Interactive comment on "Evaluation of Arctic warming in mid-Pliocene climate simulations" by Wesley de Nooijer et al.
Anonymous Referee #1
Received and published: 19 May 2020

Review of "Evaluation of Arctic warming in mid-Pliocene climate simulations" by de Nooijer et al.

The authors provide a good and well written summary of several aspects of the results of the latest round of Pliocene simulations. These simulations and their comparison with available geological records are important since this period is one of the few that provides an estimate of climatic changes that are to first order driven by changes in greenhouse-gas concentrations. In the manuscript there are several aspects that should be looked at more closely and some that should be discussed more clearly. Below I will detail my concerns.

Main concerns:
Impact and/or importance of values orbital parameters:
Lines 118- 130: The authors mention that for the PlioMIP2 simulations a specific time-slice was chosen in order to have values of the orbital parameters that are similar to today. The shorter orbital cycles are 20 and 40 kiloyears, meaning that an uncertainty in the estimate age of a mPWP temperature reconstruction of 10 kiloyear could already imply quite different values of the orbital parameters. I'm not an expert on that topic, but it seems to me the age constraints that are needed to make a firm statement about the orbital parameters that accompany the climate reconstructions that are used in this work, are very difficult to obtain. The authors also mention various experiments that have been done for the Pliocene to investigate the impact of different orbital parameters. Could those results be combined with the model-data comparison provided in this paper for a more extensive discussion on the topic?

The reviewer is correct that the age estimates of the reconstructions are not resolved to the temporal resolution required to state that the reconstructions represent a specific set of orbital parameters, such as the similar-to-modern parameters within the KM5c time slice. In the introduction section, we mention that the focus on the KM5c time slice was useful for SST data-model comparisons, as SST estimates could be resolved to that resolution.

This resolution is not (currently) possible for SAT estimates. We mention the uncertainties with the SAT estimates in the methods, and reiterate in the conclusions that our ability to evaluate the Arctic SAT anomalies is constrained by the limited availability and uncertainties of the reconstructions. No changes were made.

However, it is an interesting suggestion to incorporate the results of other studies to see what the magnitude of the errors due to different orbital parameters could be. Feng et al. (2017) investigated the effects of changing orbital parameters, by performing sensitivity experiments that included respectively the minimum and maximum possible insolation at 65N in July. In their conclusions they mention "Individual forcings of elevated CO2 level (by 50 ppm), high summer/annual insolation of NHL, and closed Arctic gateways may explain 1–2 °C of the terrestrial model-proxy data mismatch in the NHL." (NHL=Northern high latitudes). We added a sentence that includes these results to give an impression of the magnitude of error associated with the orbital parameters.

Line 174-176: “Feng et al. (2017) investigated the effects of different orbital configurations, as well as elevated atmospheric CO2 concentrations (+50ppm) and closed Arctic gateways in PlioMIP1 simulations, and found that they may change the outcomes of data-model comparisons in the northern high latitudes by 1-2 °C.”

Uncertainty of proxy-based climate reconstructions:
Lines 301-304: Please shortly reiterate how this maximum uncertainty range is estimated as it is quite important for the discussion that follows. Does it for instance include any discussion on the interpretation of the climate reconstructions? Any seasonal biases? From reading the referenced literature it appears that changes in for instance the growing season are considered important drivers of the temperature reconstructions, but I don’t see a discussion on this topic in the current paper. How strong is the evidence that the reconstructed temperatures reflect changes in the annual mean rather than a value that is biased towards certain seasons? To investigate the importance of this issue, many studies resort to comparing the paper temperature reconstructions with both simulated annual mean and simulated summer temperatures, has that been considered?

Changed Line 303 to reiterate how we calculate the maximum uncertainty range:

“To investigate how these uncertainties may have affected the outcomes of the data-model comparison, we calculate the minimum and maximum temperature within the uncertainty, using the uncertainties for the temperature estimates as given by Feng et al. (2017).”

In the methods we state: “Reconstructed mPWP SATs are taken from Feng et al. (2017), who updated and combined an earlier compilation made by Salzmann et al. (2013) (Table S1). Hence, the uncertainties were all indirectly derived, they were derived from compilations. It is beyond the scope of this paper to investigate these uncertainties further. For clarity, we add later in the paragraph the following sentence: “The uncertainties in the reconstructions were derived by Feng et al. (2017) and Salzmann et al. (2013) from relevant literature.”

Good point about the potential bias towards seasons. As mentioned above, we will not go into this in detail but it is worth mentioning. In the following sentence:
“Further uncertainties arise due to bioclimatic ranges of fossil assemblages, errors in pre-industrial temperatures from the observational record, and additional unquantifiable factors.” We add “potential seasonal biases”. (At the end of this paragraph we refer to Salzmann et al. (2013) for a more detailed description of the uncertainties)

While it would definitely be interesting to compare the results to summer temperatures, in the discussion we merely try to give an indication of how the magnitude of the uncertainties associated with the reconstructions may have affected the outcomes of the data-model comparison, rather than investigating the causes and validity of these uncertainties.

Robust changes in NAO and/or NAM?
Line 435: Please be more clear about whether the RCP4.5 simulations show robust changes in NAO and/or NAM. Do you have grounds to conclude that this is the case for the PlioMIP2 simulations?

Upon further inspection and thorough discussion, we decide to remove the section about the NAO/NAM. Based on comments of both Reviewer 1 and Reviewer 2. With the following reasons:

- The results for both the PlioMIP2 and the RCP4.5 simulations are not very robust. There is a low signal-to-noise ratio.
- The comparison of the PlioMIP2 and RCP4.5 simulations is significantly hindered by the different nature of the simulations: Equilibrium versus transient. As pointed out by reviewer 2.
- The comparison is further hindered by the potential strong effect orography has on Arctic variability in the mPWP simulation. Hill et al. (2011) ascribed most of the change in the NAM they observed in the mid-Pliocene simulation to changes in orography. Since the changes in orography in PlioMIP2 are non-analogous with future climate change we do not feel that this comparison is useful.

We therefore remove Section 7.3, and make appropriate changes in the abstract, introduction, the start of Section 7, and the conclusions to represent this.

Similarly on lines 436-445: Are the changes in NAM and NAO significant? So it depends on the metrics that is used to calculate these modes of variability whether or not the changes are significantly? What does that mean? And while the temperature changes in the RCP4.5 simulations are smaller, the NAO/NAM changes are larger? Please clarify.

Thank you for your comments. We refer to the comments above for our response.

Minor comments:
Lines 124-130: So how many models did actually close the Bering Strait? From figure 2 it can be concluded that not all did, but you mention that this change in experimental design improved the model-data fit so it is important to state this clearly.

Good spot. After checking, we found that all models do have a closed Bering Strait. Furthermore, a closed Bering Strait is part of both the standard and the enhanced boundary condition datasets (Haywood et al., 2016; www.clim-past.net/12/663/2016/) in PlioMIP2 and thus part of each model's simulation. Evidently, a mistake was made with the stippling. This has been updated. Stippling became redundant and hence has been removed. Description of stippling in Figure 2 caption has been removed.

Lines 270-271: How is this conclusion reached? Why is it not important to correctly simulate SAT anomalies for the SIE anomalies?

Indeed, this conclusion cannot be reached from this data alone. It has been removed.

Lines 334-357: Of course the authors realize that having only three data points in the whole Arctic Ocean doesn't make for a particularly strong model-data comparison, but we have to work with what we have. Nonetheless, the text should clearly reflect this. The site in the Iceland Sea appears to be very close to the boundary between the regions that are never covered by sea ice and those that are covered at least one month a year. One cannot expect a course resolution climate model to put this boundary at the exact right location and thus no strong conclusions can be attached to a model-data comparison at such a site.

Agreed, three datapoints do not make for a strong data-model comparison. At the start of the sea ice data-model comparison we added: "The limited availability of proxy evidence (three reconstructions) severely limits our ability to evaluate the simulation of mPWP sea ice in PlioMIP2 simulations. Nevertheless, a data-model comparison is still worthwhile, as the few reconstructions that are available may form an interesting out-of-sample test for the simulation of sea ice in the PlioMIP2 models."

Additionally, the reviewer is correct about that the coarse resolution of the climate models and the location(s) of sea ice proxies on the maximum monthly sea ice extent boundary.

In the sentence "The majority of the models simulate a maximum SIE that extends, **or nearly extends**, into the Fram Strait and Iceland Sea Figure 10b) in at least one month (in winter) per year (Fig. 10b)," the part ", or nearly extends," was included in the paper to allow for some room for error spatially. No change was made here.

The following paragraph describes the models that match the proxy evidence completely, and does not allow for this room for error. Many models nearly match the reconstructions, and others just barely match them, and changing the definition for sea ice from a SIC of 15% to, for example, 10% would already give substantially different results. This indicates that it is too arbitrary to conclude whether a model completely agrees or completely disagrees with a specific reconstruction. Hence, the paragraph was removed.

Lines 382-390: the authors should more clearly state what the differences are between the paleo and future simulations. Both are forced with greenhouse-gas concentration changes, but the paleo runs are further forced by changes in the icesheets, vegetation, gate-ways? As for the changes in ice-sheets, vegetation and also the AMOC, one could argue that these simulation give a true long-term equilibrium response to greenhouse-gas changes. This is not the case for the impact of changing the Arctic gate-ways. Is there a way to quantify the impact of the latter as to make the comparison with future simulations more meaningful?

All major differences between the future and mid-Pliocene simulations are listed in lines 376-382.

It is an interesting idea to try to isolate the effects of orography, under the assumption that future climate will look similar to the mid-Pliocene in terms of CO2, ice sheets, and vegetation. Several papers have isolated the effects of the implementation of mid-Pliocene orography in their PlioMIP2 simulations and we have added these results to this paragraph.

Changed the paragraph to:

"Using PlioMIP2 simulations for potential lessons about future warming may be improved by isolating the effects of the changes in orograph. Similar changes in ice sheets and vegetation may occur in future equilibrium warm climates, but the changes in orography are definitively non-analogous to future warming. Several groups isolated the effects of the changed orography on global warming in PlioMIP2 simulations and found that it contributes, respectively, around 23% (IPSL6-CM6A-LR; Tan et al., 2020), 27% (COSMOS; Stepanek et al., 2020), and 41% (CCSM4-UoT; Chendan and Peltier, 2018) to the annual mean global warming in the mPWP simulations. Furthermore, this warming was strongest in the high latitudes (Chandan and Peltier, 2018; Tan et al., 2020) indicating that the additional Arctic warming in PlioMIP2 simulations, as compared to future climate simulations, are likely partially caused by changes in orography that are non-analogous with the modern-day orography. These findings highlight the caution that has to be taken when using palaeoclimate simulations as analogues for future climate change."

Lines 391-398: There are a number of studies discussing simulations of the impact of closing the Bering Strait on the AMOC strength, do they also show a moderate to strong increase in AMOC strength?

These studies did not fully implement the PlioMIP2 boundary conditions, and not all of them closed both the Bering Strait and the Canadian Archipelago Seaway. Otto-Bliesner et al. (2017) closed both Arctic Ocean gateways and found an increase of 4.5 Sv in the AMOC (~18% increase). We do not include this result in the paper as it does not implement the other PlioMIP2 boundary conditions, which may influence the magnitude of change. We do mention the papers, as the direction of change (increase in AMOC strength) corresponds. No changes made.

Line 430: what is meant with an 'active NAO strength'? It appears that the models do not provide robust support of a change in NAO amplitude.

Indeed this is not clear. We refer to our earlier response to comments about this section.

Line 433: Why are RCP6.0 and RCP8.5 simulations not used in the comparison if those provide a better comparison in terms of temperature changes?

Good point, ideally we would compare the simulations to RCP6.0 and RCP8.5, because they are more similar in terms of temperature change, but data was only available for the RCP4.5 projections. Since this section has since been removed (see earlier comments), we do not address this comment further.

Line 455: What would such improvements in boundary conditions be? Don't the authors think that all changes in boundary conditions that are likely to have a significant effect are already included?

We do think that the most important changes in boundary conditions are incorporated, but there are still large uncertainties surrounding them. E.g. It is unclear whether the atmospheric CO2 concentration was actually 400ppm. Reducing these uncertainties could improve the simulations. We change the wording "enhanced boundary conditions" to "reducing uncertainties in boundary conditions".
Furthermore, we suggest later in the conclusions that more sensitivity experiments could be carried out to quantify the effects of these uncertainties on the simulations. No changes were made here.

Figure 11: limited data availability? The data between 60N and 67.5N is missing?

The IPCC (Masson-Delmotte et al., 2013 in this case) use 67.5-90N as their definition of the Arctic region and listed their data for this region. Changed the phrasing of "here 67.5-90 N, due to limited data availability" to "here 67.5-90 N, the definition used by Masson-Delmotte et al. (2013) and the area for which they listed data"

Technical comments:
Line 110: For me forcings are not part of model physics. Please clarify.

Indeed, this can be phrased better. Changed *"Uncertainties in model physics include unconstrained forcings and uncertainties in model parameters"* to "Uncertainties in model physics include processes that are not incorporated in the model and uncertainties in model parameters."

Line 145: missing space

Good spot, fixed.

Line 451: "11 out of 16", just for clarity.

Good suggestion, added "out of 16" for increased clarity.

Table 1: It would be good to add to this table if the model is also used in CMIP5, CMIP6 or neither of those.

All models participating in PlioMIP2 are participants of CMIP6. The pre-industrial simulation is the piControl simulation of the CMIP6 DECK experiments, and PlioMIP2 is part of PMIP4 which is one of the projects of CMIP6. As CMIP6 models are generally different versions of their equivalent CMIP5 counterparts, we do not add information about the CMIP5 models, as this is not relevant for the current paper.

Caption figure 3: shouldn't that be "compared to the annual mean in a given month"?

The figure depicts the ratio between the warming in a given month respective to the annual mean, for each model individually. Adjusted the caption to: "Ratio between the mean Arctic (a) SAT and (b) SST warming in a given month and the annual mean Arctic warming, for each model (and MMM) individually"

Figure 6: what does the 'p' stand for?

Added "Depicted for both correlations are the correlation coefficient (R), the slope and the probability value (p) that when the variables are not related, a statistical result equal to or greater than observed would occur."

Figure 11: What is shown for the RCP simulations, an average over year xx to yy?

Added (2081-2100 average) to the figure caption, and "end-of-century (2081-2100) average" to the text preceding the figure.

Interactive comment on "Evaluation of Arctic warming in mid-Pliocene climate simulations" by Wesley de Nooijer et al.

Anonymous Referee #2
Received and published: 7 June 2020 Review of manuscript "Evaluation of Arctic warming in mid-Pliocene climate simulations " - de Nooijer et al.

