrees with the ECS reported in IPCC, the exact value that this analysis should be testing.

This discussion arose from two points:

1. *The Pliocene temperature anomaly at many model gridpoints is correlated with the model's ECS.*
2. *Some of the gridpoints where this correlation occurs have proxy data – providing an independent estimate of the Pliocene climate.*

While models provide a range of estimates of climate sensitivity, combining points 1 and 2 allow us to estimate climate sensitivity from proxy data at certain locations.

We did not intend to justify our method based on it agreeing with IPCC, rather we intended to compare our results with IPCC estimates (and also with earlier PlioMIP1 studies) for completeness. We will rewrite the text around line 571-572 to make this clear.

The additional models that have been added to PlioMIP2 since the initial submission improve the relationship between Pliocene temperature anomaly and ECS. This means that the number of proxy data points that can be used to estimate ECS will increase in a revised version, and will cover a wider region.

L. 629. This is not a conclusion of this study, and therefore shouldn't be included in the Conclusions.

We agree with this point and this conclusion will be removed in a revised version

Response to Tim Herbert (reviewer 3):

We would like to thank Tim Herbert for providing helpful comments on this manuscript, which will help improve a revised version. We provide a response below to those comments that have been incorporated into version 1 of the manuscript:

[TDH1]: This is absolute scaling, but it would be more useful to provide a context: either add the absolute mean SAT, SST, or the scaling between delta SST and delta SAT(land). [line 41]

In the revised version we will specify by how much SAT increases over land and how much it increases over the ocean.

[TDH2]: Clarify? Meaning of "constraints" not evident. [line 44]

In the previous version there was a statistically significant relationship between a model's Pliocene temperature response and the Equilibrium Climate Sensitivity when we excluded the EC-Earth3.1 model and averaged together the CCSM4 models. These were the modelling constraints we referred to.

However, since the previous version of the manuscript the EC-Earth3.1 model has been withdrawn from PlioMIP2 and two further models have been added. This means that we now obtain a statistically significant relationship between a model's Pliocene temperature response and the Equilibrium Climate Sensitivity without any 'modelling constraints'. This sentence is therefore no longer needed in the revised version.

[TDH3]: Data shows polar amplification well equatorward of +-60 so this suggests model not able to capture all the physics? [line 43]

We reported the polar amplification as the ratio of warming poleward of 60° to the global mean warming. This is a standard metric for calculating polar amplification (Smith et al 2019). We do not state that there is no polar amplification equatorward of 60°, and figure 1b also shows that polar amplification occurs equatorward of 60°. Therefore, our paper does not imply model-data disagreement or any problems with model physics in this respect.

[TDH4]: This is all in the context of a constant 400 ppm forcing? E.g. no uncertainty in pCO2? Relevant to deducing ECS from SST- requires pCO2 estimate, correct? [line 48]

This is a good point and in a revised version we will add some further discussion about CO_2 uncertainty to the text.

The ECS estimate uses two inputs:

1. *The data, which we assume comes from a 400ppm CO_2 world. However, we note that uncertainties may mean the data represents a world where CO_2 was slightly different.*
2. *The models which were all run with CO_2 of 400ppmv.*

Given that we do not currently have a range of model simulations with different CO_2 values, the only possible way of estimating ECS requires using regressions derived from a CO_2 = 400ppmv modelling world. We therefore currently have no other option other than to assume that the data also represents CO_2 of 400ppmv. If the data represented a world with a different CO_2 value it could not be related to the model outputs.

However, the PlioMIP2 model design did accept that there are uncertainties on the KM5c CO_2 value (Haywood et al; this issue) and suggested that modelling groups also carried out experiments with CO_2 set to 350ppmv and 450ppmv in order to quantify CO_2 uncertainties. As the simulations with different CO_2 values become available, we will be able to add reliable error bars to our ECS estimates that will account for CO_2 uncertainty.

[TDH5]: Of course, given present proxy CO2 data, we don't really know if this window was say +- 20 ppm higher than the canonical 400 ppm… [line 115]

Please see response to TDH4.

[TDH6]: I would have thought that reduction in winter sea ice and/or lower land surface albedo would have generated a larger winter warming relative to the mean anomaly. [lines 230-240].

The amount of winter sea ice has very little effect on temperature. This is because there is no sunlight over the winter pole, and so the value of the surface albedo is irrelevant.

The reduction in winter snow cover away from the pole will affect only be a small proportion of the northern hemisphere surface. Therefore, it can only have a limited effect on hemisphere averaged temperatures.

[TDH7]: I am a bit mystified here on the Foley/Dowsett data source precision. For example, see Figure 9 of Caballero-Gill et al.: at 3.205 Ma, SST at Site 1125 is ~21oC and 594 is 14oC. The Foley/Dowsett table gives 19.5 and 12.2 respectively. [lines 382-387]

*Figure 9 of Caballero-Gill et al, includes data from 2.6Ma – 4.2Ma and as such data near KM5c only represents a very small proportion of the figure. The figure is therefore not of sufficient temporal resolution to obtain SST directly. The dataset upon which the figure is based (*https://doi.pangaea.de/10.1594/PANGAEA.898162) *shows that SST does nearly reach 21°C at 3.213Ma, however at 3.205 the SST is 19.5 as reported by Foley and Dowsett. A similar argument could be applied to site 594.*

[TDH8]: Suggest moving to McClymont et al. data set in lieu of comments above? [line 397]

Sites which are included in both Foley and Dowsett and McClymont et al. datasets generally show the same or very similar SST estimates. There are some very small differences (0.1-0.2°C) which are undoubtedly due to the different time windows used (10ky and 30ky in Foley and Dowsett, 20ky in McClymont et al.). In some cases, McClymont et al. used previously unpublished data that could not have been included in Foley and Dowsett.

As modellers, we need to validate our models against a wide range of different datasets. Here we choose to validate against the Foley and Dowsett dataset because the model results are already compared with the McClymont dataset elsewhere (McClymont et al., this issue). However, we note that the first order outcomes of the PlioMIP2 model-data comparison is the same, regardless of whether model results are compared with the Foley and Dowsett dataset of the McClymont et al. dataset. This will be noted in a revised version of our manuscript.

[TDH9]: To me, the pattern of data anomalies exceeding model anomalies near the gyre boundaries is quite robust (see MyClymont or example) and is a general feature of "warm climate" reconstructions- see Brierley for an earlier Pliocene time slice, and many others for the Miocene and Eocene. I think there is a fundamental model deficiency here. [line 401-406]

We agree that there could be a fundamental model deficiency in these regions and have stated near line 520 "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". However, the interpretation of data in upwelling regions is also not trivial and we also discuss this in our paper.

[TDH10] and [TDH11]: See comments above – I think this SST data set is particularly problematic. [Lines 480-496]

Please see response to TDH7 and TDH8.

[TDH12]: F&D data set is more likely to be the issue…[lines 498-503].

Here the reviewer is referring to our analysis which shows that the mPWP-PI temperature anomaly can depend on which dataset we use to represent the preindustrial data. We note that if two different datasets give different values for the PI climate, than these datasets will lead to different values of the mPWP-PI temperature anomaly. This difference will be independent of which dataset was chosen to represent the mPWP. Because this paragraph was discussing the choice of preindustrial dataset, we do not agree that the mPWP F&D dataset will be an issue here.

[TDH13]: However this is not born out with modern sediments in the region so I think this is special pleading! Large compilations, the most recent being Tierney and Tingley (2018) do not identify an upwelling bias in the modern data set. [line 540]

The sentence referred to in the comment is:

"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)."

The sentence was written as a possible explanation for the data-model mismatch. We also give other possible reasons for the data-model mismatch (i.e. that the model's struggle in upwelling regions).

Although there is not a seasonal upwelling bias in the modern dataset, it is not implausible that such a bias could exist in the Pliocene. In a revised version of the manuscript we will continue to include this as a possible reason for the Pliocene model-data mismatch but will also include that no such upwelling bias exists in the modern dataset in order to include additional information for the reader.

[TDH14]: My concern here would be how much the use of the F&D data set alters this new estimate in comparison to Hargreaves and Annan? [lines 555-565]

This point refers to how the estimate of ECS depends on the dataset chosen. We agree that the estimate of ECS depends on the dataset chosen, but note that it also depends on the relationship between the published ECS and the Pliocene warming in the models. We suggest that using the new F&D dataset, instead of the PRISM3 SST anomaly field used by Hargreaves and Annan would improve the estimate of ECS. This is because the PRISM3 SST anomaly field was based on warm peak averaging over a 240,000-year timeslab while the F&D dataset better represents the MIS KM5c timeslice that our models are set up to represent.

However, we will note in the revised version of the manuscript that as more orbitally tuned SST data becomes available it will be important to revisit the ECS analysis to ensure maximum accuracy.

[TDH15]: The quality of the estimation also depends on the reliability of the paleo-CO2 estimates

We agree that the relationship between ECS and ESS will only be robust if we are using the correct palaeo CO_2 *value. Please see the response to TDH4, which explains why we are not able to quantify errors due to* CO_2 *uncertainties at the moment, but how we plan to do this in the future.*

The Pliocene Model Intercomparison Project Phase 2: Large scale climate features and constraining climate sensitivity.

