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 CO_2 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 CO_2, 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.