In the present manuscript, de Nooijer et al. present an analysis of Arctic climate as simulated by the coupled models ensemble from the PLIOMIP2 initiative. PLIOMIP2 focuses on the specific KM5c interval, peak of the mPWP. Notable improvements have also being done for the boundary conditions (e.g. closed Arctic gateways during this period). Models generally simulate an Arctic amplification larger than 2.5, increase in SAT and SST. Comparison with the few existing proxies suggest that only few models of the ensemble are able to fit the warm climatic conditions of the particular KM5c interval. However, the lack of proxies prevent a more detailed comparison. An attempt is made to compare those new results to projections. Conclusion of the authors is that using the simulated mPWP KM5c is not yet informative for the future, given the current state of models and limitations of the design of the experiments and lack of proxies to validate the paleo-simulations. In general, what this phase 2 of the PLIOMIP initiative shows is that boundary conditions improvements and focus on a specific interval of the mPWP generally increase the agreement with the few existing proxies. However, the paper remains rather very elusive and not detailed too much about the causes of the simulated anomalies. In addition, there is a distinct dichotomy within the models with only few models increasing the MMM. An aspect that is really unclear throughout this manuscript is the impact of the models that do not used closed gateway in their simulations and how much this impact on the interpretation of the entire metrics presented here. In addition to closed gateways, individual model resolution might also have an impact on the representation of those gateways and this is not discuss here. The attempt made to compare with CMIP5 projections is to my opinion unsuccessful given the striking difference in gateways between the modern geography and that of the mPWP. In addition, the authors attempt to compare the mode of variability which is a non-sens here since the paleoclimatic simulations are equilibrium simulations while projections are transient short-term simulations. Authors warn about the lack of "slow-feedbacks" in the projections, but the contrary is also true, the short-term variability present some limitation in the paleoclimate runs. I do not advise to remove it. However, some improvements are needed to strengthen those parts and to make them meaningful in a way or in the other. The manuscript is written quite well (though in some places that I have indicated in my comments below, some improvements in the writing is needed to clarify). My impression is that this paper remains superficial and does not provide a real analysis of the Arctic warming. There is no real analysis of the causes/consequences of this warming (i.e. albedo, seasonal cycle in temperature, snow cover, westerlies etc.). . . Even if the number of proxies is limited, the authors could deepen their analysis to compare the different models together to provide partial answers to some of the questions posed paper by the authors themselves within the different sub-section of the manuscript. They should also explore the dichotomy amongst the models visible in almost all the figures of this manuscript and the impact this dichotomy has on the MMM and thus the overall interpretation of the MMM. I therefore recommend moderate revisions.

Author response:
Thank you for your review. We have addressed each of your comments one-by-one and we feel that there were some substantial improvements following your comments. A small

response on the main summary of the reviewer, with regards to the paper remaining superficial:
This paper mainly describes results and highlights differences between models. To investigate causes of the differences, e.g. because of albedo/seasonal cycle, we would need sensitivity experiments. This is not plausible for a multi-model analysis. We agree that the paper remained somewhat superficial, but we do not think it is possible to do deep analyses without sensitivity experiments.

Comments:
Line 68: I would remove "future" and just write "as warming in the Arctic directly affects. . .". This is because this is always true, not only for future. Or perhaps just reformulate in "as it is shown that projected Arctic warming affects. . .".

Good spot. The sentence has been adjusted to the suggested reformulation.

Line 84: Would it be worth mentioning that the interest of the KM5c interval is because orbitals are similar to present? I think this is important and relevant to the comparison with projections.

Indeed, it is an important feature of the KM5c time slice that it has a similar-to-modern orbital forcing and we have added emphasis on this.

"Additionally, the KM5c time slice is characterized by a similar-to-modern orbital forcing (Haywood et al., 2013b; Prescott et al., 2014). These factors give lessons learned from the mPWP, and the KM5c time slice in particular, potential relevance for future climate change (Burke et al., 2018; Tierney et al., 2019), and this is one of the guiding principles of PlioMIP (Haywood et al., 2016)."

Line 141: correct "model resultsaere calculated" in "model results are calculated"

Good spot, fixed.

Line 196-203: I find interesting to note that most of the models simulate air and sea temperature values below the mean and that only a couple of models exhibit values much higher than the mean. It could also be worth mentioning this somewhere (though it is not a paper about individual model performances) because it also impacts on the interpretation that one does about the ensemble mean.

Indeed, good observation, a subset of the ensemble simulates much larger temperature anomalies than the rest of the ensemble. To note readers on the potential impacts this may have on the multi-model mean results we added the following:

"There is a large variation in the magnitude of the simulated Arctic SAT anomalies, with five out of sixteen models, namely CCSM4-Utrecht, CCSM4-UoT, CESM1.2, CESM2, and EC-Earth 3.3 all simulating much stronger anomalies than the rest of the ensemble. This subset of the ensemble raises the MMM substantially and this has to be taken into account when interpreting the MMM results. The MMM SAT anomaly for the PlioMIP2 ensemble excluding this subset of five models is 5.8 °C."

Additionally, we added a sentence about SST, as the same five models are seen here to raise the MMM. “Furthermore, the five models that simulated the largest Arctic SAT anomalies also simulate the largest Arctic SST anomalies.”

In section 5.1 we discuss that this subset of the ensemble generally matches the SAT proxies best. No change was made here.

Line 209: but did not you write that also the Bering Strait is closed in some of the models? We don’t see a particularly large anomaly around this area.

The Bering Strait is closed in the PlioMIP2 simulations (mentioned in line 122) as a part of both the standard and enhanced boundary conditions. Indeed, the closure of the Bering Strait did not lead to a large SAT anomaly. Upon closer inspection, the largest SAT anomalies are mostly above the Baffin Bay, rather than over the Canadian Archipelago. The first part of the paragraph has been adjusted accordingly.

“The greatest MMM SAT anomalies in the Arctic are found in the regions with reduced ice sheet extent on Greenland (Haywood et al., 2016), which generally show warming of over 10 °C and even up to 20°C. Additionally, temperature anomalies of over 10 °C are simulated around the Baffin Bay”

Line 212: and thus? What causes such an increase in the Baffin Bay? The lack of sea ice due to no arctic waters flowing through the CA? If yes, it would be good to mention.

This line was meant as a description of the results and of the figure. While it would be interesting to know the underlying mechanisms for the warming in this location, and while we do discuss a potential mechanism later in the paper (AMOC), we did not mean to describe the causes of the temperature increase in the Baffin Bay here. No changes were made.

Line 196 - 215: How does the discrepancy in land sea mask, especially in the Bering Strait, affect the interpretation of the MMM in Figure 2? I would find very informative to indicate which models closed the Bering Strait and or the Canadian archipelago in Table 1. It seems from Figure 2b that only a few models keep the Bering Strait open. Are the models with open Bering Strait the ones with highest SST and SAT values (e.g. In Fig.1)?

Sorry for the confusion, all models have a closed Bering Strait and Canadian Archipelago as part of the PlioMIP2 boundary conditions. We added “in the mPWP simulation” to line 122 to emphasize this and to avoid future confusions for other readers. The stippling in Figure 2b has been removed as it was found to be incorrect after comments from reviewer 1 and they became redundant in the updated version. Description of stippling in Figure 2 caption has been removed.

Lines 272-289: How much is the MMM-proxy comparison valid in the Canadian archipelago? I mean, in Figure 7 the proxies there are very closed to each others (while already slightly shifted for better understanding) and, how many grid points are there in in the simulations this area? Is the comparison here valid? Or not resolutiondependent? Same for Alaska?

Valid point. Given the coarse resolution of global climate models it could be impossible for simulations of SAT anomalies to match all five reconstructions in the Canadian Archipelago. We added the following:

"It has to be noted, however, that SAT anomalies are underestimated at three other sites within the Canadian Archipelago. Given the resolution of global climate models and the close proximity of the sites, it may be impossible for simulations to match all five of these SAT estimates."

Figure 8: Since the beginning, there are two distinct groups amongst the models and the MMM is shifted to higher value because of 7 models. This discrepancy between the two groups is very neat. Thus I really wonder what are the causes of such dichotomy and what is the impact on the interpretation of the MMM in the paper in general?

Indeed, good spot, there are two distinct groups amongst the models. We would argue, however, that the first group consists of the five previously discussed models (CCSM4-Utrecht, CCSM4-UoT, CESM1.2, CESM2, and EC-Earth 3.3) when looking at the median bias (rather than the extent of the box-whiskers). We already mention here that these are the models with the highest Arctic SAT anomalies. We added some emphasis on these five models and that it may be interesting to uncover why they are simulating distinctly larger anomalies than the other simulations.

"Future research into the underlying mechanisms for the increased Arctic warming in these five simulations, compared to the remaining eleven simulations in the ensemble, may form a way to uncover factors that contribute to improved data-model agreement."

Added the following to the eleventh line of the abstract;
"although the degree of underestimation varies strongly between the simulations"

Added the following at the second line of the conclusion:
"although large differences in the degree of underestimation exist between the simulations. The models that simulate the largest Arctic SAT anomalies tend to match the reconstructions better, and investigation into the mechanisms underlying the increased Arctic warming in these simulations may help uncover factors that could contribute to improved data-model agreement."

Lines 320-321: but also models should also all use the same boundary conditions. Because if some fo the models do not close some fo the straits, or if they have no sufficient resolution to capture the width of some passages etc. . .how can we interpret the misfit between data and models correctly? I mean, as it is now, it is impossible to determine wether or not in some models the different boundary conditions or different physics affect the misfit and in which proportion. I know it is very difficult to modify the land-sea mask in coupled models and in some cases it will also require more computational resources to increase spatial resolution enough to capture the different gateways properly. However, at some points, we will need to do it to further advance those types of data-model exercises.

Sorry for the confusion. All models used the same boundary conditions, quoted from Haywood et al. (2020*): "All model groups participating in PlioMIP2 were required to use standardised boundary condition data sets for the core midPliocene-eoi400 experiment".* We added a sentence in the methods section to emphasize this.

“All model groups incorporated the standardised set of boundary conditions from the PlioMIP2 experimental design in their simulations (Haywood et al., 2016).”

Figure 10: yellow and white squares are reconstructions from proxies? I guess yes. . . but this is not mentioned in the caption.

This information has been added to the figure caption. “Depicted squares represent the locations of the reconstructions and their respective colour the inferred mPWP sea ice conditions at that location.”

Figure 11: is the vertical Y scale in frame b) the same as in frame a)? In any case, please add the ticks for dSAT values on the graph for projections.

Indeed, the Y scales are the same. It is a good idea to add the ticks in b) also, for added clarity. This has been done and the figure has been updated.

Lines 377- 381: When reading those lines, it seems that only CO2 forcing matters here. But in many of your models, some gateways are closed, and as you cite Otto-Bliesner et al. (2017), this matters. . . Thus I disagree with the formulation of those sentences. Please also discuss the difference in Arctic geography and how this impact ton the comparison with the projections.

Indeed, CO2 is not the only forcing that matters. The dominant mechanism of warming in both ensembles is CO2 (for PlioMIP2 this can be found in papers from Tan et al. (2020), Chendan and Peltier (2018), and Stepanek et al. (2020)). We simply state here that this is the dominant mechanism of warming, but that there are additional mechanisms for warming in PlioMIP2. We discuss that this may be due to changes in Arctic ocean gateways or other changes in orography in the following sentence. No changes were made.

Lines 396-400: Given the different boundary conditions, I find very difficult to make a direct comparison here. In most of PLIOMIP2 models, the Arctic gateways are closed and this generates a strengthening of the AMOC. While under modern geography, the Arctic gateways are open and a weakling of the AMOC is projected. You cannot compare those two situations here directly. In general, this short paragraph is not very clear. If you state more clearly at the beginning and in Table 1 that not all models prescribed closed gateway, this would definitely improve the reader understanding of the paper.

The main point of this paragraph is to show that there are differences between the two ensembles, regardless of their cause, in AMOC strength and thus one of the mechanisms underlying Arctic warming. We mention in the previous paragraph that strengthening of the AMOC in PlioMIP2 is likely due to the closure of the Arctic ocean gateways. As the purpose of this paragraph is to highlight the difference, rather than investigate it, we did not make any change. Sorry again for the confusion that not all models have the closed gateway, they all do, and previous comments of the reviewer led us to put more emphasis on this to avoid confusion for future readers.

Line 397: “This is consistent” To what does “this” refer to?

Indeed, we could make this more clear. We changed “this is consistent” to “The strengthening of the AMOC in the PlioMIP2 ensemble is consistent” and added a space to make it a separate paragraph. We also added “compared to the future climate ensembles” in the previous sentence for improved clarity.

Subsection 7.3: To my opinion, it is very difficult to compare transient short-term projections variability with equilibrium climate variability of a few centuries (as just say line 440). Thus I find not very much straight forward and informative the conclusions from this comparison here.

Upon further inspection and thorough discussion, we decide to remove the section about the NAO/NAM. Based on comments of both Reviewer 1 and Reviewer 2. With the following reasons:

- The results for both the PlioMIP2 and the RCP4.5 simulations are not very robust. There is a low signal-to-noise ratio.
- The comparison of the PlioMIP2 and RCP4.5 simulations is significantly hindered by the different nature of the simulations: Equilibrium versus transient. As pointed out by reviewer 2.
- The comparison is further hindered by the potential strong effect orography has on Arctic variability in the mPWP simulation. Hill et al. (2011) ascribed most of the change in the NAM they observed in the mid-Pliocene simulation to changes in orography. Since the changes in orography in PlioMIP2 are non-analogous with future climate change we do not feel that this comparison is useful.

We therefore remove Section 7.3, and make appropriate changes in the abstract, introduction, the start of Section 7, and the conclusions to represent this.

Lines 427-429: This sentence is very unclear, please reformulate.

Thank you for the comment. The section in which this sentence was stated has been removed based on earlier comments.

Lines 455 - 458: You state about the discrepancies between mPWP and projections simulations: “firstly the incomplete manifestation of slow responses in transient simulations”. But not only, I would say also vice-versa: “the lack of transient variability in equilibrium climate”. Then you state “secondly the observed differences in Arctic climate features between the ensembles”: which ensembles are you referring too here? PLIOMIP1 versus PLIOMIP2 or PLIOMIP2 versus projections? If this is the second option, then I would say the entire sentence does not make sense because of course they are different, besides equilibrium versus transient, boundary conditions also differ. . .

Good point, there is a difference between comparing simulations of different climates, and different climates themselves. Indeed, both the nature (transient versus equilibrium) and boundary conditions of the ensembles differ. We focus on the differences in Arctic climate features we observe between the ensembles, and their implication for attempting to use mPWP simulations to learn about future climate change.