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 1.7~~4~~ and 5.2~~4.7~~°C relative to pre-industrial with a multi-model mean value of ~~2.8~~ 3.2°C. Annual mean total precipitation rates increase by ~~6~~ 7% (range: 2%-13%). On average, surface air temperature (SAT) increases by~~are 1.3~~ 4.3°C ~~greater~~ over the land and 2.8°C~~than~~ over the oceans. T~~, and t~~here 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 2.3~~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, t~~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 E~~e~~arth system response to doubling of CO_2 (including ice sheet feedbacks) is ~~approximately 50%~~ 67% greater than ECS, this is larger than the increase of 47% obtained from the~~consistent with results from the~~ PlioMIP1 ensemble. Proxy-derived estimates of Pliocene sea-surface temperatures are used to assess model estimates of ECS and give a range of ECS between~~indicate a range in ECS from~~ 2.6 and 4.8~~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 pre-industrial, 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 driving polar amplification of the warming. ~~.~~ The reduced meridional temperature gradient influenced atmospheric circulation in a number of ways, ~~.~~ such as the poleward shift of the~~For example,~~ m~~M~~id-latitude westerly winds ~~shifted poleward~~ ~~, a response consistent with a poleward shift in meridional circulation~~ (Li et al., 2015). In addition, 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 some~~specific~~ climate models also a weaker Walker Circulation, ~~. This~~ a 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. Li et al. (2018) reported that the global land monsoon system during the MP simulated in the PlioMIP1 ensemble generally expanded poleward with increased monsoon precipitation over land.

The modelled response in ocean circulation was also studied in PlioMIP1. ~~he response of~~ T~~t~~he Atlantic Meridional Overturning Circulation (AMOC) was analysed by Zhang et al. (2013~~4~~). 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 determined to be predominantly a response to direct CO_2 forcing, while at high-latitudes changes clear sky albedo became the dominant contributor, with the warming being only partially offset by cooling driven by cloud albedo changes. ~~Hill et al. (2014) analysed the dominant components of MP warming across the PlioMIP1 ensemble using an energy balance analysis. Instead of enhanced OHT as a driving mechanism warming iIn the tropics, increased temperatures were determined to be predominantly a response to direct CO2 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 ~~-~~ (Pliocene Research, Interpretation and Synoptic Mapping version 3) 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 was used to~~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~~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 Shelves~~f.~~. ~~In order to~~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~~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~~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 climate 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). We also present key differences between PlioMIP2 and PlioMIP1. 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)~~.~~. They~~is~~ include~~s~~ 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~~ (enhanced boundary condition version).~~. 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). However, we acknowledge that there are uncertainties on the KM5c CO_2 value, hence the specification of Tier 1 PlioMIP2 experiments (Haywood et al. 2016; this volume), which have CO_2 of ~350 ppmv and 450ppmv ~~that~~ will be ~~are also planned~~ used to ~~to~~ investigate CO_2~~isese~~ uncertainty at a later date~~in the future~~ies. 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

There are currently 16~~Fifteen~~ climate models that have completed the *Plio*$_{Core}$ experiment to comprise the PlioMIP2 ensemble. These models were developed at different times and have differing levels of complexity and spatial resolution. A further model HadGEM3 is currently running the *Plio*$_{Core}$ experiment and results from this model will be compared with the rest of the PlioMIP2 ensemble in a subsequent paper. ~~, which were developed at different times, having differing levels of complexity and spatial resolution, completed the *Plio*$_{Core}$ experiment to comprise the PlioMIP2 ensemble.~~ The current~~,~~ 16 model~~,~~ ~~is~~ ensemble is ~~almost~~ double the size of the coupled atmosphere-ocean ensemble~~at~~ 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 modelling 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 400 ppmv 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 560 ppmv-280 ppmv; ESS) can then be estimated as follows:

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

T~~Strictly speaking, t~~here 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 likely 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 lake~~s~~ distributions in a climate model had an ~~would have an~~ insignificant effect on global temperature (even though changes regionally ~~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 canould 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.~~To assess the relationship across the ensemble between the reported ECS and the modelled ESS, we correlate reported ECS across the ensemble with the associated $Plio_{Core}$ - PI_{Ctrl} temperature anomalies. We do this on a global, zonal mean, and local scale. A strong correlation at a particular location would suggest that MP proxy data at that location could be used to derive a proxy-data constrained estimate of ECS (similar to an "emergent constraint"), 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)

Fig. 1a shows the global mean surface air temperature (SAT) for each model. The top panel shows the $Plio_{Core}$ and PI_{Ctrl} SATs while the lower panel shows the anomaly between them. In this, and all subsequent figures, the models are ordered by ECS (see Table 2) such that the model with the highest published ECS (i.e. CESM2; ECS=5.3) is shown on the~~to the~~ left, while the model with lowest published ECS (i.e. NorESM1-F; ECS = 2.3) is on the right. Increases in $Plio_{Core}$ global annual mean SATs, compared to each of the contributing models PI_{Ctrl} experiment, range from 1.7 to 5.2 °C (Fig. 1a; Table 2), with an ensemble mean ΔT of 3.2~~2.8~~°C. The multi-model median ΔT is 3.0°C, while the 10th and 90th percentiles are 2.1°C and 4.8°C, respectively. Analogous results from individual models of the PlioMIP1 ensemble are shown by the grey horizontal lines on Fig. 1a, and have a mean warming of 2.7°C. Pliocene warming for individual PlioMIP1 models falls into two distinct anomaly bands that~~these~~ are~~: ~~ 1.8 - 2.2°C (CCSM4, GISS-E2-R, IPSLCM5A, MRI2.2) and 3.2 - 3.6°C (COSMOS, HadCM3, MIROC4m, NorESM-L). ~~With the exception of~~~~Apart from~~In general, PlioMIP1 models that were in the lower anomaly band show a larger temperature anomaly in PlioMIP2, while those in the upper anomaly band show a lower temperature anomaly in PlioMIP2. The only exception to this is COSMOS, which is the only PlioMIP2 model to use dynamic vegetation (Table 1), the effect of dynamic vegetation on temperature anomalies is discussed in Stepanek et al. (2020). PlioMIP2 shows a greater range of responses than PlioMIP1, and PlioMIP2 results are more evenly scattered over the ensemble range. The larger ensemble mean in PlioMIP2 is due to the addition of new and more sensitive models, rather than an increase in the temperature anomaly due to the change in boundary conditions.

PlioMIP2 shows i~~I~~ncreased SATs~~ are predicted~~ over~~throughout~~ the whole globe (Fig. 1b), with an ensemble average warming of ~2.0~~1.7~~°C for the tropical oceans (20°N-20°S), which increases~~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 (G~~R~~IS), and the simulated reduction in Northern Hemisphere sea-ice cover (de Nooijer et al., 2020~~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 CESM2 model has the greatest apparent sensitivity to imposing MP boundary conditions with a simulated ΔT of 5.2 °C (Fig 1a). ~~, t~~This model was published in 2020 and has the highest ECS of all the PlioMIP2 models. This model was not included in PlioMIP1, and ~~it's~~its response to Pliocene boundary conditions lies outside the range of all PlioMIP1 models both in global mean and for every latitude band (F~~f~~igure 1a, 1c). I~~i~~t is also warmer than the PlioMIP2 multi-model mean at nearly all gridboxes (S~~s~~upplementary Fig. 1). Other particularly sensitive models~~,~~ (EC-Earth3.3, CESM1.2, CCSM4-Utr and CCSM4-UoT;~~)~~ ~~(~~shown as an anomaly from the multi-model mean in S~~s~~upplementary Fig. 1) are also new to ~~PlioMIP2 and~~PlioMIP2 and this explains why the simulated ΔT from PlioMIP2 exceeds that from PlioMIP1. The model with the lowest response to PlioMIP2 boundary conditions is the NorESM1-F model, which is also the model with the lowest published ECS. Although there is clearly some correlation between a model's ECS and its $Plio_{Core} – PI_{Ctl}$ temperature anomaly, the relationship is not exact. In particular, the versions of CCSM4 that were run by Utrecht University (CCSM4-Utr) and the University of Toronto (CCSM-UoT) both show a large Pliocene response but have a modest ECS compared to the other models.

Three different versions of CCSM4 contributed to PlioMIP2 (see T~~t~~able 1): the standard version run at NCAR (hereafter referred to as CCSM) has a simulated ΔT = 2.6°C, while CCSM4-Utr has a simulated ΔT = 4.7°C and CCSM4-UoT has a simulated ΔT = 3.8°C.~~ 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 60°S - 90°S band ~~Southern Hemisphere (SH) at ~70°S,~~ where the ~~zonal~~ mean warming in the CCSM4-Utr~~2deg~~ simulation is 4~~.0~~ °C higher than in the CCSM4-UoT simulation and 6.6 °C higher than in the CCSM4~~-1deg~~ simulation (Fig. 1c; Supplementary Fig. 1). Supplementary T~~t~~able 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 similar~~the same~~ ECS (T~~t~~able 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-Utr~~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 parti~~rti~~cular 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 80% ~~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. ~~, ~~I~~i~~t 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 Northern Hemisphere (NH) averaged SAT anomaly over the seasonal cycle ~~of the SAT anomaly for each model~~ is presented in Fig. 3a. Overall, the ensemble mean anomaly (black dashed line) is fairly constant throughout the year,~~ displays a small enhancement in SAT warming in the Northern Hemisphere (NH) summer and early autumn.~~ ~~H~~however, models within the ensemble have very different characteristics in terms of the monthly and seasonal distribution of the warming. Some members of the ensemble have a relatively flat seasonal cycle in ΔSAT (e.g. NorESM-L, NorESM1-F, COSMOS), however others show a very strong seasonal cycle. The models that show a very strong seasonal cycle do not agree on the timing of the peak warming. For example, EC-Earth3.3 has the peak warming in October, CESM2 has peak warming in July and MRI2.3 has peak warming in Jan/Feb, ~~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 grey shaded area in Fig 3a shows the range of NH temperature response in PlioMIP1, with the PlioMIP1 ensemble average shown by the black dotted line. Although the ensemble average from PlioMIP2 and PlioMIP1 both show a relatively flat seasonal cycle, the range of responses is very different between the two ensembles. PlioMIP1 predicted a large range of temperature responses in the NH winter, which reduced in the summer. In PlioMIP2, however, the summer range is amplified compared to the winter. Indeed 7 of the 16 PlioMIP2 models show a NH summer temperature anomaly that is noticeably above that seen in any of the PlioMIP1 simulations. Some of these models (CESM2, EC-Earth3.3, CCSM4-Utr, CCSM4-UoT and CESM1.2) did not contribute to PlioMIP1, showing that which models are included in an ensemble can strongly affect the ensemble response. However other models (MIROC4m and HadCM3) that show an enhanced summer response in PlioMIP2 were also included in PlioMIP1, showing that there is also an impact of the change in boundary conditions on seasonal temperature. None of the PlioMIP2 models ~~are able tocan~~replicate the lowest warming seen in DJF in the PlioMIP1 ensemble, this lowest value was derived from the GISS-E2-R model in PlioMIP1 which did not contribute to PlioMIP2.

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.~~ The land amplification factor is similar in PlioMIP2 and PlioMIP1, and models in both ensembles cluster near a land amplification factor of ~1.5. There is also no relationship between a model's climate sensitivity and the land amplification factor. ~~were included in,.~~ The multi-model median (10th percentile / 90th percentile) warming over the land and ocean is 4.5°C (2.6°C / 6.1 °C) and 2.5°C (1.9°C / 4.4 °C) respectively. ~~S~~

The extratropical NH (45°N-90°N) warms more than the extratropical Southern Hemisphere (SH) (45°S-90°S) in 5 of the 8 models (62%) from PlioMIP1 and in 11 of the 16 models (69%) from PlioMIP2 (Fig. 3c). This shows that neither the change in boundary conditions nor the addition of newer models to PlioMIP2 affects the ensemble proportion of enhanced NH warming. Neither does the published ECS have any obvious impact on whether the warming is concentrated in the NH or the SH. ~~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 (CCSM4-Utr, GISS2.1G, NorESM-L), are among 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.~~(Fig 3b).

Polar amplification (PA) can be~~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 mean (median)~~average~~ PA is 2.3 (2.2) ~~and is similar~~ in both the NH and the SH, suggesting that across the ensemble PA is hemispherically symmetrical. This result is very similar to PlioMIP1 (not shown), which suggests that the enhanced warming in the PlioMIP2 ensemble does not affect the PA. For PlioMIP2, the NH median PA is 2.2, with the 10th and 90th percentiles at 1.9 and 2.8 respectively, while in the SH the median PA is 2.2, with the 10th and 90th percentiles at 1.8 and 3.1 respectively. Polar amplification is lower over the land than the ocean (Supplementary Figure 2) in both hemispheres. The NH mean (10th / 50th / 90th percentiles) PA is 1.6 (1.4 / 1.6 / 1.9) and 2.7 (2.4 / 2.7 / 3.3) over the land and ocean respectively, while the SH mean (10th / 50th / 90th percentiles) PA is 0.9 (0.5 / 0.8 / 1.5) and 1.9 (1.1 / 1.9 / 2.5) over the land and ocean respectively. Note that in the SH total PA is higher than both land and ocean PA because of the change in the area of land between the $Plio_{Core}$ and PI_{Cntl} experiments. There appears to be a weak relationship between the PA factor and a model's ECS. Those models which have a lower published ECS (those to the right of Fig. 3d) have a tendency towards higher PA. This is not because these models have excess warming at high latitudes, rather these models have less tropical warming than other models.

~~.SFF, rather~~

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 zonal 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 and~~&~~ Fedorov, 2014; Tierney et al., 2019).

The multi-model mean meridional profile of zonal mean SSTs in 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 1.5-2.5°C. This difference increases to ~5.0°C in the NH at ~55°N, with an inter-model range of 2°C - 11°C. The Pliocene and Pre-industrial meridional SST profile in the Pacific (Fig. 4b) is similar to that of the Atlantic, but with little indication from the multi-model mean for a high latitude enhancement in meridional temperature. However, a large range in the ensemble response is noted, and the importance of an adjustment of the vertical mixing parameterization towards simulation of a reduced Pliocene meridional gradient has been recently shown (Lohmann et al., in review).

~~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.°C. This difference increases to ~.0°C in the NH at ~55°N, where the range in warming across the models also increases to between °C and °C. The Pliocene and Pre-industrial meridional SST profiles in the Pacific isare similar to that of the Atlantic, but with little indication from the multi-model mean for a high latitude enhancement in meridional temperature in the Pacific compared to the Atlantic (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 ~1.9°C (ensemble range from 0.8°C to 3~~.~~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 both 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~~ is not consistent with a permanent El-Niño (see Supplementary Fig. 3). The PlioMIP2 ensemble supports a recent proxy-derived reconstruction for the Pacific that~~, through careful assessments of proxy uncertainty,~~ found Pliocene ocean temperatures increased in both the eastern and western Tropical Pacific ~~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 robust over ~~significant in~~ nearly all ~~most~~ ocean grid cells (Supplementary Fig. ~~S~~4). Supplementary Fig. ~~S~~3 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 being larger than the inter-model standard deviation ~~across the ensemble~~ there are still many regions (e.g. Southern Ocean, North Atlantic Ocean, Arctic Ocean) where there is a notable~~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 (shown as horizontal grey lines on Fig. 5a). The PlioMIP2 ensemble mean ΔPrecip is 0.19 mm/day. The increase in the globally averaged precipitation anomaly in PlioMIP2 is due to the addition of new models to the ensemble, which have high ECS and are also more sensitive to the PlioMIP2 boundary conditions. Models that were included in PlioMIP1 (COSMOS, IPSLCM5A, MIROC4m, HadCM3, CCSM4, NorESM-L and MRI2.3) show PlioMIP2 precipitation anomalies that are similar to PlioMIP1 results. 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 (W~~w~~est 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 (CESM2, CCSM4-Utr~~2deg~~, EC-Earth3.3~~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 inter-model coherence in many places because at least 80% of ~~12 of the 15~~ models agree on the sign of change. However, most of 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). ~~,~~ T~~t~~his 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. ~~S~~5 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

Figure 6a shows the seasonal cycle of the precipitation anomaly averaged over the Northern Hemisphere. As was the case for SAT, the monthly and seasonal distribution of precipitation anomalies are highly model dependent, although the ensemble average shows a clear~~small~~ NH late spring to autumn *Plio*$_{Core}$ enhancement in precipitation (Fig. 6a)~~.~~. This is most strongly evident ~~evident~~ in the models CESM2, EC-Earth3.3 and CCSM4-Utr, however it is also evident in other models~~CSM4-2deg, CCSM4-1deg, CESM1.2, CCSM4-UoT and IPSLCM6A~~. Some models show a different seasonal cycle to the annual mean, ~~Most other models have no clear structure to the monthly distribution of precipitation changes, other than~~ for example the GISS2.1G model ~~that~~ simulates the NH late spring to autumn ΔPrecip being supressed compared to the rest of the year, and HadCM3 which has a ~~biomodal~~bimodal distribution. An increase in NH summer precipitation is consistent with a general trend of West African, Indian and East Asian Summer monsoon enhancement, and this will be discussed in detail in a forthcoming PlioMIP2 paper. In PlioMIP1 (ensemble average - dotted black line and model range – shaded grey area in Fig. 6a) the seasonal cycle in precipitation was much more muted. PlioMIP1 results in the boreal winter are similar to PlioMIP2, however the mean precipitation anomaly in PlioMIP2 between June and November is 40% larger than PlioMIP1. This increase is mainly due to the inclusion of new and more sensitive models into PlioMIP2 (e.g. CESM2 and EC-Earth3.3).~~;~~ H~~h~~owever, some models with enhanced summer precipitation (e.g. COSMOS) contributed to both PlioMIP1 and PlioMIP2 suggesting a role of boundary condition changes in enhancing the NH boreal summer precipitation. It is noted, however, that not all the new models in PlioMIP2 show enhanced summer precipitation relative to PlioMIP1. The GISS2.1G model, which was new to PlioMIP2, shows the most muted summer precipitation response in the NH in all PlioMIP2 and PlioMIP1 models. This means that the range of summer/autumn NH precipitation responses as shown by the ensemble increases significantly in PlioMIP2. For example, PlioMIP1 showed a NH precipitation response in October to be 0.13-0.42 mm/day while in PlioMIP2 this has increased to 0.05-0.70 mm/day.

In terms of the land/sea ΔPrecip contrast the PlioMIP2~~LIOMIP2~~ 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 (EC-Earth3.3, MIROC4m, HadCM3, CCSM4, CCSM4-Utr, CCSM4-UoT, NorESM-L and NorESM-F)~~CCSM4-1deg, CCSM4-2deg, EC-Earth3.1, HadCM3, MIROC4m, NorESM1-F, NorESM-L and CCSM4-UoT),~~ and the other where there is either a small or~~is~~ no ~~clear~~ enhancement in the land versus oceans (CESM2, IPSLCM6A, COSMOS, CESM1.2,~~ COSMOS,~~ IPSLCM5A, IPSLCM5A2, GISS2.1G and MRI2.3).~~, 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.~~ Models which show the greatest precipitation enhancement over the land are generally those with a lower published ECS (those to the right of Fig 6b), which have a small precipitation response over the oceans but have a land precipitation anomaly similar to other models. Models with higher ECS (e.g. CESM2) show a similar precipitation anomaly over the land and ocean. Grey horizontal lines on F~~f~~ig. 6b shows the land/sea ΔPrecip amplification for the PlioMIP1 models. None of the PlioMIP1 models have a ΔPrecip amplification factor > 2, however half of the PlioMIP2 models do. Further, four models which contributed to both PlioMIP1 and PlioMIP2 (MIROC4m, HadCM3, CCSM4, NorESM-L) show a much greater land amplification in PlioMIP2, showing that the change in boundary conditions strongly affects this diagnostic.