Changed the sentence to: “Lastly, we find differences in Arctic climate features between the PlioMIP2 ensemble and future climate ensembles, including the magnitude of Arctic amplification, changes in AMOC strength, and northern modes, which highlight that caution has to be taken when attempting to use simulations of the mPWP to learn about future climate change.”

Evaluation of Arctic warming in mid-Pliocene climate simulations

Wesley de Nooijer[1], Qiong Zhang[1], Qiang Li[1], Qiang Zhang[1], Xiangyu Li[2,3], Zhongshi Zhang[4,3,2], Chuncheng Guo[3], Kerim H. Nisancioglu[3], Alan M. Haywood[5], Julia C. Tindall[5], Stephen J. Hunter[5], Harry J. Dowsett[6], Christian Stepanek[7], Gerrit Lohmann[7], Bette L. Otto-Bliesner[8], Ran Feng[9], Linda E. Sohl[10,11], Mark A. Chandler[10,11], Ning Tan[12,13], Camille Contoux[13], Gilles Ramstein[13], Michiel L. J. Baatsen[14], Anna S. von der Heydt[14,15], Deepak Chandan[16], W. Richard Peltier[16], Ayako Abe-Ouchi[17], Wing-Le Chan[17], Youichi Kamae[18], Chris M. Brierley[19]

1. Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
2. Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
3. NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway
4. Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan, China
5. School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds, West Yorkshire, UK
6. Florence Bascom Geoscience Center, U.S. Geological Survey, Reston, VA 20192, USA
7. Alfred Wegener Institute - Helmholtz-Zentrum für Polar und Meeresforschung, Bremerhaven, Germany
8. Palaeo and Polar Climate Division, National Center for Atmospheric Research, Boulder, Colorado, USA
9. Department of Geosciences, College of Liberal Arts and Sciences, University of Connecticut, Connecticut, USA
10. Center for Climate Systems Research, Columbia University, New York, USA
11. NASA Goddard Institute for Space Studies, New York, USA
12. Key Laboratory of Cenozoic Geology and Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China
13. Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
14. Centre for Complex Systems Science, Utrecht University, Utrecht, The Netherlands
15. Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, Utrecht, The Netherlands.
16. Department of Physics, University of Toronto, Toronto, Ontario, Canada
17. Centre for Earth Surface System Dynamics (CESD), Atmosphere and Ocean Research Institute (AORI), University of Tokyo, Tokyo, Japan
18. Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
19. Department of Geography, University College London, London, UK

Correspondence to: Qiong Zhang (qiong.zhang@natgeo.su.se)

Abstract. Palaeoclimate simulations improve our understanding of the climate, inform us about the performance of climate models in a different climate scenario, and help to identify robust features of the climate system. Here, we analyse Arctic warming in an ensemble of 16 simulations of the mid-Pliocene Warm Period (mPWP), derived from the Pliocene Model Intercomparison Project Phase 2 (PlioMIP2).

The PlioMIP2 ensemble simulates Arctic (60-90° N) annual mean surface air temperature (SAT) increases of 3.7 to 11.6 °C compared to the pre-industrial, with a multi-model mean (MMM) increase of 7.2 °C. The Arctic warming amplification ratio relative to global SAT anomalies in the ensemble ranges from 1.8 to 3.1 (MMM is 2.3). Sea ice extent anomalies range from -3.0 to -10.4 x 10^6 km^2 with a MMM anomaly of -5.6 x10^6 km^2, which constitutes a decrease of 53 % compared to the pre-industrial. The majority (11 out of 16) models simulate summer sea ice-free conditions ($\leq$ 1 x 10^6 km^2) in their mPWP simulation. The ensemble tends to underestimate SAT in the Arctic when compared to available reconstructions although the degree of underestimation varies strongly between the simulations. The simulations with the highest Arctic SAT anomalies tend to match the proxy dataset in its current form better. The ensemble shows some agreement with reconstructions of sea ice, particularly with regards to seasonal sea ice. Large uncertainties limit the confidence that can be placed in the findings and the compatibility of the different proxy datasets. We show that, while reducing uncertainties in the reconstructions could decrease the SAT data-model discord substantially, further improvements are likely to be found in enhanced boundary conditions or model physics. Lastly, we compare the Arctic warming in the mPWP to projections of future Arctic warming and find that the PlioMIP2 ensemble simulates greater Arctic amplification than CMIP5 future climate simulations and an increase instead of a decrease in AMOC strength compared to pre-industrial. The results highlight the importance of slow feedbacks in equilibrium climate simulations, and that caution must be taken when using simulations of the mPWP as an analogue for future climate change.

1 Introduction

The simulation of past climates improves our understanding of the climate system, and it provides an opportunity for the evaluation of the performance of climate models beyond the range of present and recent climate variability (Braconnot et al., 2012; Harrison et al., 2014, 2015; Masson-Delmotte et al., 2013; Schmidt et al., 2014). Comparisons of palaeoclimate simulations and palaeoenvironmental reconstructions have been carried out for several decades (Braconnot et al., 2007; Joussaume and Taylor, 1995) and show that while climate models can reproduce the direction and large-scale patterns of changes in climate, they tend to underestimate the magnitude of specific changes in regional climates (Braconnot et al., 2012; Harrison et al., 2015). The comparison of palaeoclimate simulations with future projections has aided in the identification of robust features of the climate system which can help constrain future projections (Harrison et al., 2015; Schmidt et al., 2014), including in the Arctic (Yoshimori and Suzuki, 2019).

One such robust feature is the Arctic amplification of global temperature anomalies (Serreze and Barry, 2011). Increased warming in the Arctic region compared to the global average is a common feature of both palaeo- and future climate simulations and is also present in the observational record (Collins et al., 2013; Masson-Delmotte et al., 2013). Arctic warming has a distinct seasonal character, with the largest sea surface temperature (SST) and the smallest surface air temperature (SAT) anomalies occurring in the summer due to enhanced ocean heat uptake following sea ice melt (Serreze et al., 2009; Zheng et al., 2019). It is critical to correctly simulate Arctic amplification as it is shown that projected Arctic warming affects ice sheet stability, global sea-level rise and carbon cycle feedbacks (e.g. through permafrost melting; Masson-Delmotte et al., 2013).

Deleted: future warming in the

Deleted: directly

Several multi-model analyses that included palaeoclimate simulations and/or future projections found that changes in northern high-latitude temperatures scale (roughly) linearly with changes in global temperatures (Bracegirdle and Stephenson, 2013; Harrison et al., 2015; Izumi et al., 2013; Masson-Delmotte et al., 2006; Miller et al., 2010; Schmidt et al., 2014; Winton, 2008).

Underestimation of Arctic SAT has been reported for several climates in the Palaeoclimate Modelling Intercomparison Project Phase 3 (PMIP3), including the mid-Pliocene Warm Period (Dowsett et al., 2012; Haywood et al., 2013a; Salzmann et al., 2013), Last Interglacial (LIG: Bakker et al., 2012; Lunt et al., 2013; Otto-Bliesner et al., 2013) and Eocene (Lunt et al., 2012a). PMIP4 simulations, however, of the LIG showed good agreement with SAT reconstructions in the Canadian Arctic, Greenland, and Scandinavia, while showing overestimations in other regions (Otto-Bliesner et al., 2020). PMIP4 simulations of the Eocene were also able to capture the polar amplification indicated by SAT proxies (Lunt et al., 2020).

In the present work, we analyze the simulated Arctic warming in a new ensemble of 16 simulations in the Pliocene Model Intercomparison Project Phase 2 (PlioMIP2) (Haywood et al., 2016). PlioMIP2 is designed to represent a discrete time slice within the mid-Pliocene Warm Period (mPWP; 3.264–3.025 Ma; sometimes referred to as mid-Piacenzian Warm Period): Marine Isotope Stage (MIS) KM5c, 3.204–3.207 Ma (Dowsett et al., 2016, 2013; Haywood et al., 2013b, 2016). The mPWP is the most recent period in geological history with atmospheric CO_2 concentrations similar to the present, therefore providing great potential to learn about warm climate states. Additionally, the KM5c time slice is characterised by a similar-to-modern orbital forcing (Haywood et al., 2013; Prescott et al., 2014). These factors give lessons learned from the mPWP, and the KM5c time slice in particular, potential relevance for future climate change (Burke et al., 2018; Tierney et al., 2019), and this is one of the guiding principles of PlioMIP (Haywood et al., 2016).

Deleted: This gives

Palaeoenvironmental reconstructions show that the elevated CO_2 concentrations in the mPWP coincided with substantial warming, which was particularly prominent in the Arctic (Brigham-Grette et al., 2013; Dowsett et al., 2012; Panitz et al., 2016; Salzmann et al., 2013). Haywood et al. (2020) discuss the large-scale outcomes of PlioMIP2 and observe a global warming that is between the best estimates of predicted end-of-century global temperature change under the RCP6.0 (+2.2 ± 0.5 °C) and RCP8.5 (3.7 ± 0.5 °C; Collins et al., 2013) emission scenarios.

The dominant mechanism for global warming in mid-Pliocene simulations is through changes in radiative forcing following increases in greenhouse gas concentrations (Chandan and Peltier, 2017; Hill et al., 2014; Hunter et al., 2019; Kamae et al., 2016; Lunt et al., 2012b; Stepanek et al., 2020; Tan et al., 2020). Polar warming is also dominated by changes in greenhouse gas emissivity (Hill et al., 2014; Tindall and Haywood, 2020). Apart from the changes in greenhouse gas concentrations, changes in boundary conditions that led to warming in previous simulations of the mPWP included the specified ice sheets, orography, and vegetation (Hill, 2015; Lunt et al., 2012b).

In PlioMIP1, the previous phase of this project, model simulations underestimated the strong Arctic warming that is inferred from proxy records was found (Dowsett et al., 2012; Haywood et al., 2013a; Salzmann et al., 2013). This data-model discord may have been caused by uncertainties in model physics, boundary conditions, or reconstructions (Haywood et al., 2013a)

Uncertainties in model physics include physical processes that are not incorporated in the models and uncertainties in model parameters. It was found that the inclusion of chemistry-climate feedbacks from vegetation and wildfire changes leads to substantial global warming (Unger and Yue, 2014, while excluding industrial pollutants and explicitly simulating aerosol-cloud interactions (Feng et al., 2019), and decreasing atmospheric dust loading (Sagoo and Storelvmo, 2017) leads to increased Arctic warming in mPWP simulations. Similarly, in simulations of the Eocene, two models that implemented modified aerosols had better skill than other models at representing polar amplification (Lunt et al., 2020). Changes in model parameters, such as the sea ice albedo parameter (Howell et al., 2016b), may provide further opportunities for increasing data-model agreement in the Arctic.

Deleted: unconstrained forcings and uncertainties

Several studies found changes in boundary conditions that could help resolve some of the data-model discord in the Arctic for PlioMIP1 simulations. The studied changes in boundary conditions include changes in orbital forcing (Feng et al., 2017; Prescott et al., 2014; Salzmann et al., 2013), atmospheric CO2 concentrations (Feng et al., 2017; Howell et al., 2016b; Salzmann et al., 2013), and palaeogeography and bathymetry (Brierley and Fedorov, 2016; Feng et al., 2017; Hill, 2015; Otto-Bliesner et al., 2017; Robinson et al., 2011).

New in the experimental design of PlioMIP2 are a closed Bering Strait and Canadian Archipelago in the mPWP simulation. The closure of these Arctic Ocean gateways has been shown to alter oceanic heat transport into the North Atlantic (Brierley and Fedorov, 2016; Feng et al., 2017; Otto-Bliesner et al., 2017). Additionally, the focus on a specific time slice within the mPWP allows for reduced uncertainties in reconstructions and boundary conditions, in particular with regards to orbital forcing. These changes have led to an improved data-model agreement for reconstructions of SST, particularly in the North Atlantic (Dowsett et al., 2019; Haywood et al., 2020; McClymont et al., 2020). Multi-model mean (MMM) SST anomalies in the North Atlantic deviate less than 3 °C from reconstructed temperatures (Haywood et al., 2020).

Deleted: .

In the following sections, we first evaluate the simulated Arctic (60–90° N) temperatures and sea ice extents (SIE) in the PlioMIP2 ensemble. We then perform a data-model comparison for SAT and an evaluation of how uncertainties in the reconstructions may affect the outcomes of the data-model comparison. We then compare the simulated sea ice to reconstructions. Lastly, we investigate two climatic features of the mPWP, namely Arctic amplification and the Atlantic Meridional Overturning Circulation (AMOC), and compare these analyses to findings of future climate studies to investigate the extent to which the mPWP can be used as an analogue for future Arctic climate change.

2 Methods

2.1 Participating models

The simulations of the mPWP by 16 models participating in PlioMIP2 were used in this study. The models included in this study are listed in Table 1. A more detailed description of each model's information and experiment setup can be found in Haywood et al. (2020). All model groups incorporated the standardised set of boundary conditions from the PlioMIP2 experimental design in their simulations (Haywood et al., 2016).

For each simulation, the last 100 years of data are used for the analysis. Individual model results are calculated on the native grid of each model. MMM results are obtained after regridding each model's output to a 2° x 2° grid using bilinear interpolation. Using a non-weighted ensemble mean theoretically averages out biases in models, assuming models are independent, and errors are random (Knutti et al., 2010). Climate models can, however, generally not be assumed to be independent (Knutti et al., 2010; Tebaldi and Knutti, 2007) and this is especially true for the PlioMIP2 ensemble where many models have common origins (Table 1). The MMM results will therefore likely be biased towards specific common errors within the models comprising the ensemble.

Table 1: Models participating in PlioMIP2 used in this study.

Model name	Institution	PlioMIP2 reference
CCSM4-NCAR	National Center for Atmospheric Research (NCAR)	Feng et al. (2020)
CCSM4-Utrecht	IMAU, Utrecht University	
CCSM4-UofT	University of Toronto, Canada	Chandan and Peltier (2017)
CESM1.2	NCAR	Feng et al. (2020)
CESM2	NCAR	Feng et al. (2020)
COSMOS	Alfred Wegener Institute	Samakinwa et al., (2020); Stepanek et al. (2020)
EC-Earth 3.3	Stockholm University	Zhang et al. (in Review)
GISS–E2–1–G	NASA/GISS	Kelley et al. (2020)
HadCM3	Hadley Centre for Climate Prediction and Research/Met Office UK	Hunter et al. (2019)
IPSLCM5A	Laboratoire des Sciences du Climat et de l'Environnement (LSCE)	Tan et al. (2020)
IPSLCM5A–2.1	LSCE	Tan et al. (2020)
IPSL-CM6A–LR	LSCE	Lurton et al. (2020)

Formatted: English (US)

MIROC4m	CCSR/NIES/FRCGC, Japan	Chan and Abe-Ouchi (2020)
MRI–CGCM2.3	Meteorological Research Institute	Kamae et al. (2016)
NorESM-L	NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway	Li et al. (2020)
NorESM1-F	NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway	Li et al. (2020)

2.2 Data-model comparisons

To evaluate the ability of climate models to simulate mPWP Arctic warming, we first perform a comparison to SAT estimates from palaeobotanical reconstructions. The data-model comparison is performed using temperature anomalies, calculated by differencing the mPWP and the pre-industrial simulation, to avoid overestimations of agreement due to strong latitudinal effects on temperature (Haywood and Valdes, 2004).

Reconstructed mPWP SATs are taken from Feng et al. (2017), who updated and combined an earlier compilation made by Salzmann et al. (2013) (Table S1). Qualitative estimates of confidence levels for each reconstruction were made by Feng et al. (2017) and Salzmann et al. (2013). Only reconstructions that are located at or northward of 60° N and for which the temporal range covers the KM5c time slice are included in the data-model comparison. Three reconstructions from Ballantyne et al. (2010) at the same location (78.3° N, -80.2° E) were averaged to avoid oversampling that location. The uncertainties in the reconstructions were derived by Feng et al. (2017) and Salzmann et al. (2013) from relevant literature.

Deleted: compilations made by Ballantyne et al. (2010) and

The data-model comparison will be a point-to-point comparison of modelled and reconstructed temperatures estimated from palaeobotanical proxies, which initially does not take the uncertainties of the reconstructions (Table S1) into account. The potential influence of the uncertainties in reconstructions on the outcomes of the data-model comparison will be investigated in a later section. The temporal range of the reconstructions is broad and certainly not resolved to the resolution of the KM5c time slice, unlike the dataset of SST estimates compiled by Foley and Dowsett (2019) used for PlioMIP2 SST data-model comparisons by Haywood et al. (2020) and McClymont et al. (2020). Prescott et al. (2014) found that peak warmth in the mPWP would be diachronous between different regions based on simulations with different configurations of orbital forcing. Orbital forcing is particularly important in the high latitudes and for proxies that may record seasonal signatures (e.g. due to recording growing season temperatures). As such, there may be significant biases in the dataset, as the temporal ranges of the proxies include periods with substantially different external forcing than during the KM5c time slice for which the simulations are run. Feng et al. (2017) investigated the effects of different orbital configurations, as well as elevated atmospheric CO_2

concentrations (+50ppm) and closed Arctic gateways in PlioMIP1 simulations, and found that they may change the outcomes of data-model comparisons in the northern high latitudes by 1-2 °C.

Further uncertainties arise due to bioclimatic ranges of fossil assemblages, errors in pre-industrial temperatures from the observational record, potential seasonal biases, and additional unquantifiable factors. Ultimately, the uncertainties constrain our ability to evaluate the Arctic warming in the PlioMIP2 simulations substantially. A more detailed description of the uncertainties in the SAT estimates can be found in the work of Salzmann et al. (2013).

The reconstructed temperatures are differenced with temperatures from the observational record to obtain proxy temperature anomalies. Observational record temperatures are obtained from the Berkeley Earth monthly land and ocean dataset (Rohde et al., 2013a, 2013b), and the average temperature in the 1870–1899 period was used.

Furthermore, the simulation of mPWP SIE will be evaluated using three palaeoenvironmental reconstructions that indicate whether sea ice was perennial or seasonal at a specific location. Darby (2008) infers that perennial sea ice was present at Lomonosov Ridge (87.5° N, 138.3° W) throughout the last 14 Ma based on estimates of drift rates of sea ice combined with inferred circum-Arctic sources of detrital mineral grains in sediments at this location. Knies et al. (2014) infer seasonal sea ice cover based on the abundance of the IP_{25} biomarker, a lipid that is produced by certain sea ice diatoms, which is similar to the modern summer minimum throughout the mid-Pliocene in sediments at two locations near the Fram Strait, of which one is chosen for this data-model comparison (80.2° N, 6.4° E). Similarly, Clotten et al. (2018) infer seasonal sea ice cover with occasional sea ice-free conditions in the Iceland Sea (69.1° N, -12.4° E) between 3.5 and 3.0 Ma using a multiproxy approach. As the sediment record studied by Clotten et al. (2018) included a peak in the abundance of the IP_{25} biomarker at 3.2 Ma, we infer seasonal sea ice cover during the KM5c time slice.

3 Arctic warming in the PlioMIP2 ensemble

3.1 Annual mean warming

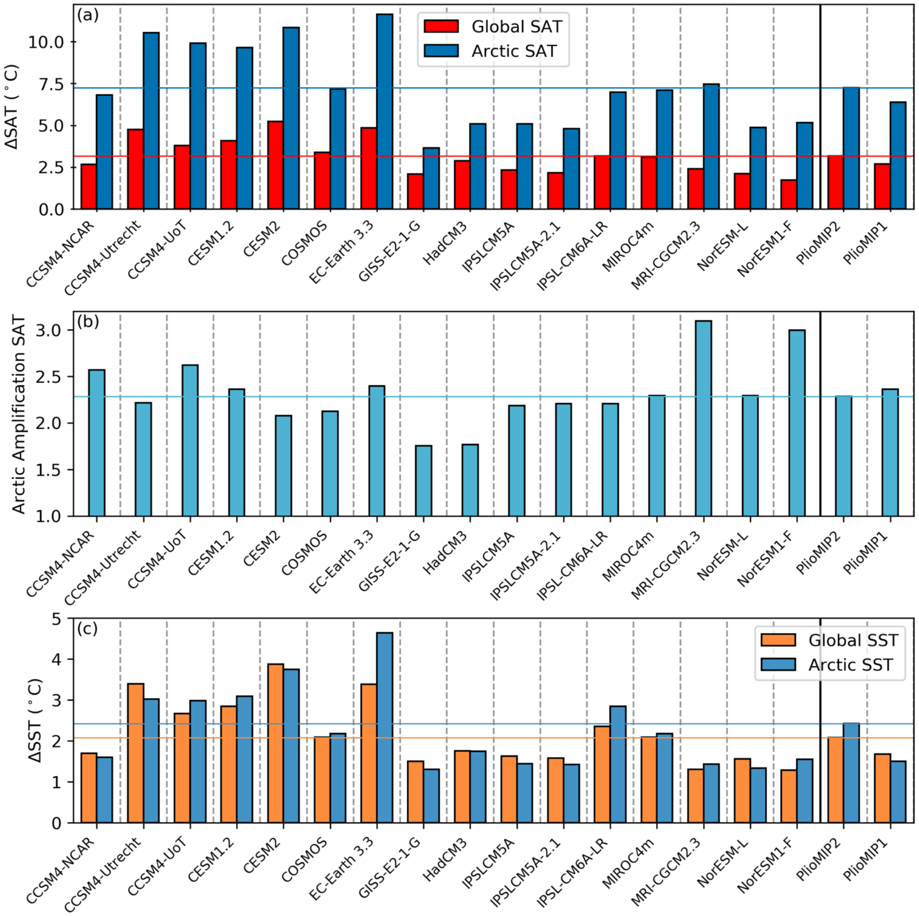


Figure 1: Simulated global and Arctic (a) SAT anomalies (mPWP minus pre-industrial), (b) Arctic amplification ratio of SAT, and (c) SST anomalies for each model and the MMMs. The horizontal lines represent PlioMIP2 MMM values.

The PlioMIP2 experiments show substantial increases in global annual mean SAT (ranging from 1.7 to 5.2 °C, with a MMM of 3.2 °C; Fig. 1a; Table S2) and SST (ranging from 0.8 to 3.9 °C, with a MMM of 2.0 °C; Fig. 1c; Table S2) in the mPWP, compared to pre-industrial.

All models show a clear Arctic amplification, with annual mean SAT in the Arctic (60–90° N) increasing by 3.7 to 11.6 °C (MMM of 7.2 °C; Fig. 1a). The magnitude of Arctic amplification, defined as the ratio between the Arctic and global SAT anomaly, ranges from 1.8 to 3.1, and the MMM shows an Arctic amplification factor of 2.3 (Fig. 1b). There is a large variation in the magnitude of the simulated Arctic SAT anomalies, with five out of sixteen models, namely CCSM4-Utrecht, CCSM4-UoT, CESM1.2, CESM2, and EC-Earth 3.3 all simulating much stronger anomalies than the rest of the ensemble. This subset of the ensemble raises the MMM substantially and this has to be taken into account when interpreting the MMM results. The MMM SAT anomaly for the PlioMIP2 ensemble excluding this subset of five models is 5.8 °C.

Moved down [1]: Annual mean SST in the Arctic increased by 1.3 to 4.6 °C (MMM of 2.4 °C; Fig. 1c).

Annual mean SST in the Arctic increased by 1.3 to 4.6 °C (MMM of 2.4 °C; Fig. 1c). Furthermore, the five models that simulated the largest Arctic SAT anomalies also simulate the largest Arctic SST anomalies. Temperature anomalies in the PlioMIP2 ensemble are similar but slightly higher, than in the PlioMIP1 ensemble. A similar magnitude of Arctic amplification is simulated by the two ensemble means.

Moved (insertion) [1]

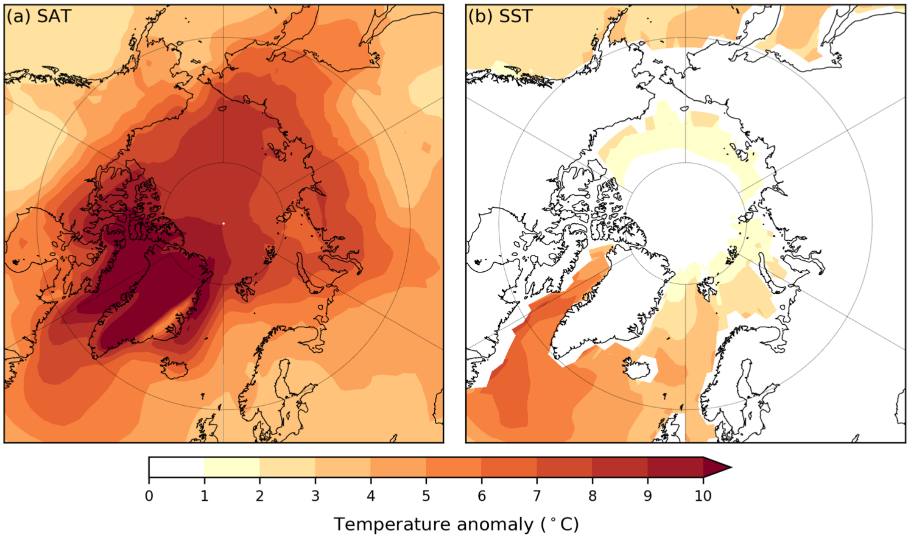


Figure 2: MMM annual temperature anomalies in the Arctic: (a) SAT, (b) SST. At least 15 out of 16 models agree on the sign of change at each location.

The greatest MMM SAT anomalies in the Arctic are found in the regions with reduced ice sheet extent on Greenland (Haywood et al., 2016), which generally show warming of over 10 °C and even up to 20°C. Additionally, temperature anomalies of over 10 °C are simulated around the Baffin Bay. SAT anomalies of around 6–9 °C are simulated over most of the Arctic Ocean regions. SST anomalies in the Arctic are strongest in the Baffin Bay and the Labrador Sea, reaching up to 7 °C (Fig. 2b).

3.2 Seasonal warming

The distinct seasonality of Arctic amplification (Serreze et al., 2009; Zheng et al., 2019) can be used to identify mechanisms causing Arctic amplification. Figure 3 depicts the seasonality of Arctic warming for each model, with monthly SAT and SST anomalies normalized by the annual mean anomaly for that specific model.

The ensemble simulates a consistent peak in Arctic SST warming between July and September (Fig. 3b). This is consistent with the response that increased seasonal heat storage from incoming heat fluxes would have upon the reduction of SIE (Serreze et al., 2009; Zheng et al., 2019). Minimum SAT warming is expected in the summer because of the increased ocean heat uptake, while maximum SAT warming is expected in the autumn and winter following the release of this heat (Pithan and Mauritsen, 2014; Serreze et al., 2009; Yoshimori and Suzuki, 2019; Zheng et al., 2019). This is not simulated by all models, however (Fig. 3a). COSMOS, GISS-E2-1-G, IPSL-CM6A-LR, and MRI-CGCM2.3 all do show this autumn and winter amplification of annual mean SAT anomalies and decreased warming in the summer. Decreased summer warming is simulated by CCSM4-Utrecht, EC-Earth 3.3 and IPSLCM5A in combination with autumn amplification, and by CESM2 and NorESM1-F in combination with winter amplification. All other models in the ensemble do not show an autumn or winter amplification in combination with decreased summer warming, suggesting a more limited role of reductions in SIE underlying the seasonal cycle of Arctic SAT anomalies.

Deleted: coincide with local prescribed changes

Deleted: boundary conditions (Fig. 2a). The

Deleted: generally show warming of over 10 °C and even up to 20°C. Additionally, temperature anomalies of over 10 °C are simulated around the Canadian Archipelago, which is closed in the mPWP simulations (Haywood et al., 2016). SAT anomalies of around 6–9 °C are simulated over most of the Arctic Ocean regions. SST anomalies in the Arctic are strongest in the Baffin Bay and the Labrador Sea, reaching up to 7 °C (Fig. 2b).

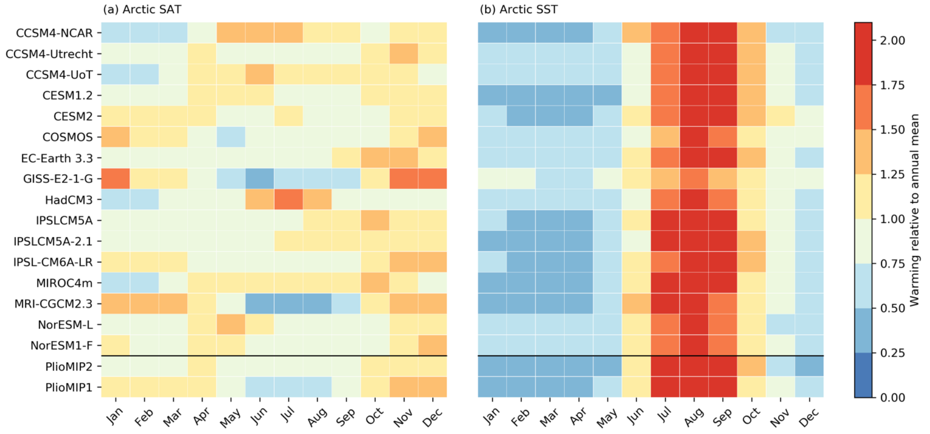


Figure 3: Ratio between the mean Arctic (a) SAT and (b) SST warming in a given month and the annual mean Arctic warming, for each model (and MMM) individually. Values of zero would imply no warming compared to pre-industrial in a given month.