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 T~~t~~able 1) and the ESS estimated from the $Plio_{Core} – PI_{Ctrl}$ temperature anomaly (equation 1). Due to the prescribed changes to ice sheets and vegetation, the $Plio_{Core}$ simulation is representing a state in which the associated feedbacks are in equilibrium. The mean ESS / ECS ratio is 1.67~~55~~, suggesting that the ESS based on the ensemble is 67~~55~~% larger than the ECS, however the range is large with the GISS2.1G model suggesting that the ESS / ECS ratio is 1.22 while the CCSM4_Utr model suggests that the ESS / ECS ratio is 2.85. ~~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. This is seen i~~o~~n ~~(~~Fig. 7a, and each cross represent the results from a different model in the PlioMIP2 ensemble. ~~).~~ There is a significant relationship between ECS and the $Plio_{Core}$ - PI_{Ctrl} temperature anomaly ~~the two~~

at the 95% confidence level ($p=0.0$1, $R^2=0.35$~~3~~) with the line of best fit: ECS = 2.3 + (0.44 × ($Plio_{Core}(SAT) - PI_{Ctrl}(SAT)$)). ~~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 whether there is a correlation across the ensemble between ECS and the $Plio_{Core} – PI_{Ctrl}$ SAT anomaly on spatial scales. ~~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) across the ensemble between ~~the~~ modelled~~annual mean~~ ECS and the modelled zonal mean~~the~~ $Plio_{Core} - PI_{Ctrl}$ SAT anomaly~~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 38~~23~~°N and 27~~12~~°S, where a high proportion of the inter-model variability in global ECS can be related to the inter-model variability in ~~the latitudinal average~~ the Pliocene SAT anomaly at an individual latitude, reaching a maximum of 65% at ~15~~10~~°N.

Next the relationship between global ECS and the local~~point based~~ $Plio_{Core} – PI_{Ctrl}$ SAT anomaly is~~was~~ assessed. ~~. This follows a similar approach to that of Hargreaves and Annan (2016).~~ In Fig. 7c~~d~~ colours show the *R-squared* correlation across the ensemble between ~~a~~ modelled~~'s~~ global ECS and modelled local~~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 local~~grid box~~ $Plio_{Core} – PI_{Ctrl}$ SAT anomaly is significant over most ~~large parts~~ of the tropics ~~(e.g. the tropical Pacific and Northern Africa)~~, and over some mid and high latitude regions including 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 time_slice~~.~~. 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 of 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). Age models used in the compilation are those originally released with the data sets, but later modifications of age models or the integration of additional data could result in mean SST values different from those reported in F&D19_30. All SST estimates are calibrated using Müller et al. (1998). 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 include~~s~~ 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 multi-model mean ~~ensemble~~ does not predict the scale of warming seen in 3 of the 4 ~~the~~ proxy reconstructions. The multi-model mean~~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 and Mediterranean regions) 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~~MIS KM5c time slice (Fig. 8d) yields a similar outcome to the comparison of SST anomalies (Fig. 8c). However, the Benguela region shows 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.

The proxy data ΔSST minus the mean ΔSST for individual models is shown in S~~s~~upplementary Fig. 6. In regions where there was a strong discrepancy between the proxy data ΔSST and the multi-model mean ΔSST none of the individual models show good model-data agreement. The EC-Earth3.3 model shows an improved agreement with the data in the Mediterranean, the Benguela upwelling system, the site along the East Coast of North America and the site to the West of Svalbard. However, this improved model-data agreement is at the expense of reduced model-data agreement elsewhere: many of the low and mid-latitude sites, which had good model-data agreement for the multi-model mean have reduced model-data agreement in EC-Earth3.3. Other models, which showed large warming in PlioMIP2 (i.e. CESM2 and CCSM4-Utr) also show a larger ΔSST than the data for some of the tropical and mid-latitude sites which were in good agreement with the multi-model mean. Models that were less sensitive to Pliocene boundary conditions (i.e. GISS2.1G and NorESM-L) do not predict the amount of warming seen in the data for some of the North Atlantic sites and the multi-model means performs better. Table 3 shows statistics for the data-model comparison for both individual models and the multi-model mean. The root mean square error (RMSE) between the model and the data is 3.72 for the multi-model mean, but is lower in some individual models (namely CESM2, IPSLCM6A, EC-Earth3.3, CESM1.2, CCSM4-UoT). In general, those models that have a lower model-data RMSE are those which have higher ECS and a higher $Plio_{Core} – PI_{Ctrl}$ warming, while less sensitive models have a higher model-data RMSE. The average difference between the data and model across all the data points shows a similar pattern. The proxy data is on average 1.5°C warmer than the multi-model mean. ~~, h~~However, some individual models have a much smaller average model-data discrepancy (e.g. CESM2 = -0.18°C). The models with a lower model-data discrepancy are those which also have a lower model-data RMSE and have higher than average $Plio_{Core} – PI_{Ctl}$ warming.

This initial analysis suggests that the most sensitive models agree better with the proxy data than the less sensitive models. However further analysis does not fully support this result. If we consider how many of the 37 sites have ‘good’ model-data agreement a different picture emerges. Table 3 shows how many sites have model ΔSST within 2°C, 1°C and 0.5°C of the data ΔSST. Using these diagnostics, the MMM performs better than any of the 16 individual models. Those models which have the lowest RMSE and the best average model-data agreement are not those models which have largest number of sites where model and data agree. For example, CESM2 and EC-Earth3.3 have a particularly low number of sites with good model-data agreement. The models with the highest number of sites with model-data agreements (e.g. ISPLCM6A and CCSM4-UoT, MIROC4m and CESM1.2) show a $Plio_{Core} – PI_{Ctl}$ warming that is closer to the MMM. The fact that the MMM has more sites with ‘good’ model-data agreement than any individual model, highlights the benefit of performing a large multi-model ensemble as we have done for PlioMIP2. It allows inherent biases within individual models to cancel out and likely provides a more accurate way of estimating climate anomalies than can be done with a single model.

~~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.~~ Models show a strong relationship between SST anomalies and global mean SAT anomalies (S~~s~~upplementary F~~f~~igure 7a; SATanom = (1.18 × SSTA) + 0.66, Rsq=0.97); and also a strong correlation between SST averaged over 60°N-60°S and global mean SAT anomalies (S~~s~~upplementary F~~f~~igure 7b; ΔSAT = (1.16 × ΔSST) + 0.74, with Rsq=0.97). This strong correlation suggests that proxy-based SST anomaly estimates can be used to infer global mean SAT anomalies, provided that enough SST proxy data is available to reliably estimate SST anomalies. The multi-model median ratio of ΔSAT / ΔSST is 1.4, while the multi-model median ratio of ΔSST to ΔSST (60°N-60°S) is 1.5.

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 1.7 to 5.2°C) is akin to the best estimate (and uncertainty bounds) of predicted global temperature change by 2100CE using the RCP4.5 to 8.5 scenarios (RCP4~~2~~.5 = 1.8 ± 0.5 °C and RCP8.5 = 3.7 ± 0.7 °C IPCC, 2013; T~~t~~able 12.2). Comparing the degree of Pliocene temperature change to predicted changes at 2300 CE, the multi-model mean SAT change is~~-~~ between RCP4.5 (2.5 +/- 0.6 °C) and RCP6.0 (4.2 +/- 1.0 °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). ~~-~~All 16 ~~Fourteen out of the fifteen~~ models within the PlioMIP2 ensemble simulate a polar amplification factor (PA; averaged over the NH and SH) between 2 and ~~of ≤~~ 3 (meaning that the high latitude temperature increase is ~~less than~~ 2-3 times the global mean temperature increase) ~~), only EC-Earth3.1 shows a PA > 3. H~~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 SH~~outhern Hemisphere~~ far more than the NH~~orthern 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 perturbations to the pre-industrial~~PI~~, modelling and observational studies have shown that land warms 30 to 70% more than the oceans (Lambert and~~&~~ 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, 11~~9~~ out of 16~~5~~ 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 3.2~~2.8~~°C increase in multi-model mean temperature is associated with a 7~~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 (F~~f~~igure 5d~~3d~~)~~.~~. These regions also experience enhanced precipitation under the RCP8.5 scenario for 2100 (IPCC, 2013; F~~f~~igure SPM.7)~~.~~. However, in other tropical and subtropical regions the $Plio_{Core}$ model response is small compared to the pre-industrial inter-model standard deviation.

Corvec and Fletcher (2017) showed that in PlioMIP1 studies the tropical overturning circulations in the mPWP~~mPWP~~ were weaker than pre-industrial 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 F~~f~~igure 5 for the $Plio_{Core}$ ensemble and in IPCC, (2013; F~~f~~igure 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), despite the different boundary conditions.

It has been seen that some of the main differences between the PlioMIP1 and PlioMIP2 ensembles are due to the inclusion of new models in PlioMIP2 that were not available at the time of PlioMIP1. We therefore assess whether recent developments in model physics lead to altered responses in Pliocene boundary conditions, in a statistically significant way. In particular, we assess whether ~~if~~newer models predict a larger Pliocene response than older models. Across the ensemble, model sensitivity to Pliocene boundary conditions does not appear to correlate with the release date of the model (left panels of S~~s~~upplementary F~~f~~igure ~~S~~8; 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 CESM2 model (release date 2020) is more sensitive than the CESM1.2 model (release date 2013), which in turn is more sensitive than CCSM4~~-1deg~~ (release date 2011), when all~~both~~ are run with the same resolution, boundary and initial conditions. However, ~~other models from the~~ CCSM4-Utr ~~family (CCSM4-UoT and CCSM4-2deg; both with~~ (release date 2011) is also very sensitive~~are the most sensitive of this family~~. This~~,~~ ~~suggesting~~ shows that within the CCSM/CESM family,~~that~~ model sensitivity is ~~also very~~ ~~more~~ strongly related to parameterisation choices and initial condition choices, in addition to the release date of the model.~~ than the release date of the model.~~

Across the ensemble there is a~~no clear~~ significant correlation between sensitivity and model resolution (right panels of s~~S~~upplementary Fig. ~~S~~8), with a larger temperature anomaly and precipitation anomaly predicted in higher resolution models ($p < 0.05$). This suggests that low resolution models may not be able to capture the full extent of climate change shown by higher resolution models. ~~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.~~ However, it is noted that these relationships are only statistical correlations and some models do not show the same pattern. For example, the CCSM4-Utr model has much greater temperature and precipitation anomalies, and CCSM4 has lower temperature and precipitation anomalies than other models of a similar 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.