Deleted: Monthly

Deleted: relative to

Deleted: respective

Deleted: in the PlioMIP2 ensemble.

4 Sea ice analysis

4.1 Annual mean sea ice extent

The MMM of Arctic annual SIE (sea ice concentration ≥ 0.15) is 11.9 x 10^6 km^2 for the pre-industrial simulations, and 5.6 x 10^6 km^2 (a 53 % decrease) for the mPWP simulations. The pre-industrial annual mean SIE ranges from 9.1 to 15.6 x 10^6 km^2 in the ensemble, while the mPWP SIE ranges from 2.3 to 10.4 x 10^6 km^2. The decrease in SIE between individual simulations ranges from -3.0 x 10^6 km^2 to -10.4 x 10^6 km^2 (Table S2). Interestingly, the PlioMIP1 MMM shows larger SIEs in both the pre-industrial and the mPWP than any individual model in the PlioMIP2 ensemble (Fig. 4). The 53% MMM decrease in SIE simulated by the PlioMIP2 ensemble is substantially greater than the 33% MMM decrease in SIE simulated by the PlioMIP1 ensemble (Howell et al., 2016a).

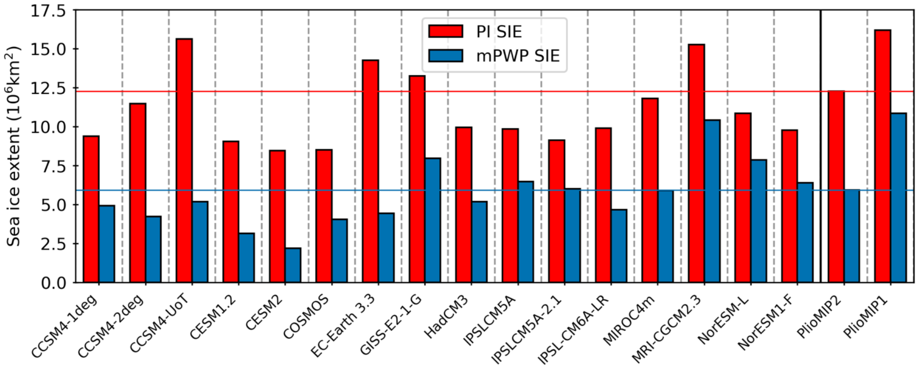
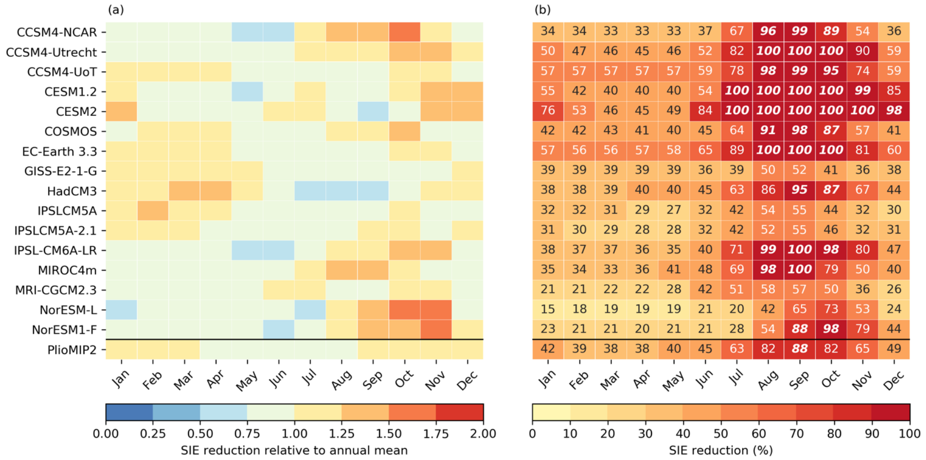


Figure 4: Mean annual SIE (10^6 km^2) for the pre-industrial and mPWP simulations. The horizontal lines represent PlioMIP2 MMM values.

4.2 Monthly mean sea ice extent

The seasonal cycle of SIE anomalies is depicted in Fig. 5a. Reductions in SIE are slightly greater in the autumn (September-November) as compared to other seasons for the MMM. There is, however, no consistent response in the seasonal character of SIE anomalies in the PlioMIP2 ensemble. CCSM4-UoT, CESM2, IPSLCM5A, IPSLCM5A-2.1 simulate the largest reductions in SIE in winter (December-February), while GISS-E2-1-G and HadCM3 simulate the largest SIE reductions in spring. The remaining 10 models simulate the greatest SIE anomalies in autumn.

A more consistent response is observed when comparing monthly mean mPWP SIEs and pre-industrial SIEs. For each model, the largest reductions in SIE in terms of percentages occur between August and October (Fig. 5b). This may be explained by the lesser amount of energy that is needed to melt a given % of the smaller SIE that is present in the summer compared to winter. 11 out of 16 models simulate sea ice-free conditions (SIE < 1×10^6 km^2) in at least one month, while five models (GISS-E2-1-G, IPSLCM5A, IPSLCM5A-2.1, MRI-CGCM2.3, and NorESM-L) do not (Fig. 5b). The NorESM1-F simulation simulates the smallest global mean warming (1.7 °C; Fig. 1a) resulting in Arctic sea ice-free conditions.



	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CCSM4-NCAR	34	34	33	33	33	37	67	***96***	***99***	***89***	54	36
CCSM4-Utrecht	50	47	46	45	46	52	82	***100***	***100***	***100***	90	59
CCSM4-UoT	57	57	57	57	57	59	78	***98***	***99***	***95***	74	59
CESM1.2	55	42	40	40	40	54	***100***	***100***	***100***	***100***	***99***	85
CESM2	76	53	46	45	49	84	***100***	***100***	***100***	***100***	***100***	***98***
COSMOS	42	42	43	41	40	45	64	***91***	***98***	***87***	57	41
EC-Earth 3.3	57	56	56	57	58	65	89	***100***	***100***	***100***	81	60
GISS-E2-1-G	39	39	39	39	39	36	39	50	52	41	36	38
HadCM3	38	38	39	40	40	45	63	86	***95***	***87***	67	44
IPSLCM5A	32	32	31	29	27	32	42	54	55	44	32	30
IPSLCM5A-2.1	31	30	29	28	28	32	42	52	55	46	32	31
IPSL-CM6A-LR	38	37	37	36	35	40	71	***99***	***100***	***98***	80	47
MIROC4m	35	34	33	36	41	48	69	***98***	***100***	79	50	40
MRI-CGCM2.3	21	21	22	22	28	42	51	58	57	50	36	26
NorESM-L	15	18	19	19	19	21	20	42	65	73	53	24
NorESM1-F	23	21	21	20	21	21	28	54	***88***	***98***	79	44
PlioMIP2	42	39	38	38	40	45	63	82	***88***	82	65	49

Figure 5: (a) Monthly SIE anomalies relative to annual mean anomalies, warmer colours highlight in which months reductions in sea ice were largest. (b) Reduction in SIE (%) in the mPWP simulations compared to the pre-industrial monthly mean SIE for each month. Highlighted in bold italics in (b) are months with sea ice-free conditions (SIE < $1x10^6$ km^2).

4.3 Sea ice and Arctic warming

There is a strong anti-correlation between annual mean Arctic SAT and SIE anomalies (R=-0.79; Fig. 6a), as well as between SST and SIE anomalies (R=-0.79; Fig. 6b). These anti-correlations are stronger than those found for the PlioMIP1 ensemble (R=-0.76, R=-0.73, respectively; Howell et al., 2016).

Deleted: The results highlight the importance of correctly simulating SIE anomalies for the Arctic SAT anomalies, which, in turn, may yield better agreement with reconstructions.

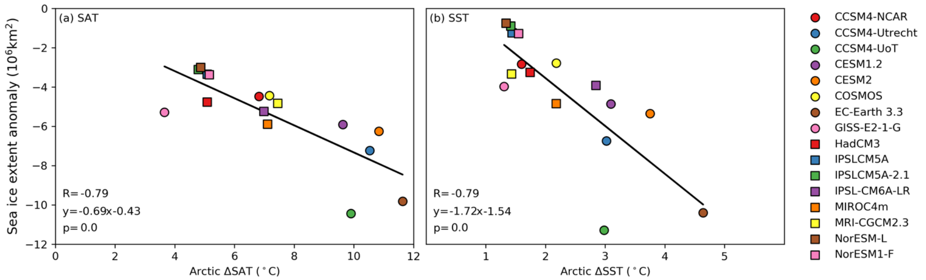


Figure 6: Correlations between annual mean SIE anomalies and (a) Arctic SAT anomalies and (b) Arctic SST anomalies. Depicted for both correlations are the correlation coefficient (R), the slope and the probability value (p) that when the variables are not related, a statistical result equal to or greater than observed would occur.

Deleted:

Formatted

5 Data-model comparison surface air temperatures

5.1 Results

To evaluate the ability of the PlioMIP2 ensemble to simulate Arctic warming, we perform a data-model comparison with the available SAT reconstructions for the mPWP. The data-model comparison hints at a substantial mismatch between models and temperature reconstructions. Mean absolute deviations (MAD) range from 5.0 to 11.2 °C (Table S3), with a MAD of 7.3 °C for the MMM. The median bias ranges from -2.0 to -13.1 °C, with a median bias of -8.2 °C for the MMM (Table S3). The PlioMIP2 MMM shows slightly improved agreement with the SAT reconstructions compared to the PlioMIP1 MMM (MAD = 7.8 °C, median bias = -8.7 °C). Figure 7 depicts the deviation from reconstructions for the MMM. Underestimations range from -17 to -2.5 °C, while at two sites in the Canadian Archipelago (80° N, 85° W and 79.85° N, 99.24° W) the MMM overestimates the reconstructed temperatures (by 2.7 and 1.2 °C, respectively). It has to be noted, however, that SAT anomalies are underestimated at three other sites within the Canadian Archipelago. Given the resolution of global climate models and the close proximity of the sites, it may be impossible for simulations to match all five of these SAT estimates.

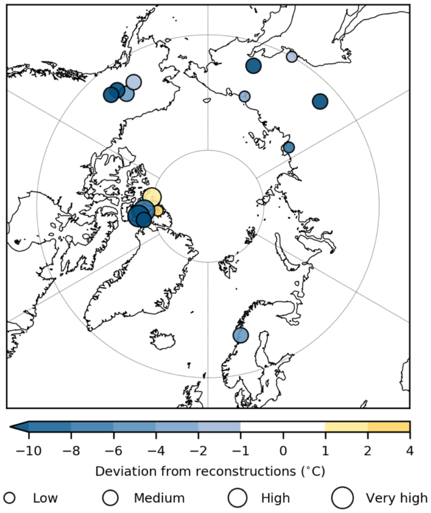


Figure 7: Point-to-point comparison of MMM and reconstructed SAT. The size of SAT reconstructions is scaled by qualitatively assessed confidence levels (Salzmann et al., 2013). Data markers for reconstructions in close proximity of each other have been slightly shifted for improved visibility.

The deviation from reconstructions for each model and the PlioMIP2 and PlioMIP1 MMMs is represented by the box-whisker plots in Fig. 8. A consistent underestimation of the temperature estimates from SAT reconstructions is present in the PlioMIP2 ensemble. CESM2 simulates the smallest deviations from reconstructions in the ensemble, with a MAD of 5.0 °C and a median bias of -2 °C. The five models that simulated the highest Arctic SAT anomalies (CCSM4-Utrecht, CCSM4-UoT, CESM1.2, CESM2, and EC-Earth 3.3) simulate the lowest median biases, indicating that the upper end of the range of simulated Arctic SAT anomalies in the PlioMIP2 ensemble tends to match proxy dataset in its current form better. Future research into the underlying mechanisms for the increased Arctic warming in these five simulations, compared to the remaining eleven simulations in the ensemble, may form a way to uncover factors that contribute to improved data-model agreement.

Deleted: with

Deleted: five

Deleted: in the ensemble

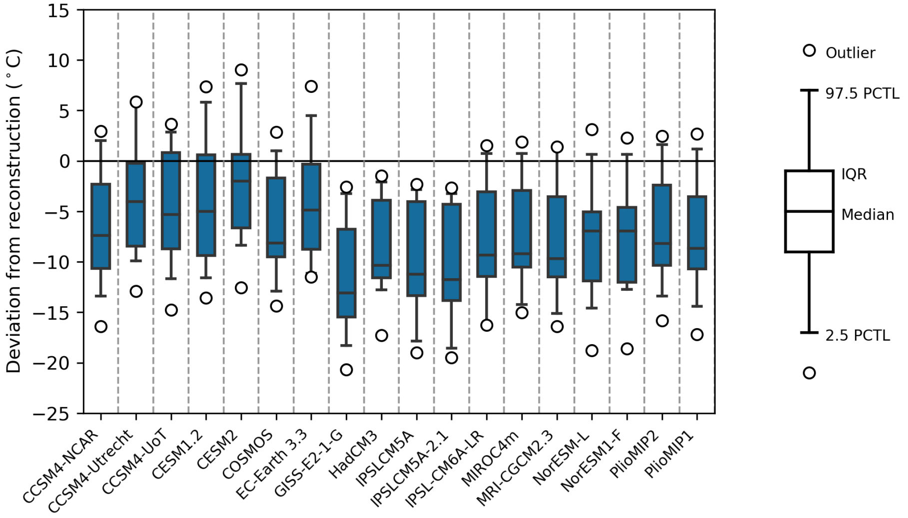


Figure 8: Box-whisker plots depicting the distribution of biases (models minus reconstruction) with biases over (under) 0 representing locations where models overestimated (underestimated) reconstructed temperatures. Boxes depict the interquartile ranges (IQR) of the distribution, whiskers extend to the 2.5th and 97.5th percentiles, the median is displayed by a horizontal line in the boxes, and outliers (outside of the 97.5th percentile) by open circles outside of the whiskers. Given the sample size of 15 reconstructions, the two outer values are depicted as outliers using these definitions.

5.2 Uncertainties

Some of the data-model discord may be caused by uncertainties in the temperature estimates (Table S1; Salzmann et al., 2013). To investigate how these uncertainties may have affected the outcomes of the data-model comparison, we construct a maximum uncertainty range. This range spans from the highest possible temperature within uncertainty and the lowest possible

Deleted: To investigate how these uncertainties may have affected the outcomes of the data-model comparison, we calculate, similar to work by Salzmann et al. (2013),

Deleted: using estimates of temporal and bioclimatic temperature within uncertainty. The uncertainties for the temperature estimates were taken from the compilation of mPWP Arctic SAT estimates from Feng et al. (2017) (Table S1).

Deleted: .

Figure 9 depicts the locations for which at least one model in the ensemble simulates a temperature within the maximum available uncertainty range of a reconstruction. For six out of the twelve reconstructions that included an uncertainty estimate, the models in the PlioMIP2 ensemble simulate temperatures that are within the uncertainty range (Fig. 9). Additionally, both over- and underestimations are present for the Magadan District reconstruction for which no uncertainty estimate is available (60° N, 150.65° E, Table S1), implying that the reconstruction falls within the range of simulated temperatures in the PlioMIP2 ensemble. For the remaining six reconstructions, including several which are assessed high or very high confidence (Figure 9), no model simulates temperatures within the uncertainty range.

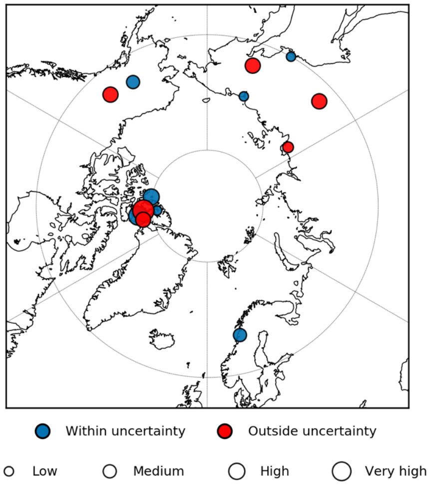


Figure 9: Blue circles highlight where at least one model in the ensemble simulates a temperature that falls within the uncertainty range of the reconstruction. The size of SAT reconstructions is scaled by qualitatively assessed confidence levels (Salzmann et al., 2013). Data markers for reconstructions in close proximity of each other have been slightly shifted for improved visibility.

Ultimately, when considering the full uncertainty ranges of the reconstructions, it becomes evident that solely reducing potential errors in SAT estimates would not fully resolve the data-model discord for several locations in the Arctic. It is thus likely that other sources of error contribute to the data-model discord, such as uncertainties in model physics (e.g. Feng et al., 2019; Howell et al., 2016b; Lunt et al., 2020; Sagoo and Storelvmo, 2017; Unger and Yue, 2014) and boundary conditions (e.g. Brierley and Fedorov, 2016; Feng et al., 2017, 2017; Hill, 2015; Howell et al., 2016b; Otto-Bliesner et al., 2017; Prescott et al., 2014; Robinson et al., 2011; Salzmann et al., 2013). The focus on the KM5c time slice has helped resolve some of the data-model discord that was present in the North Atlantic for SST (Haywood et al., 2020), and similar work for SAT reconstructions may thus be beneficial. However, this may not always be possible given the lack of precise dating and chronologies available. It is at this moment unclear whether the underestimation of Arctic SAT is specific to the mid-Pliocene, through uncertainties in reconstructions or boundary conditions, or an indicator of common errors in model physics.

6 Evaluation of sea ice

The limited availability of proxy evidence (three reconstructions) severely limits our ability to evaluate the simulation of mPWP sea ice in PlioMIP2 simulations. Nevertheless, a data-model comparison is still worthwhile, as the few reconstructions that are available may form an interesting out-of-sample test for the simulation of sea ice in the PlioMIP2 models.

Figure 10a depicts the number of models per grid box that simulate perennial sea ice. Six models simulate the inferred perennial sea ice (mean sea ice concentration ≥ 0.15 in each month) at Lomonosov Ridge (87.5° N, 138.3° W; Darby, 2008), while the remaining ten simulate sea ice-free conditions in at least one month per year at this site. The majority of the models simulate a maximum SIE that extends, or nearly extends, into the Fram Strait and Iceland Sea (Figure 10b) in at least one month (in winter) per year (Fig. 10b), consistent with proxy evidence (Clotten et al., 2018; Knies et al., 2014).

Deleted: ¶

Deleted: Three models: HadCM3, IPSLCM5A, MRI-CGCM2.3, simulate both the perennial sea ice at Lomonosov Ridge and a sea ice minimum that extends into the Fram Strait and Iceland Sea. MRI-CGCM2.3, however, simulates perennial sea ice in the Fram Strait, which does not correspond to seasonally sea ice-free conditions at this site that are inferred from the reconstruction (Knies et al., 2014). Ultimately, only HadCM3 and IPSLCM5A simulate sea ice that is consistent with all three available sea ice reconstructions. Interestingly, in an analysis of simulated sea ice in the PlioMIP1 ensemble, the HadCM3 simulation was also found to be consistent with proxy evidence, and it was the only model fulfilling this criterion (Howell et al., 2016a). The HadCM3 and IPSLCM5A simulations, however, do not show improved agreement with regard to SAT reconstructions than the PlioMIP2 MMM (Fig. 8), indicating that matching the sea ice reconstructions does not necessarily imply a better match with the SAT reconstructions.¶

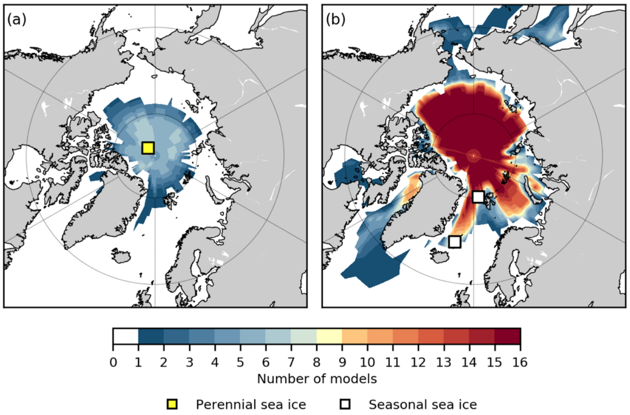


Figure 10: Number of models simulating (a) annual mean perennial sea ice (sea ice concentration of ≥ 0.15) at any given location in the Arctic in the mPWP simulations and (b) monthly mean sea ice in any month of the year. Depicted squares represent the locations of the reconstructions and their respective colour the inferred mPWP sea ice conditions at that location.

The uncertainties in both the SAT and SIE reconstructions are large, and it may not be possible to match both datasets in their current forms. This would require increased Arctic annual terrestrial warming compared to the mean model (Sect. 5.1) as well as perennial sea in the summer and a large SIE in winter (extending at least into the Iceland Sea). Moreover, McClymont et al. (2020) found that the warmest model values in the PlioMIP2 ensemble tend to align best with North Atlantic SST reconstructions, further indicating that strong Arctic warming is required for data-model agreement. If there was no perennial sea ice in the mPWP like most models in the PlioMIP2 ensemble, the different proxy records may be more compatible, but this would be in disagreement with findings from (Darby, 2008). The CCSM4-Utrecht model, which simulated a relatively high Arctic SAT anomaly (10.5 °C; Figure 1a) and low median bias (-4 °C) in the point-to-point SAT data-model comparison compared to the rest of the ensemble, simulates a maximum winter SIE that extends both into the Fram Strait and Iceland Sea. This highlights that models with higher Arctic SAT anomalies and better SAT data-model agreement can still match both seasonal sea ice proxies. Ultimately, more reconstructions of sea ice are needed for a more robust evaluation of mPWP sea ice and Arctic warming in general.

7 Comparison to future climates

Research into the mPWP is often motivated by a desire to understand future climate change (Burke et al., 2018; Haywood et al., 2016; Tierney et al., 2019). Here, we analyze how the mPWP may teach us about future Arctic warming by comparing two climatic features of the mPWP simulations to simulations of future climate. The climatic features include Arctic amplification, and a feature for which there is some proxy evidence available that may also aid in model evaluation: the AMOC.

Deleted: some

7.1 Arctic amplification

A linear relationship between global and Arctic temperature anomalies is present in the PlioMIP2 ensemble (R=0.93, Fig. 11a). This is consistent with findings from multi-model analyses of other climates (Bracegirdle and Stephenson, 2013; Harrison et al., 2015; Izumi et al., 2013; Masson-Delmotte et al., 2006; Miller et al., 2010; Schmidt et al., 2014; Winton, 2008) and indicates that global temperature anomalies are a good index for Arctic SAT anomalies in mPWP simulations.

For four ensembles of future climate simulations, from the previous phase of the Coupled Model Intercomparison Project (CMIP), CMIP5, data for MMM Arctic (defined there as 67.5–90° N) temperature anomalies are available (Masson-Delmotte et al., 2013; Table S4). The PlioMIP2 MMM shows global warming that falls between the RCP6.0 and RCP8.5 MMMs in terms of magnitude (Fig. 11b). Even though PlioMIP underestimates mPWP SAT reconstructions (Sect. 5.1), the simulations do simulate stronger Arctic temperature anomalies per degree of global warming compared to future climate ensembles (Fig.

11b). The future climate ensemble MMMs simulate end-of-century (2081-2100) average Arctic (67.5–90° N) amplification ratios that range from 2.2 to 2.4, while PlioMIP2 and PlioMIP1 simulate mean ratios of 2.8 and 2.7, respectively (Table S4).

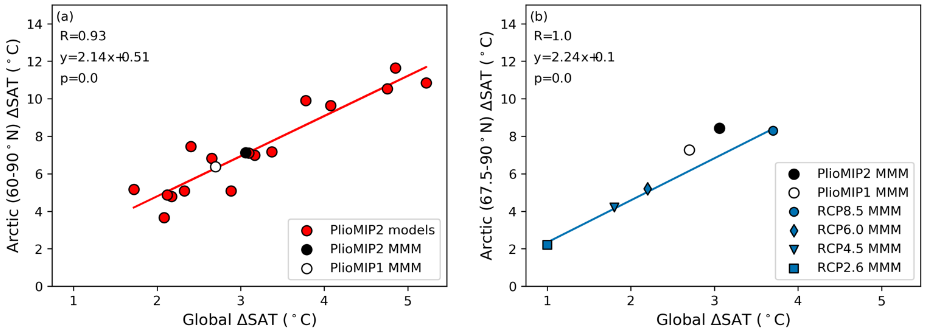
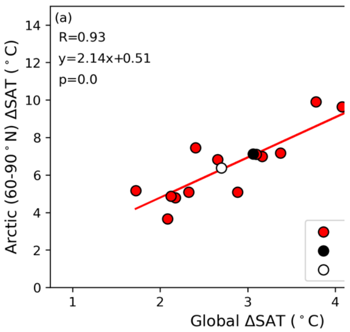


Figure 11: (a) The relationship between global and Arctic (60–90° N) temperature anomalies in the PlioMIP2 ensemble. The red trendline is constructed based on this relationship for the individual models. (b) The relationship between global and Arctic (here 67.5–90° N, the definition used by Masson-Delmotte et al. (2013) and the area for which they listed data) for the MMMs of the two PlioMIP and the four CMIP5 future climate ensembles (2081-2100 average). The blue trendline highlights this relationship for the RCP MMMs.

The increased Arctic warming per degree of global warming indicates that apart from warming through changes in atmospheric CO_2 concentrations, which is the dominant mechanism for warming in both ensembles, different or additional mechanisms underly the simulated mPWP Arctic warming compared to the future climate simulations. The difference between the PlioMIP2 and future climate ensembles may be explained by slow responses to changes in forcings that fully manifest in equilibrium climate simulations, such as the response to reduced ice sheets, but not in transient, near-future, climate simulations. Additional Arctic warming in the mPWP simulations may arise due to the changes in orography (Brierley and Fedorov, 2016; Feng et al., 2017; Haywood et al., 2016; Otto-Bliesner et al., 2017), ice sheets, and vegetation in the boundary conditions (Hill, 2015; Lunt et al., 2012b).

Using PlioMIP2 simulations for potential lessons about future warming may be improved by isolating the effects of the changes in orograph. Similar changes in ice sheets and vegetation may occur in future equilibrium warm climates, but the changes in orography are definitively non-analogous to future warming. Several groups isolated the effects of the changed orography on global warming in PlioMIP2 simulations and found that it contributes, respectively, around 23% (IPSL6-CM6A-LR; Tan et al., 2020), 27% (COSMOS; Stepanek et al., 2020), and 41% (CCSM4-UoT; Chendan and Peltier, 2018) to the annual mean global warming in the mPWP simulations. Furthermore, this warming was strongest in the high latitudes (Chendan and Peltier, 2018; Tan et al., 2020) indicating that the additional Arctic warming in PlioMIP2 simulations, as compared to future climate



Deleted:

Deleted: due to limited

Deleted: availability of the future climate simulations

Deleted: .

Deleted: Arctic ocean gateways

Deleted: other changes in orography

simulations, are likely partially caused by changes in orography that are non-analogous with the modern-day orography. These findings highlight the caution that has to be taken when using palaeoclimate simulations as analogues for future climate change.

7.2 Atlantic meridional overturning circulation

The AMOC, a major oceanic current transporting heat into the Arctic (Mahajan et al., 2011), is inferred to have been significantly stronger in the mPWP compared to pre-industrial based on proxy evidence (Dowsett et al., 2009; Frank et al., 2002; Frenz et al., 2006; McKay et al., 2012; Ravelo and Andreasen, 2000; Raymo et al., 1996). An analysis of AMOC changes in PlioMIP2 simulations shows that, indeed, the maximum AMOC strength increases: by 4 to 53% (Fig. 12; Table S2: Li et al., in prep.). The closure of the Arctic Ocean gateways, in particular the Bering Strait, likely contributed to the increase in AMOC strength (Brierley and Fedorov, 2016; Feng et al., 2017; Haywood et al., 2016; Otto-Bliesner et al., 2017).

Strengthening of the AMOC contrasts projections of future changes by CMIP5 models which predict a weakening of the AMOC over the 21st century, with best estimates ranging from 11 to 34% depending on the chosen future emission scenario (Collins et al., 2013). These opposing responses may help explain some of the additional Arctic warming that is observed in the PlioMIP2 ensemble compared to the future climate ensembles (Fig. 11b).

Deleted: (Fig.

The strengthening of the AMOC in the PlioMIP2 ensemble is consistent with the additional 0.4 °C increase in SST warming in the Arctic (Figure 1c) and the better data-model agreement in the North Atlantic that is observed for the PlioMIP2 MMM (Dowsett et al., 2019; Haywood et al., 2020; McClymont et al., 2020) compared to the PlioMIP1 MMM (Fig. 1c), which did not show any substantial changes in AMOC strength compared to pre-industrial (Zhang et al., 2013).