Data model c~~C~~omparisons undertaken for PlioMIP1 ~~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~~.~~. In addition ~~However~~, comparing the PlioMIP2 results to an additional ~~the~~ dataset of published SSTs for the timeslice ~~in~~ (McClymont et al., this issue), we see that the first order outcome of model-data comparison is the same as that shown by the comparison to F&D19_30.

The Pliocene minus pre-industrial SST anomaly will not only depend on which SST dataset is chosen to represent the Pliocene, but also on the choice of ~~However, the choice of the observed~~ observed SST data set used for the pre-~~-~~industrial~~to create the Pliocene minus pre-industrial SST anomaly can also be important.~~. Supplementary Fig. 9 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~~several 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~~several 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, leading to warming of 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~~more 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~~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). However, we note that no seasonal bias has been identified in the modern dataset in the Benguela region (Tierney and Tingley, 2018).

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 anomalies in the tropics~~y~~ and specific~~the~~ gridbox based temperature anomalies over large portions of the globe~~y. 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~~. Across the ensemble, the tropical Pliocene temperature anomaly is more strongly related to ~~a models~~ ECS than other latitudes, both as a latitudinal mean and also when considering individual gridpoints. 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~~ modelled~~s~~ ECS, according to:

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

Where α and C are constants, and 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 proxy data is more sparsely distributed than PlioMIP1 proxy 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 (F~~f~~igure 7c). ~~Hence~~Hence, we apply equation (2) with~~ll~~ ~~now use~~ ΔSST from~~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 F~~f~~ig. 9,~~,~~ and have a range of 2.6°C - 4.8°C with a mean ECS of 3.6°C and a standard deviation of 0.6°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~~several 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 local $Plio_{Core} - PI_{Crtl}$ and a model's ECS was significant at the 95% confidence level ($p < 0.05$; these regions are hatched in F~~f~~ig.~~ure~~ 7d). Secondly, we required that at least one of the models in the PlioMIP2 ensemble was within 1°C of the data~~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 two~~one~~ site from Benguela – despite these sites showing a theoretical relationship between $Plio_{Core} - PI_{Crtl}$ and ECS. Altogether~~Although only~~ 13 datapoints fulfilled both these conditions and could be used to estimate ECS.~~,~~ T~~t~~he range of estimates of ECS from PlioMIP2 (2.6°C-4.8°C) are similar to ~~in good agreement with~~ IPCC (1.5°C – 4.5°C) but are ~~and are similar to (but~~ slightly larger than was estimated from~~)~~ PlioMIP1 (1.9°C -3.7°C). It is not currently possible to add reliable error bars to the range of ECS estimates from PlioMIP2. However, as the T~~t~~ier1 PlioMIP2 experiments with CO_2 set to 350ppmv and 450ppmv become available we will be able to provide an indication as to how uncertainties in the KM5c CO_2 would affect the PlioMIP2/PRISM4 constrained estimates of ECS. In addition, as more orbitally tuned SST data becomes available, it will be necessary to revisit the ECS analysis in order to ensure maximum accuracy.

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 with enhanced~~under a contemporary doubling of~~ atmospheric CO_2 concentration. T~~Whilst t~~his may appear to be a reasonable assumption~~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 ~~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 2.0, 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 but slightly larger (1.1 to 2.9 and 1.7 respectively)~~.~~. Therefore, new modelling and new constraints on the data for PlioMIP2 suggests a slight increase in estimates of both the ESS/CS ratio and data constrained estimates of ECS between PlioMIP1 and PlioMIP2.

6. Conclusions

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

- Global annual mean surface air temperatures increases by 1.7 to 5.2°C compared to the pre-industrial, with a multi-model average increase of 3.2°C.
- The multi-model mean annual total precipitation rate increases by 7% compared to the pre-industrial, while~~ith~~ the modelled range of precipitation increases by between 2% and 13%.
- The ~~predicted~~ multi-model mean anomaly between Pliocene and pre-industrial is statistically robust for surface air temperature and sea surface temperature over most of the globe~~.~~. The multi-model mean~~modelled~~ precipitation anomaly

is robust at mid-high latitudes and in monsoon regions but is smaller than inter-model standard deviation 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 NH warming does not show a clear seasonal cycle, but a clear seasonal cycle is seen in many individual models. ~~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, ~~T~~t~~he difference in the average warming between the hemispheres is subdued, relative to simulations of 2100 CE climate. 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 67%.
- 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 2.6°C-4.8°C. This is larger than the values suggested from PlioMIP1 (1.9°C -3.7°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 large 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, grant #. 024.002.001). Simulations with CCSM4-Utr~~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. –XL acknowledges financial support from China Scholarship Council (201804910023) and the China Postdoctoral Science Foundation funded project (2015M581154). 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., and Scipal, K.: 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, https://doi.org/10.1175/2008jhm1068.1, 2009.

Balsamo G., Pappenberger F., Dutra E., Viterbo P., and van der Hurk B.: 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, https://doi.org/10.1002/hyp.7808, 2011.

Bitz, C.M., Shell, K.M., Gent, P.R., Bailey, D.A., Danabasoglu, G., Armour, K.C., Holland, M.M., and Kiehl, J.T.: Climate Sensitivity of the Community Climate System Model, Version 4., Journal of Climate, 25(9), 3053-3070, https://doi.org/10.1175/JCLI-D-11-00290.1, 2012.

Burls, N.J., and Fedorov, A.V.: Simulating Pliocene warmth and a permanent El Nino-like state: The role of cloud albedo, Paleoceanography, 29(10), 893-910, https://doi.org/10.1002/2014pa002644, 2014.

Byrne, M.P., and O'Gorman, P.A.: Land-Ocean Warming Contrast over a Wide Range of Climates: Convective Quasi-Equilibrium Theory and Idealized Simulations, Journal of Climate, 26(12), 4000-4016, https://doi.org/10.1175/jcli-d-12-00262.1, 2013.

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

Chan, W.-L. and Abe-Ouchi, A.: PlioMIP2 simulations using the MIROC4m climate model, Climate of the Past, in prep, 2019.

Chandan, D. and Peltier, W. R.: 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, 2017.

Chandan, D. and Peltier, W. R.: 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, 2018.

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

Corvec, S., and Fletcher, C.G.: Changes to the tropical circulation in the mid-Pliocene and their implications for future climate, Climate of the Past, 13(2), 135-147, https://doi.org/10.5194/cp-13-135-2017, 2017.

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

Craig, A.P., Vertenstein, M., and Jacob, R.: A new flexible coupler for earth system modeling developed for CCSM4 and CESM1, International Journal Of High Performance Computing Applications, 26(1), 31-42, https://doi.org/10.1177/1094342011428141, 2012.

Danabasoglu, G., Bates, S.C., Briegleb, B.P., Jayne, S.R., Jochum, M., Large, W.G., Peacock, S., and Yeager, S.G.: The CCSM4 Ocean Component, Journal of Climate, 25(5), 1361-1389, https://doi.org/10.1175/JCLI-D-11-00091.1, 2012.

Danabasoglu, G., Lamarque, J.F., Bacmeister, J., Bailey, D.A., DuVivier, A.K., Edwards, J., Emmons, L.K., Fasullo, J., Garcia, R., Gettelman, A. and Hannay, C.: The community earth system model version 2 (CESM2). Journal of Advances in Modeling Earth Systems, 12(2), p.e2019MS001916. 2020.

De La Vega, E., Chalk, T.B., Foster, G.L., Bysani, R., and Wilson, ~~P.A. :~~P.A.: 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. 2018.

de Nooijer, W., Zhang, Q., Li, Q., Zhang, Q., Li, X., Zhang, Z., Guo, C., Nisancioglu, K. H., Haywood, A. M., Tindall, J. C., Hunter, S. J., Dowsett, H. J., Stepanek, C., Lohmann, G., Otto-Bliesner, B. L., Feng, R., Sohl, L. E., Tan, N., Contoux, C., Ramstein, G., Baatsen, M. L. J., von der Heydt, A. S., Chandan, D., Peltier, W. R., Abe-Ouchi, A., Chan, W.-L., Kamae, Y., and Brierley, C. M.: Evaluation of Arctic warming in mid-Pliocene climate simulations, Clim. Past Discuss., https://doi.org/10.5194/cp-2020-64, in review, 2020.

De Schepper, S., Schreck, M., Beck, K.M., Matthiessen, J., Fahl, K., and Mangerud, G.: Early Pliocene onset of modern Nordic Seas circulation related to ocean gateway changes, Nature Communications, 6, Article Number 8659, https://doi.org/10.1038/ncomms9659, 2015.

Doescher R. and the EC-Earth Consortium: The EC-Earth3 Earth System Model for the Climate Model Intercomparison Project 6. Manuscript in preparation, 2020.

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

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

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., and Riesselman, C.R.: Assessing confidence in Pliocene sea surface temperatures to evaluate predictive models. Nature Climate Change, 2, (5), 365-371, https://doi.org/10.1038/NCLIMATE1455, 2012.