Deleted: This

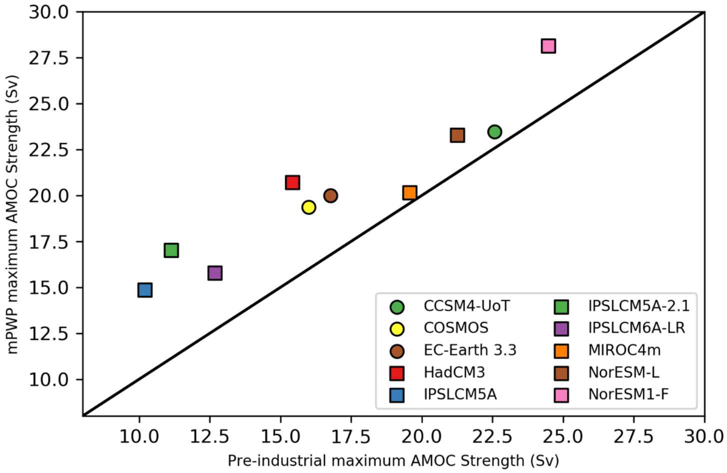


Figure 12: Maximum pre-industrial and mPWP AMOC strength (Sv). The black line indicates equal pre-industrial and mPWP maximum AMOC strength.

8 Conclusions

The PlioMIP2 ensemble simulates substantial Arctic warming and 11 out of 16 models simulate summer sea ice-free conditions. Comparisons to reconstructions show, however, that the ensemble tends to underestimate the available reconstructions of SAT in the Arctic, although large differences in the degree of underestimation exist between the simulations. The models that simulate the largest Arctic SAT anomalies tend to match the reconstructions better, and investigation into the mechanisms underlying the increased Arctic warming in these simulations may help uncover factors that could contribute to improved data-model agreement. We find that, while some of the SAT data-model discord may be resolved by reducing uncertainties in proxies, additional improvements are likely to be found in reducing uncertainties in boundary conditions or model physics. Furthermore, there is some agreement with reconstructions of sea ice in the ensemble, especially for seasonal sea ice. The limited availability of proxy evidence and the uncertainties associated with them severely constrain the compatibility of the different proxy datasets and our ability to evaluate the Arctic warming in PlioMIP2. Increased proxy evidence of different climatic variables, and additional sensitivity experiments, among others, are needed for a more robust evaluation of Arctic warming in the mPWP. Lastly, we find differences in Arctic climate features between the PlioMIP2 ensemble and future climate ensembles, that include the magnitude of Arctic amplification and changes in AMOC strength. These differences highlight that caution has to be taken when attempting to use simulations of the mPWP to learn about future climate change.

Deleted: .

Deleted: enhanced

Supplement.

Formatted: Normal

Formatted: Font: Bold, Not Italic

Author contributions. Qiong Z. and Wesley de Nooijer designed the work, Wesley de Nooijer did the analyses and wrote the manuscript under supervision from Qiong Z., Q. L. and Qiang Z. performed the simulations with EC-Earth3. X. L. and Z.Z. provided input on AMOC analysis. H. D. provided the input on reconstructions. All the other co-authors provided the PlioMIP2 model data and commented on the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work is supported by the Swedish Research Council (VR) funded projects 2013-06476 and 2017-04232. The EC-Earth3 simulations are performed on the Swedish National Infrastructure for Computing (SNIC) at the National Supercomputer Centre (NSC). COSMOS PlioMIP2 simulations have been conducted at the Computing and Data Center of the Alfred-Wegener-Institut– Helmholtz-Zentrum für Polar und Meeresforschung on a NEC SX-ACE high performance vector computer. G. L. and C. S. acknowledge funding via the Alfred Wegener Institute's research programme PACES2. C. S. acknowledges funding by the Helmholtz Climate Initiative REKLIM. C.C. and G.R. thank ANR HADOC ANR-17-CE31-0010. Authors 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. W.-L.C. and A.A.-O. acknowledge funding from JSPS KAKENHI grant 17H06104 and MEXT KAKENHI grant 17H06323, and JAMSTEC for use of the Earth Simulator supercomputer. The PRISM4 reconstruction and boundary conditions used in PlioMIP2 were funded by the U.S. Geological Survey Climate and Land Use Change Research and Development Program. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Bakker, P., Stone, E. J., Charbit, S., Gröger, M., Krebs-Kanzow, U., Ritz, S. P., Varma, V., Khon, S., Lunt, D. J., Mikolajewicz, U., Prange, M., Renssen, H., Schneider, B. and Schulz, M.: Last interglacial temperature evolution – a model inter-comparison, Clim. Past Discuss., 8(5), 4663–4699, doi:10.5194/cpd-8-4663-2012, 2012.

Ballantyne, A. P., Greenwood, D. R., Damsté, J. S. S., Csank, A. Z., Eberle, J. J. and Rybczynski, N.: Significantly warmer Arctic surface temperatures during the Pliocene indicated by multiple independent proxies, Geology, 38(7), 603–606, doi:10.1130/G30815.1, 2010.

Bracegirdle, T. J. and Stephenson, D. B.: On the Robustness of Emergent Constraints Used in Multimodel Climate Change Projections of Arctic Warming, J. Clim., 26(2), 669–678, doi:10.1175/JCLI-D-12-00537.1, 2013.

Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, Th., Hewitt, C. D., Kageyama, M., Kitoh, A., Laîné, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., Yu, Y. and Zhao, Y.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features, Clim. Past, 3(2), 261–277, 2007.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B. and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nat. Clim. Change, 2(6), 417–424, doi:10.1038/nclimate1456, 2012.

Brierley, C. M. and Fedorov, A. V.: Comparing the impacts of Miocene–Pliocene changes in inter-ocean gateways on climate: Central American Seaway, Bering Strait, and Indonesia, Earth Planet. Sci. Lett., 444, 116–130, doi:10.1016/j.epsl.2016.03.010, 2016.

Brigham-Grette, J., Melles, M., Minyuk, P., Andreev, A., Tarasov, P., DeConto, R., Koenig, S., Nowaczyk, N., Wennrich, V., Rosén, P., Haltia, E., Cook, T., Gebhardt, C., Meyer-Jacob, C., Snyder, J. and Herzschuh, U.: Pliocene Warmth, Polar Amplification, and Stepped Pleistocene Cooling Recorded in NE Arctic Russia, Science, 340(6139), 1421–1427, doi:10.1126/science.1233137, 2013.

Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J. and Otto-Bliesner, B. L.: Pliocene and Eocene provide best analogs for near-future climates, Proc. Natl. Acad. Sci., 115(52), 13288–13293, doi:10.1073/pnas.1809600115, 2018.

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

Chan, W. L., and Abe-Ouchi, A.: Pliocene Model Intercomparison Project (PlioMIP2) simulations using the Model for Interdisciplinary Research on Climate (MIROC4m), Clim. Past, 16, 1523-1545, 10.5194/cp-16-1523-2020, 2020.

Christensen, J. H., Kanikicharla, K. K., Aldrian, E., An, S. I., Cavalcanti, I. F. A., Castro, M. de, Dong, W., Goswami, P., Hall, A., Kanyanga, J. K., Kitoh, A., Kossin, J., Lau, N. C., Renwick, J., Stephenson, D. B., Xie, S. P., Zhou, T., Abraham, L., Ambrizzi, T., Anderson, B., Arakawa, O., Arritt, R., Baldwin, M., Barlow, M., Barriopedro, D., Biasutti, M., Biner, S., Bromwich, D., Brown, J., Cai, W., Carvalho, L. V., Chang, P., Chen, X., Choi, J., Christensen, O. B., Deser, C., Emanuel, K., Endo, H., Enfield, D. B., Evan, A., Giannini, A., Gillett, N., Hariharasubramanian, A., Huang, P., Jones, J., Karumuri, A., Katzfey, J., Kjellström, E., Knight, J., Knutson, T., Kulkarni, A., Kundeti, K. R., Lau, W. K., Lenderink, G., Lennard, C., Leung, L. yung R., Lin, R., Losada, T., Mackellar, N. C., Magaña, V., Marshall, G., Mearns, L., Meehl, G., Menéndez, C., Murakami, H., Nath, M. J., Neelin, J. D., Oldenborgh, G. J. van, Olesen, M., Polcher, J., Qian, Y., Ray, S., Reich, K. D., Fonseca, B. R. de, Ruti, P., Screen, J., Sedláček, J., Solman, S., Stendel, M., Stevenson, S., Takayabu, I., Turner, J., Ummenhofer, C., Walsh, K., Wang, B., Wang, C., Watterson, I., Widlansky, M., Wittenberg, A., Woollings, T., Yeh, S. W., Zhang, C., Zhang, L., Zheng, X. and Zou, L.: Climate phenomena and their relevance for future regional climate change, Clim. Change 2013 Phys. Sci. Basis Work. Group Contrib. Fifth Assess. Rep. Intergov. Panel Clim. Change, 1217–1308, doi:10.1017/CBO9781107415324.028, 2013.

Clotten, C., Stein, R., Fahl, K. and De Schepper, S.: Seasonal sea ice cover during the warm Pliocene: Evidence from the Iceland Sea (ODP Site 907), Earth Planet. Sci. Lett., 481, 61–72, doi:10.1016/j.epsl.2017.10.011, 2018.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., Wehner, M. F., Allen, M. R., Andrews, T., Beyerle, U., Bitz, C. M., Bony, S. and Booth, B. B. B.: Long-term Climate Change: Projections, Commitments and Irreversibility, Clim. Change 2013 - Phys. Sci. Basis Contrib. Work. Group Fifth Assess. Rep. Intergov. Panel Clim. Change, 1029–1136, 2013.

Darby, D. A.: Arctic perennial ice cover over the last 14 million years, Paleoceanography, doi:10.1029/2007PA001479@10.1002/(ISSN)1944-9186.CENTARC1, 2008.

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., Pound, M., Salzmann, U., Robinson, M., Chandler, M., Foley, K. and Haywood, A.: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction, Clim. Past, 12(7), 1519–1538, doi:https://doi.org/10.5194/cp-12-1519-2016, 2016.

Dowsett, H. J., Robinson, M. M. and Foley, K. M.: Pliocene three-dimensional global ocean temperature, , 28, 2009.

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, Nat. Clim. Change, 2(5), 365–371, doi:10.1038/nclimate1455, 2012.

Dowsett, H. J., Robinson, M. M., Stoll, D. K., Foley, K. M., Johnson, A. L. A., Williams, M. and Riesselman, C. R.: The PRISM (Pliocene palaeoclimate) reconstruction: time for a paradigm shift, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 371(2001), 20120524, doi:10.1098/rsta.2012.0524, 2013.

Dowsett, H. J., Robinson, M. M., Foley, K. M., Herbert, T. D., Otto-Bliesner, B. L. and Spivey, W.: Mid-piacenzian of the north Atlantic Ocean, Stratigraphy, 16(3), 119144, doi:10.29041/strat.16.3.119-144, 2019.

Feng, R., Otto-Bliesner, B. L., Fletcher, T. L., Tabor, C. R., Ballantyne, A. P. and Brady, E. C.: Amplified Late Pliocene terrestrial warmth in northern high latitudes from greater radiative forcing and closed Arctic Ocean gateways, Earth Planet. Sci. Lett., 466, 129–138, doi:10.1016/j.epsl.2017.03.006, 2017.

Feng, R., Otto-Bliesner, B. L., Xu, Y., Brady, E., Fletcher, T. and Ballantyne, A.: Contributions of aerosol-cloud interactions to mid-Piacenzian seasonally sea ice-free Arctic Ocean, Geophys. Res. Lett., doi:10.1029/2019GL083960, 2019.

Feng, R., Otto-Bliesner, B. L., Brady, E. C., and Rosenbloom, N.: Increased Climate Response and Earth System Sensitivity From CCSM4 to CESM2 in Mid-Pliocene Simulations, Journal of Advances in Modeling Earth Systems, 12, e2019MS002033, 10.1029/2019ms002033, 2020.

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

Frank, M., Whiteley, N., Kasten, S., Hein, J. R. and O'Nions, K.: North Atlantic Deep Water export to the Southern Ocean over the past 14 Myr: Evidence from Nd and Pb isotopes in ferromanganese crusts, Paleoceanography, 17(2), 12-1-12–9, doi:10.1029/2000PA000606, 2002.

Frenz, M., Henrich, R. and Zychla, B.: Carbonate preservation patterns at the Ceará Rise – Evidence for the Pliocene super conveyor, Mar. Geol., 232(3), 173–180, doi:10.1016/j.margeo.2006.07.006, 2006.

Harrison, S. P., Bartlein, P. J., Brewer, S., Prentice, I. C., Boyd, M., Hessler, I., Holmgren, K., Izumi, K. and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Clim. Dyn., 43(3), 671–688, doi:10.1007/s00382-013-1922-6, 2014.

Harrison, S. P., Bartlein, P. J., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P. and Kageyama, M.: Evaluation of CMIP5 palaeo-simulations to improve climate projections, Nat. Clim. Change, 5(8), 735–743, doi:10.1038/nclimate2649, 2015.

Haywood, A. M. and Valdes, P. J.: Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere, Earth Planet. Sci. Lett., 218(3), 363–377, doi:10.1016/S0012-821X(03)00685-X, 2004.

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, Clim. Past, 9(1), 191–209, doi: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. R. Soc. Math. Phys. Eng. Sci., 371(2001), 20120515, doi: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, Clim. Past, 12(3), 663–675, doi:https://doi.org/10.5194/cp-12-663-2016, 2016.

Haywood, A. M., Tindall, J. C., Dowsett, H. J., Dolan, A. M., Foley, K. M., Hunter, S. J., Hill, D. J., Chan, W.-L., Abe-Ouchi, A., Stepanek, C., Lohmann, G., Chandan, D., Peltier, W. R., Tan, N., Contoux, C., Ramstein, G., Li, X., Zhang, Z., Guo, C., Nisancioglu, K. H., Zhang, Q., Li, Q., Kamae, Y., Chandler, M. A., Sohl, L. E., Otto-Bliesner, B. L., Feng, R., Brady, E. C., Heydt, A. S. von der, Baatsen, M. L. J. and Lunt, D. J.: A return to large-scale features of Pliocene climate: the Pliocene Model Intercomparison Project Phase 2, Clim. Past Discuss., 1–40, doi:https://doi.org/10.5194/cp-2019-145, 2020.

Hill, D. J.: The non-analogue nature of Pliocene temperature gradients, Earth Planet. Sci. Lett., 425, 232–241, doi:10.1016/j.epsl.2015.05.044, 2015.

Hill, D. J., Csank, A. Z., Dolan, A. M. and Lunt, D. J.: Pliocene climate variability: Northern Annular Mode in models and tree-ring data, Palaeogeogr. Palaeoclimatol. Palaeoecol., 309(1), 118–127, doi:10.1016/j.palaeo.2011.04.003, 2011.

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, Clim. Past, 10(1), 79–90, doi:https://doi.org/10.5194/cp-10-79-2014, 2014.

Howell, F. W., Haywood, A. M., Otto-Bliesner, B. L., Bragg, F., 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, Clim. Past, 12(3), 749–767, doi:https://doi.org/10.5194/cp-12-749-2016, 2016a.