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

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., and Haywood, A.M.: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. Climate of the Past, 12, (7), 1519-1538, https://doi.org/10.5194/cp-12-1519-2016, 2016.

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

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.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123-2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Feng, R., Otto-Bliesner, B.L., Xu, Y.Y., Brad, E., Fletcher, T., and Ballantyne, A.: Contributions of aerosol-cloud interactions to mid-Piacenzian seasonally sea ice-free Arctic Ocean, Geophysical Research Letters, 46(16), 9920-9929, https://doi.org/10.1029/2019GL083960, 2019.

Feng, R., Otto-Bliesner, B.L, Brady, E.C. and Rosenbloom, N.A. Increasing Earth System Sensitivity in mid-Pliocene simulations from CCSM4 to CESM2. (in review at Journal of Advances in Modeling Earth Systems 2020)

Fedorov, A.V., Brierley, C.M., and Emanuel, K.: Tropical cyclones and permanent El Nino in the early Pliocene epoch. Nature, 463 (7284), 1866-U84, https://doi.org/10.1038/nature08831, 2010.

Fedorov, A.V., Brierley, C.M., Lawrence, K.T., Liu, Z., Dekens, P.S., and Ravelo, A.C.: Patterns and mechanisms of early Pliocene warmth, Nature, 496, (7443), 43-49, https://doi.org/10.1038/nature12003, 2013.

Fichefet, T., and Maqueda, M.A.M.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, Journal of Geophysical Research-Oceans, 102(C6), 12609-12646, https://doi.org/10.1029/97JC00480, 1997.

Fichefet, T., and Maqueda, M.A.M.: 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, https://doi.org/10.1007/s003820050280, 1999.

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

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.: The community climate system model version 4, Journal of Climate, 24(19), 4973-4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Gettelman, A., Hannay, C., Bacmeister, J.T., Neale, R.B., Pendergrass, A.G., Danabasoglu, G., Lamarque, J.F., Fasullo, J.T., Bailey, D.A., Lawrence, D.M. and Mills, M.J.: High climate sensitivity in the Community Earth System Model Version 2 (CESM2). Geophysical Research Letters, 46(14), .8329-8337. 2019

Gettelman, A., Kay, J.E., and Shell, K.M.: The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model, Journal of Climate, 25(5), 1453-1469, https://doi.org/10.1175/JCLI-D-11-00197.1, 2012.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., and Wood, R.A.: 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, https://doi.org/10.1007/s003820050010, 2000.

Guo, C.C. Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Ottera, O.H.: Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM), Geoscientific Model Development, 12(1), 343-362, https://doi.org/10.5194/gmd-12-343-2019, 2019.

Hagemann, S., and Gates, L.D.: 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, https://doi.org/10.1007/s00382-003-0349-x, 2003.

Hagemann, S., and Dumenil, L.: A parametrization of the lateral waterflow for the global scale, Climate Dynamics, 14(1), 17-31, https://doi.org/10.1007/s003820050205, 1998.

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

Haywood, A.M., Valdes, P.J., and Sellwood, B.W.: Global scale palaeoclimate reconstruction of the middle Pliocene climate using the UKMO GCM: initial results., Global and Planetary Change, 25, (3-4), 239-256, https://doi.org/10.1016/S0921-8181(00)00028-X, 2000.

Haywood, A.M., and Valdes, P.J.: Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere, Earth and Planetary Science Letters, 218, (3-4), 363-377, https://doi.org/10.1016/S0012-821X(03)00685-X, 2004.

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., and Sohl, L.E.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1), Geosci. Mod. Dev., 3, (1), 227-242, https://doi.org/10.5194/gmd-3-227-2010, 2010.

Haywood, A.M., Dowsett, H.J., Robinson, M.M., Stoll, D.K., Dolan, A.M., Lunt, D.J., Otto-Bliesner, B., and Chandler, M.A.: Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 2), Geosci. Mod. Dev., 4, 571-577, https://doi.org/10.5194/gmd-4-571-2011, 2011.

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., and Zhang, Z.: Large-scale features of Pliocene climate: Results from the Pliocene Model Intercomparison Project., Climate of the Past, 9, (1), 191-209, https://doi.org/10.5194/cp-9-191-2013, 2013a.

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., and Valdes, P.J.: On the identification of a Pliocene time slice for data-model comparison, Philos Trans A Math Phys Eng Sci., 371, https://doi.org/10.1098/rsta.2012.0515, 2013b.

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., and Salzmann, U.: The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: Scientific Objectives and Experimental Design, Climate of the Past, 12, (3), 663-675, https://doi.org/10.5194/cp-12-663-2016, 2016.

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., and van der Wiel., K.: EC-Earth V2.2: description and validation of a new seamless earth system prediction model, Climate Dynamics, 39(11), 2611-2629, https://doi.org/10.1007/s00382-011-1228-5, 2012.

Held, I.M., and Soden, B.J.: Robust responses of the hydrological cycle to global warming, Journal of Climate, 19, (21), 5686-5699, https://doi.org/10.1175/JCLI3990.1, 2006.

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., and Ueda, H.: Evaluating the dominant components of warming in Pliocene climate simulations, Climate of the Past, 10, (1), 79-90, https://doi.org/10.5194/cp-10-79-2014, 2014.

Hill, D.J.: The non-analogue nature of Pliocene temperature gradients, Earth and Planetary Science Letters, 425, 232-241, https://doi.org/10.1016/j.epsl.2015.05.044, 2015.

Hind, A., Zhang, Q., and Brattstrom, G.: Problems encountered when defining Arctic amplification as a ratio, Scientific Reports, 6, article 30469, https://doi.org/10.1038/srep30469, 2016.

Holland, M.M., Bailey, D.A., Briegleb, B.P., Light, B., and Hunke, E.: 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, https://doi.org/10.1175/JCLI-D-11-00078.1, 2012.

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., and Lott, F.: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, Climate Dynamics, 27(7-8), 787-813, https://doi.org/10.1007/s00382-006-0158-0, 2006.

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., and Roehrig, R.: LMDZ5B: the atmospheric component of the IPSL climate model with revisited parameterizations for clouds and convection, Climate Dynamics, 40(9-10), 2193-2222, https://doi.org/10.1007/s00382-012-1343-y, 2013.

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., and Zhang, Z.: Arctic sea ice simulation in the PlioMIP ensemble, Climate of the Past, 12 (3), 749-767, https://doi.org/10.5194/cp-12-749-2016, 2016.

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, and Zhang, H.M.: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons, Journal of Climate, 30, (20), 8179-8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Hunke, E.C., Lipscomb, W.H., and Turner, A.K.: Sea-ice models for climate study: retrospective and new directions, Journal of Glaciology, 56 (200), 1162-1172, https://doi.org/10.3189/002214311796406095, 2010.

Hunke, E.C., Lipscomb, W.H., Turner, A.K., Jeffery, N. and Elliott, S.: CICE: The Los Alamos Sea Ice Model. Documentation and Software User's Manual. Version 5.1. T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Tech. Rep. LA-CC-06-012, 2015.

Hunter, S.J., Haywood, A.M., Dolan, A.M., and Tindall, J.C.: The HadCM3 contribution to PlioMIP phase 2. Climate of the Past, 15, (5), 1691-1713, https://doi.org/10.5194/cp-15-1691-2019, 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.: The community earth system model: A framework for collaborative research, B. Am. Meteorol. Soc., 94, 1339–1360. https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

IPCC: *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.* 2013.

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

Jungclaus, J.H., Keenlyside, N., Botzet, M., Haak, H., Luo, J.J., Latif, M., Marotzke, J., Mikolajewicz, U., and Roeckner, E.: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM, Journal of Climate, 19(16), 3952-3972, https://doi.org/10.1175/JCLI3827.1, 2006.

K-1 model developers: 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. 2004.

Kamae, Y., Yoshida, K., and Ueda, H.: Sensitivity of Pliocene climate simulations in MRI-CGCM2.3 to respective boundary conditions, Climate of the Past, 12(8), 1619-1634, https://doi.org/10.5194/cp-12-1619-2016, 2016.

Krinner, G., Viovy, N., de Noblet-Ducoudre, N., Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I.C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochemical Cycles, 19, GB1015, https://doi.org/10.1029/2003GB002199, 2005.

Lambert, F.H., Webb, M.J., and Joshi, M.M.: The Relationship between Land-Ocean Surface Temperature Contrast and Radiative Forcing, Journal of Climate, 24, (13), 3239-3256, https://doi.org/10.1175/2011JCLI3893.1, 2011.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M., and Levrard, B.: A long-term numerical solution for the insolation quantities of the Earth, Astronomy and Astrophysics, 428, 261-285. https://doi.org/10.1051/0004-6361:20041335, 2004.

~~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., and Slater, A.G.: 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, https://doi.org/10.1029/2011MS000045, 2011.~~

Lawrence, D.M., Oleson, K.W., Flanner, M.G., Fletcher, C.G., Lawrence, P.J., Levis, S., Swenson, S.C., and Bonan, G.B.: The CCSM4 Land Simulation, 1850-2005: Assessment of Surface Climate and New Capabilities, Journal of Climate, 25(7), 2240-2260, https://doi.org/10.1175/JCLI-D-11-00103.1, 2012.

Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D. and Kluzek, E.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. Journal of Advances in Modeling Earth Systems. 2019.

Leduc, G., Garbe-Schonberg, D., Regenberg, M., and Contoux, C.: The late Pliocene Benguela upwelling status revisited by means of multiple temperature proxies, Geochemistry Geophysics Geosystems, 15(2), 475-491, https://doi.org/10.1002/2013GC004940, 2014.

Li, X.Y., Jiang, D.B., Zhang, Z.S., Zhang, R., Tian, Z.P., and Yan, Q.: Mid-Pliocene westerlies from PlioMIP simulations. Adv. Atmos. Sci., 32, (7), 909-923, https://doi.org/10.1007/s00376-014-4171-7, 2015.

Li, X.Y., Jiang, D.B., Tian, Z.P., Yang, Y.B.: Mid-Pliocene global land monsoon from PlioMIP1 simulations. Palaeogeography, Palaeoclimatology, Palaeoecology, 512, 56–70, 2018.

Li, X.Y., Guo, C.C., Zhang, Z.S., Otterå O.H., Zhang, R.: PlioMIP2 simulation with NorESM-L and NorESM1-F, Climate of the Past, 16, 183-197, https://doi.org/10.5194/cp-16-183-2020, 2020.

Lohmann, G., Knorr, G., Hossain, A., and Stepanek, C.: Flat Miocene and Pliocene temperature gradients and reduced seasonality: sensitivity experiments with atmospheric CO2 concentrations and ocean mixing, Paleoceanography and Paleoclimatology (in review).

Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., and Dowsett, H.J.: Earth system sensitivity inferred from Pliocene modelling and data, Nature Geoscience, 3, (1), 60-64, https://doi.org/10.1038/NGEO706, 2010.

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., and Winguth, C.: A model-data comparison for a multi-model ensemble of early Eocene atmosphere-ocean simulations: EoMIP, Climate of the Past, 8, (5), 1717-1736, https://doi.org/10.5194/cp-8-1717-2012, 2012.

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

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

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

Marsland, S.J., Haak, H., Jungclaus, J.H., Latif, M., and Roske, F.: The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates, Ocean Modelling, 5(2), 91-127, https://doi.org/10.1016/S1463-5003(02)00015-X, 2003.

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., and Talandier, C.: Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution, Climate Dynamics, 34(1), 1-26, https://doi.org/10.1007/s00382-009-0640-6, 2010.

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., and Dosio, A.: 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, https://doi.org/10.1088/1748-9326/aab048, 2018.

Mellor, G.L., and Kantha, L.: An Ice-Ocean Coupled Model, Journal of Geophysical Research-Oceans, 94(C8), 10937-10954, https://doi.org/10.1029/JC094iC08p10937, 1989.

Müller, P.J., Kirst, G., Ruhland, G., von Storch, I., and Rosell-Mele, A.: Calibration of the alkenone paleotemperature index U-37(K ') based on core-tops from the eastern South Atlantic and the global ocean (60 degrees N-60 degrees S), Geochimica Et Cosmochimica Acta, 62(10), 1757-1772, https://doi.org/10.1016/S0016-7037(98)00097-0, 1998.

Neale, R.B., Richter, J., Park, S., Lauritzen, P.H., Vavrus, S.J., Rasch, P.J., and Zhang, M.H.: The Mean Climate of the Community Atmosphere Model (CAM4) in Forced SST and Fully Coupled Experiments, Journal of Climate, 26(14), 5150-5168, https://doi.org/10.1175/JCLI-D-12-00236.1, 2013.

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.: Description of the NCAR community atmosphere model (CAM 5.0). *NCAR Tech. Note NCAR/TN-486+ STR*, *1*(1), pp.1-12. 2010b.

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.: Description of the NCAR Community Atmosphere Model (CAM 4.0). 2010a.

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, and Somot S.: The effects of 1.5 and 2 degrees of global warming on Africa in the CORDEX ensemble, Environmental Research Letters, 13(6), 065003, https://doi.org/10.1088/1748-9326/aab1b1, 2018.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., … Zeng, X. Technical Description of version 4.0 of the Community Land Model (CLM) (No. NCAR/TN-478+STR). University Corporation for Atmospheric Research. http://dx.doi.org/10.5065/D6FB50WZ, 2010.

~~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., and Qian, T.: Improvements to the Community Land Model and their impact on the hydrological cycle, Journal of Geophysical Research-Biogeosciences, 113(G1), G01021, https://doi.org/10.1029/2007JG000563, 2008.~~

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

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

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

Pound, M.J., Tindall, J., Pickering, S.J., Haywood, A.M., Dowsett, H.J., and Salzmann, U.: 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, https://doi.org/10.5194/cp-10-167-2014, 2014.

Prescott, C.L., Haywood, A.M., Dolan, A.M., Hunter, S.J., Pope, J.O., and Pickering, S.J.: Assessing orbitally-forced interglacial climate variability during the mid-Pliocene Warm Period, Earth and Planetary Science Letters, 400, 261-271, https://doi.org/10.1016/j.epsl.2014.05.030, 2014.

Raddatz, T.J., Reick, C.H.,Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.G., Wetzel, P. and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, Climate Dynamics, 29(6), 565-574, https://doi.org/10.1007/s00382-007-0247-8, 2007.

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

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.: The atmospheric general circulation model ECHAM5. PART I: Model description, Report 349, Max-Planck-Institut für Meteorologie, Hamburg. 2003.

Rosell-Mele, A., and Prahl, F.G.: Seasonality of U-37(K)' temperature estimates as inferred from sediment trap data, Quaternary Science Reviews, 72, 128-136, https://doi.org/10.1016/j.quascirev.2013.04.017, 2013.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthelemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F.: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities, Geoscientific Model Development, 8(10), 2991-3005, https://doi.org/10.5194/gmd-8-2991-2015, 2015.

Russell, J.L., Dixon, K.W., Gnanadesikan, A., Stouffer, R.J., and Toggweiler, J.R.: The Southern Hemisphere Westerlies in a Warming World: Propping Open the Door to the Deep Ocean, J. Climate, 19, (24), 6382-6390, https://doi.org/10.1175/JCLI3984.1, 2006.

Sagoo, N., and Storelvmo, T.: Testing the sensitivity of past climates to the indirect effects of dust, Geophysical Research Letters, 44(11), 5807-5817, https://doi.org/10.1002/2017GL072584, 2017.

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., and Zhang, Z.: Challenges in reconstructing terrestrial warming of the Pliocene revealed by data-model discord, Nature Climate Change, 3, 969-974, https://doi.org/10.1038/NCLIMATE2008, 2013.

Salzmann, U., Haywood, A.M., Lunt, D.J., Valdes, P.J., and Hill, D.J.: A new global biome reconstruction and data-model comparison for the Middle Pliocene, Global Ecol. Biogeogr., 17, (3), 432-447, https://doi.org/10.1111/j.1466-8238.2008.00381.x, 2008.

Sato, N., Sellers, P. J., Randall, D. A., Schneider, E. K., Shukla, J., Kinter, J. L., Hou, Y.-Y., and Albertazzi, E.: Effects of implementing the simple biosphere model in a general circulation model, J. Atmos. Sci., 46, 2757–2782, https://doi.org/10.1175/1520-0469(1989)046<2757:EOITSB>2.0.CO;2, 1989.

Sellers, P. J., Mintz, Y., Sud, Y. C., and Dalcher, A.: A simple biosphere model (SiB) for use within general circulation models, J. Atmos. Sci., 43, 505–531, https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2, 1986.

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

Smith, R., Jones, P., Briegleb, B.; et al.: 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 2010.

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, https://doi.org/10.5194/gmd-12-1139-2019, 2019.

Stepanek, C., Samakinwa, E., and Lohmann, G.: Contribution of the coupled atmosphere–ocean–sea ice–vegetation model COSMOS to the PlioMIP2, Clim. Past Discuss., https://doi.org/10.5194/cp-2020-10, in review, 2020.

Stouffer, R.J., Manabe, S. and Bryan, K.: Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO_2, Nature, 342, 660-662, https://doi.org/10.1038/342660a0, 1989.

Sun, Y., Ramstein, G., Contoux, C., and Zhou, T.: 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. https://doi.org/10.5194/cpd-9-1449-2013, 2013.

Tan, N., Contoux, C., Ramstein, G., Sun, Y., Dumas, C., and Sepulchre, P.: 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, 2019.

Tierney, J.E., Haywood, A.M., Feng, R., Bhattacharya, T., and Otto-Bliesner, B.L.: Pliocene Warmth Consistent with Greenhouse Gas Forcing, Geophysical Research Letters, 46, (15), 9136-9144, https://doi.org/10.1029/2019GL083802, 2019.

Tierney, J.E, and Tingley, M.P.: BAYSPLINE: A New Calibration for the Alkenone Paleothermometer, Paleoceanography and Paleoclimatology, 33 (3), 281-301, https://doi.org/10.1002/2017PA003201, 2018.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., and Maqueda, M.A.M.: Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation, Ocean Modelling, 27(1-2), 33-53, https://doi.org/10.1016/j.ocemod.2008.10.005, 2009.

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., and Döscher, R.: On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-282, in review, 2019.

Yukimoto, S., Noda, A., Kitoh, A., Hosaka, M., Yoshimura, H., Uchiyama, T., Shibata, K., Arakawa, O., and Kusunoki, S.: 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, https://doi.org/10.2151/jmsj.84.333, 2006.

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., and Ueda, H.: Mid-Pliocene East Asian monsoon climate simulated in the PlioMIP, Climate of the Past, 9, (5), 2085-2099, https://doi.org/10.5194/cp-9-2085-2013, 2013.

Zhang, R., Zhang, Z.S., Jiang, D.B., Yan, Q., Zhou, X., and Cheng, Z.G.: Strengthened African summer monsoon in the mid-Piacenzian, Advances in Atmospheric Sciences, 33, (9), 1061-1070, https://doi.org/10.1007/s00376-016-5215-y, 2016.

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., and Ueda, H.: Mid-Pliocene Atlantic Meridional Overturning Circulation not unlike modern, Climate of the Past, 9, (4), 1495-1504, https://doi.org/10.5194/cp-9-1495-2013, 2013.