Howell, F. W., Haywood, A. M., Dowsett, H. J. and Pickering, S. J.: Sensitivity of Pliocene Arctic climate to orbital forcing, atmospheric CO2 and sea ice albedo parameterisation, Earth Planet. Sci. Lett., 441, 133–142, doi:10.1016/j.epsl.2016.02.036, 2016b.

Hunter, S. J., Haywood, A. M., Dolan, A. M. and Tindall, J. C.: The HadCM3 contribution to PlioMIP phase 2, Clim. Past, 15(5), 1691–1713, doi:10.5194/cp-15-1691-2019, 2019.

Hurrell, J. W. and Deser, C.: North Atlantic climate variability: The role of the North Atlantic Oscillation, J. Mar. Syst., 79(3), 231–244, doi:10.1016/j.jmarsys.2009.11.002, 2010.

Izumi, K., Bartlein, P. J. and Harrison, S. P.: Consistent large-scale temperature responses in warm and cold climates, Geophys. Res. Lett., 40(9), 1817–1823, doi:10.1002/grl.50350, 2013.

Joussaume, S. and Taylor, K. E.: Status of the paleoclimate modeling intercomparison project (PMIP), World Meteorol. Organ.-Publ.-WMO TD, 425–430, 1995.

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

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Miller, R. L., Murray, L. T., Oinas, V., Orbe, C., García-Pando, C. P., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., and Yao, M.-S.: GISS-E2.1: Configurations and Climatology, Journal of Advances in Modeling Earth Systems, 12, e2019MS002025, 10.1029/2019ms002025, 2020.

Knies, J., Cabedo-Sanz, P., Belt, S. T., Baranwal, S., Fietz, S. and Rosell-Melé, A.: The emergence of modern sea ice cover in the Arctic Ocean, Nat. Commun., 5, 5608, doi:10.1038/ncomms6608, 2014.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. and Meehl, G. A.: Challenges in Combining Projections from Multiple Climate Models, J. Clim., 23(10), 2739–2758, doi:10.1175/2009JCLI3361.1, 2010.

Li, J. and Wang, J. X. L.: A modified zonal index and its physical sense, Geophys. Res. Lett., 30(12), doi:10.1029/2003GL017441, 2003.

Li, X., Zhang, Z., Guo, C., Otterå, O. H. and others: Pliocene Atlantic Meridional Overturning Circulation simulated in PlioMIP2, in prep.

Li, X., Guo, C., Zhang, Z., Otterå, O. H. and Zhang, R.: PlioMIP2 simulations with NorESM-L and NorESM1-F, Clim. Past, 16(1), 183–197, doi:https://doi.org/10.5194/cp-16-183-2020, 2020.

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, Clim. Past, 8(5), 1717–1736, doi:10.5194/cp-8-1717-2012, 2012a.

Lunt, D. J., Haywood, A. M., Schmidt, G. A., Salzmann, U., Valdes, P. J., Dowsett, H. J. and Loptson, C. A.: On the causes of mid-Pliocene warmth and polar amplification, Earth Planet. Sci. Lett., 321–322, 128–138, doi:10.1016/j.epsl.2011.12.042, 2012b.

Lunt, D. J., Abe-Ouchi, A., Bakker, P., Berger, A., Braconnot, P., Charbit, S., Fischer, N., Herold, N., Jungclaus, J. H., Khon, V. C., Krebs-Kanzow, U., Langebroek, P. M., Lohmann, G., Nisancioglu, K. H., Otto-Bliesner, B. L., Park, W., Pfeiffer, M., Phipps, S. J., Prange, M., Rachmayani, R., Renssen, H., Rosenbloom, N., Schneider, B., Stone, E. J., Takahashi, K., Wei, W., Yin, Q. and Zhang, Z. S.: A multi-model assessment of last interglacial temperatures, Clim. Past, 9(2), 699–717, doi:10.5194/cp-9-699-2013, 2013.

Lunt, D. J., Bragg, F., Chan, W.-L., Hutchinson, D. K., Ladant, J.-B., Niezgodzki, I., Steinig, S., Zhang, Z., Zhu, J., Abe-Ouchi, A., de Boer, A. M., Coxall, H. K., Donnadieu, Y., Knorr, G., Langebroek, P. M., Lohmann, G., Poulsen, C. J., Sepulchre, P., Tierney, J., Valdes, P. J., Dunkley Jones, T., Hollis, C. J., Huber, M. and Otto-Bliesner, B. L.: DeepMIP: Model intercomparison of early Eocene climatic optimum (EECO) large-scale climate features and comparison with proxy data, preprint, Climate Modelling/Marine Archives/Cenozoic., 2020.

Lurton, T., Balkanski, Y., Bastrikov, V., Bekki, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Contoux, C., Cozic, A., Cugnet, D., Dufresne, J.-L., Éthé, C., Foujols, M.-A., Ghattas, J., Hauglustaine, D., Hu, R.-M., Kageyama, M., Khodri, M., Lebas, N., Levavasseur, G., Marchand, M., Ottlé, C., Peylin, P., Sima, A., Szopa, S., Thiéblemont, R., Vuichard, N., and Boucher, O.: Implementation of the CMIP6 Forcing Data in the IPSL-CM6A-LR Model, Journal of Advances in Modeling Earth Systems, 12, e2019MS001940, 10.1029/2019ms001940, 2020.

Mahajan, S., Zhang, R. and Delworth, T. L.: Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic Surface Air Temperature and Sea Ice Variability, J. Clim., 24(24), 6573–6581, doi:10.1175/2011JCLI4002.1, 2011.

Masson-Delmotte, V., Kageyama, M., Braconnot, P., Charbit, S., Krinner, G., Ritz, C., Guilyardi, E., Jouzel, J., Abe-Ouchi, A., Crucifix, M., Gladstone, R. M., Hewitt, C. D., Kitoh, A., LeGrande, A. N., Marti, O., Merkel, U., Motoi, T., Ohgaito, R., Otto-Bliesner, B., Peltier, W. R., Ross, I., Valdes, P. J., Vettoretti, G., Weber, S. L., Wolk, F. and YU, Y.: Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints, Clim. Dyn., 26(5), 513–529, doi:10.1007/s00382-005-0081-9, 2006.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez Rouco, J. F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X. and Timmermann, A.: Information from paleoclimate archives, in Climate change 2013: the physical science basis, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 383–464, Cambridge University Press., 2013.

McClymont, E. L., Ford, H. L., Ho, S. L., Tindall, J. C., Haywood, A. M., Alonso-Garcia, M., Bailey, I., Berke, M. A., Littler, K., Patterson, M., Petrick, B., Peterse, F., Ravelo, A. C., Risebrobakken, B., Schepper, S. D., Swann, G. E. A., Thirumalai, K., Tierney, J. E., Weijst, C. van der and White, S.: Lessons from a high CO_2 world: an ocean view from ~ 3 million years ago, Clim. Past Discuss., 1–27, doi:https://doi.org/10.5194/cp-2019-161, 2020.

McKay, R., Naish, T., Carter, L., Riesselman, C., Dunbar, R., Sjunneskog, C., Winter, D., Sangiorgi, F., Warren, C., Pagani, M., Schouten, S., Willmott, V., Levy, R., DeConto, R. and Powell, R. D.: Antarctic and Southern Ocean influences on Late Pliocene global cooling, Proc. Natl. Acad. Sci., 109(17), 6423–6428, doi:10.1073/pnas.1112248109, 2012.

Miller, G. H., Alley, R. B., Brigham-Grette, J., Fitzpatrick, J. J., Polyak, L., Serreze, M. C. and White, J. W. C.: Arctic amplification: can the past constrain the future?, Quat. Sci. Rev., 29(15–16), 1779–1790, doi:10.1016/j.quascirev.2010.02.008, 2010.

Otto-Bliesner, B. L., Rosenbloom, N., Stone, E. J., McKay, N. P., Lunt, D. J., Brady, E. C. and Overpeck, J. T.: How warm was the last interglacial? New model–data comparisons, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 371(2001), 20130097, doi:10.1098/rsta.2013.0097, 2013.

Otto-Bliesner, B. L., Jahn, A., Feng, R., Brady, E. C., Hu, A. and Löfverström, M.: Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways, Geophys. Res. Lett., 44(2), 957–964, doi:10.1002/2016GL071805, 2017.

Otto-Bliesner, B. L., Brady, E. C., Zhao, A., Brierley, C., Axford, Y., Capron, E., Govin, A., Hoffman, J., Isaacs, E., Kageyama, M., Scussolini, P., Tzedakis, P. C., Williams, C., Wolff, E., Abe-Ouchi, A., Braconnot, P., Ramos Buarque, S., Cao, J., de Vernal, A., Guarino, M. V., Guo, C., LeGrande, A. N., Lohmann, G., Meissner, K., Menviel, L., Nisancioglu, K., O&apos;ishi, R., Salas Y Melia, D., Shi, X., Sicard, M., Sime, L., Tomas, R., Volodin, E., Yeung, N., Zhang, Q., Zhang, Z. and Zheng, W.: Large-scale features of Last Interglacial climate: Results from evaluating the <i>lig127k</i> simulations for CMIP6-PMIP4, preprint, Climate Modelling/Modelling only/Pleistocene., 2020.

Panitz, S., Salzmann, U., Risebrobakken, B., De Schepper, S. and Pound, M.: Climate variability and long-term expansion of peatlands in Arctic Norway during the late Pliocene (ODP Site 642, Norwegian Sea), Clim. Past, 12, 1043–1060, doi:https://doi.org/10.5194/cp-12-1043-2016, 2016.

Pithan, F. and Mauritsen, T.: Arctic amplification dominated by temperature feedbacks in contemporary climate models, Nat. Geosci., 7(3), 181–184, doi:10.1038/ngeo2071, 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 Planet. Sci. Lett., 400, 261–271, doi:10.1016/j.epsl.2014.05.030, 2014.

Ravelo, A. C. and Andreasen, D. H.: Enhanced circulation during a warm period, Geophys. Res. Lett., 27(7), 1001–1004, doi:10.1029/1999GL007000, 2000.

Raymo, M. E., Grant, B., Horowitz, M. and Rau, G. H.: Mid-Pliocene warmth: stronger greenhouse and stronger conveyor, Mar. Micropaleontol., 27(1–4), 313–326, doi:10.1016/0377-8398(95)00048-8, 1996.

Robinson, M. M., Valdes, P. J., Haywood, A. M., Dowsett, H. J., Hill, D. J. and Jones, S. M.: Bathymetric controls on Pliocene North Atlantic and Arctic sea surface temperature and deepwater production, Palaeogeogr. Palaeoclimatol. Palaeoecol., 309(1–2), 92–97, doi:10.1016/j.palaeo.2011.01.004, 2011.

Rohde, R., A. Muller, R., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D. and Wickham, C.: A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, Geoinformatics Geostat. Overv., 01(01), doi:10.4172/2327-4581.1000101, 2013a.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C. and Mosher, S.: Berkeley Earth Temperature Averaging Process, Geoinformatics Geostat. Overv., 01(02), doi:10.4172/2327-4581.1000103, 2013b.

Sagoo, N. and Storelvmo, T.: Testing the sensitivity of past climates to the indirect effects of dust, Geophys. Res. Lett., 44(11), 5807–5817, doi: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 quantifying Pliocene terrestrial warming revealed by data–model discord, Nat. Clim. Change, 3(11), 969–974, doi:10.1038/nclimate2008, 2013.

Samakinwa, E., Stepanek, C. and Lohmann, G.: Sensitivity of mid-Pliocene climate to changes in orbital forcing, and PlioMIP's boundary conditions, Clim. Past Discuss., 1–35, doi:https://doi.org/10.5194/cp-2020-5, 2020.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, É., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A. and Yiou, P.: Using palaeo-climate comparisons to constrain future projections in CMIP5, Clim. Past, 10(1), 221–250, doi:10.5194/cp-10-221-2014, 2014.

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Glob. Planet. Change, 77(1), 85–96, doi:10.1016/j.gloplacha.2011.03.004, 2011.

Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. N. and Holland, M. M.: The emergence of surface-based Arctic amplification, The Cryosphere, 9, 2009.

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(3), 1139–1164, doi: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., 1–72, doi:https://doi.org/10.5194/cp-2020-10, 2020.

Stephenson, D. B., Pavan, V., Collins, M., Junge, M. M., Quadrelli, R. and Participating CMIP2 Modelling Groups: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Clim. Dyn., 27(4), 401–420, doi:10.1007/s00382-006-0140-x, 2006.

Tan, N., Contoux, C., Link to external site, this link will open in a new window, Ramstein, G., Sun, Y., Link to external site, this link will open in a new window, Dumas, C., Sepulchre, P., Link to external site, this link will open in a new window and Guo, Z.: Modeling a modern-like pCO2 warm period (Marine Isotope Stage KM5c) with two versions of an Institut Pierre Simon Laplace atmosphere–ocean coupled general circulation model, Clim. Past Katlenburg-Lindau, 16(1), 1–16, doi:http://dx.doi.org.ezp.sub.su.se/10.5194/cp-16-1-2020, 2020.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 365(1857), 2053–2075, doi:10.1098/rsta.2007.2076, 2007.

Thompson, D. W. J. and Wallace, J. M.: Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability, J. Clim., 13, 17, 2000.

Tierney, J. E., Haywood, A. M., Feng, R., Bhattacharya, T. and Otto-Bliesner, B. L.: Pliocene Warmth Consistent With Greenhouse Gas Forcing, Geophys. Res. Lett., 46(15), 9136–9144, doi:10.1029/2019GL083802, 2019.

Tindall, J. C. and Haywood, A. M.: Modelling the mid-Pliocene warm period using HadGEM2, Glob. Planet. Change, 186, 103110, doi:10.1016/j.gloplacha.2019.103110, 2020.

Unger, N. and Yue, X.: Strong chemistry-climate feedbacks in the Pliocene, Geophys. Res. Lett., 41(2), 527–533, doi:10.1002/2013GL058773, 2014.

Winton, M.: Sea Ice-Albedo Feedback and Nonlinear Arctic Climate Change, in Geophysical Monograph Series, edited by E. T. DeWeaver, C. M. Bitz, and L.-B. Tremblay, pp. 111–131, American Geophysical Union, Washington, D.C., 2008.

Yoshimori, M. and Suzuki, M.: The relevance of mid-Holocene Arctic warming to the future, Clim. Past, 15(4), 1375–1394, doi:10.5194/cp-15-1375-2019, 2019.

Zhang et al., Simulating the mid-Holocene, Last Interglacial and mid-Pliocenen climate with EC-Earth3-LR, submitted to GMD.

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, Clim. Past, 9(4), 1495–1504, doi: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(1), 291–305, doi:https://doi.org/10.5194/cp-15-291-2019, 2019.