Zheng, J., Zhang, Q., Li, Q., Zhang, Q., and Cai, M.: Contribution of sea ice albedo and insulation effects to Arctic amplification in the EC-Earth Pliocene simulation, Clim. Past, 15, 291–305, https://doi.org/10.5194/cp-15-291-2019, 2019.

(a) **Model ID, Vintage**	**(b)** **Sponsor(s), Country**	**(c)** **Atmosphere** **Top ~~-~~Resolution and~~&~~ Model References**	**(d)** **Ocean*** **Resolution Vertical~~Z~~ Coord., Top BC, & Model References**	**(e)** **Sea Ice*** **Dynamics, Leads & Model References**	**(f)** **Coupling*** **Flux adjustments and~~&~~ Model References**	**(g)** **Land** **Soils, Plants, Routing & Model References**	**(h)** **PlioMIP~~2~~ ~~Experiment2~~ Experiment Eoi400** **(Boundary Conditions & Experiment Citation)**	**(i)** **Vegetation (Static ~~-~~ ~~Salzmann~~ Salzmann et al. 2008 or Dynamic)**	**(j)** **Climate Sensitivity (ECS) °C (incl. source)**
CCSM4 (CESM 1.0.5) 2011	National Center for Atmospheric Research	Top = 2~~3.6~~ hPa ~~at the mid-point, 2.3 hPa at the interface~~ 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. (2010~~08~~)	Enhanced Feng et al. 2020 (in review~~prep~~)	Salzmann et al. (2008)	3.2 (Bitz et al. 2012)
CCSM4_ Utrecht (CESM 1.0.5) 2011	IMAU, Utrecht University, the Netherlands	As CCSM4 except FV (2.5°x 1.9°) ~~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)~~As CCSM4 but with parameterisation changes described in section 3.1	~~CICE4~~ ~~Hunke and Lipscomb (2008)~~as CCSM4	CPL7 Craig et al. (2012)	~~CLM4~~ ~~Oleson et al (2010);~~ ~~Lawrence et al. (2011)~~as CCSM4	Enhanced ~~TBA?~~	Salzmann et al. (2008)	3.2 (Baatsen et al, in prep) ~~(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)~~As CCSM4	As CCSM4 but with parameterisation changes described in section 3.1~~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)~~as CCSM4	~~No adjustment Craig et al., (2012)~~As CCSM4	~~Layers, canopy, routing Lawrence et al. (2011)~~as CCSM4	Enhanced Chandan and Peltier (2017, 2018)	Salzmann et al. (2008)	3.2 (Chandan and Peltier, 2018) ~~(Bitz et al. 2012)~~

(a) Model ID, Vintage	(b) Sponsor(s), Country	(c) Atmosphere Top ~~-~~Resolution and~~&~~ Model References	(d) Ocean* Resolution Vertical~~Z~~ Coord., Top BC, & Model References	(e) Sea Ice* Dynamics, Leads & Model References	(f) Coupling* Flux adjustments and~~&~~ Model References	(g) Land Soils, Plants, Routing & Model References	(h) PlioMIP~~2~~ ~~Experiment2~~ Experiment Eoi400 (Boundary Conditions & Experiment Citation)	(i) Vegetation (Static ~~-~~ ~~Salzmann~~ Salzmann et al. 2008 or Dynamic)	(j) Climate Sensitivity (ECS) °C (incl. source)
CESM1.2 2013	National Center for Atmospheric Research	Top = ~~3.6 hPa at the mid-point,~~ 2~~.3~~ hPa ~~at the interface~~ 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)~~as CCSM4	No adjustments Hurrell et al. (2013)	~~Layers, prescribed vegetation type with prognostic phenology, carbon cycle, routing Oleson et al. (2008)~~as CCSM4	Enhanced Feng et al. 2020 (in review)	Salzmann et al. (2008)	4.1 (Gettelman et al. 2012)
CESM2 2020	National Center for Atmospheric Research	Top = ~~3.6 hPa at the mid-point,~~ 2~~.3~~ hPa ~~at the interface~~ FV0.9x1.25 (~1°), L32 (CAM6) Danabasoglu et al. (2020) ~~(Neale et al. 2010a)~~	G16 (~1°), L60 depth, rigid lid, updated mixing scheme	Rheology, melt ponds, mushy physics (Hunke et al., 2015)	No adjustment Danabasoglu et al. (2020)	Layers, prescribed vegetation type with prognostic phenology, carbon and nitrogen cycle, routing (Lawrence et al., 2019)	Enhanced Feng et al. (2020, in review)	Salzmann et al. (2008)	5.3 ~~(~~Gettelman et al. (2019)
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 Stepanek et al. (2020)~~(Uploaded 2 x CO$_2$ minus PI experiment)~~
EC-Earth 3.3 2019	Stockholm University, Sweden	IFS cycle 36r4 Top = 5 hPa 1.125° x 1.125°, L62 Döscher et al. (2020)	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)	4.3 Wyser et al. (2020)
GISS2.1G 2019	Goddard Institute for Space Studies, USA	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)

(a) Model ID, Vintage	(b) Sponsor(s), Country	(c) Atmosphere Top -Resolution and~~&~~ Model References	(d) Ocean* Resolution Vertical~~Z~~ Coord., Top BC, & Model References	(e) Sea Ice* Dynamics, Leads & Model References	(f) Coupling* Flux adjustments and~~&~~ Model References	(g) Land Soils, Plants, Routing & Model References	(h) PlioMIP~~2~~ ~~Experiment2~~ Experiment Eoi400 (Boundary Conditions & Experiment Citation)	(i) Vegetation (Static - ~~Salzmann~~- Salzmann et al. 2008 or Dynamic)	(j) Climate Sensitivity (ECS) °C (incl. source)
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)

(a) **Model ID, Vintage**	(b) **Sponsor(s), Country**	(c) **Atmosphere Top -Resolution and~~&~~ Model References**	(d) **Ocean* Resolution Vertical~~Z~~ Coord., Top BC, & Model References**	(e) **Sea Ice* Dynamics, Leads & Model References**	(f) **Coupling* Flux adjustments and~~&~~ Model References**	(g) **Land Soils, Plants, Routing & Model References**	(h) **PlioMIP~~2~~ ~~Experiment2~~ Experiment Eoi400 (Boundary Conditions & Experiment Citation)**	(i) **Vegetation (Static - ~~Salzmann~~- Salzmann et al. 2008 or Dynamic)**	(j) **Climate Sensitivity (ECS) °C (incl. source)**
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)

(a) **Model ID, Vintage**	(b) **Sponsor(s), Country**	(c) **Atmosphere Top ~~-~~Resolution and~~&~~ Model References**	(d) **Ocean* Resolution Vertical~~Z~~ Coord., Top BC, & Model References**	(e) **Sea Ice* Dynamics, Leads & Model References**	(f) **Coupling* Flux adjustments and~~&~~ Model References**	(g) **Land Soils, Plants, Routing & Model References**	(h) **PlioMIP~~2~~ ~~Experiment~~2 Experiment Eoi400 (Boundary Conditions & Experiment Citation)**	(i) **Vegetation (Static - ~~Salzmann~~- Salzmann et al. 2008 or Dynamic)**	(j) **Climate Sensitivity (ECS) °C (incl. source)**
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 Plio_{Core} experiment (a to g), plus details of boundary conditions (h), treatment of vegetation (i)
1565 and Equilibrium Climate Sensitivity values (j) (°C).

~~**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).~~

Model Name	**ECS**	**Eoi400 SAT**	**E280 SAT**	**Eoi400-E280 SAT**	**ESS (eqn 1)**	**ESS/CS Ratio**
		~~SAT~~	~~SAT~~	~~SAT~~	~~ESS~~	~~Ratio~~
CCSM4-Utrecht~~2deg~~	3.2	18.9	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
CESM2	5.3	19.3	14.1	5.2	10.0	1.88
COSMOS	4.7	16.9	13.5	3.4	6.5	1.39
EC-Earth3.3	4.3	18.2	13.3	4.8	9.4	2.18
GISS2.1G	3.3	15.9	13.8	2.1	4.0	1.22
HadCM3	3.5	16.9	14.0	2.9	5.6	1.60
IPSLCM6A	4.8	16.0	12.6	3.4	6.5	1.36
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	**3.7**	**16.5**	**13.3**	**3.2**	**6.2**	**1.67**

Table 2: Details of the relationship between the equilibrium climate sensitivity (ECS) and the Earth System Sensitivity (ESS) for each model. MMM denotes the multi-model mean.

~~**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~~

Model name	Root mean squared error (RMSE)	Average difference between data and model	Number of sites where model and data are within 2°C	Number of sites where model and data are within 1°C	Number of sites where model and data are within 0.5°C
CESM2	3.44	-0.18	16	9	2
IPSLCM6A	3.38	1.17	24	15	8
COSMOS	3.92	1.99	20	13	4
EC-Earth3.3	3.34	-0.45	18	5	1
CESM1.2	3.44	0.94	22	13	8
IPSLCM5A	3.83	1.76	22	17	6
MIROC4m	4.05	1.95	20	12	9
IPSLCM5A2	3.99	1.96	23	17	7
HadCM3	4.51	1.96	21	13	6
GISS2.1G	4.22	2.58	19	9	3
CCSM4	4.09	2.07	21	14	5
CCSM4-Utr	3.87	0.18	19	13	6
CCSM4-UoT	3.71	1.12	21	17	9
NorESM-L	4.12	2.35	21	12	5
MRI2.3	4.78	2.13	16	10	8
NorESM1-F	4.51	2.62	18	10	5
MMM	**3.72**	**1.51**	**23**	**17**	**10**

Table 3: Statistical relationships between the proxy data ΔSST and the model ΔSST at each of the individual grid points. The average difference is calculated as Σ|(SSTA(model) – SSTA(data)| / n, where n is the number of sites.