

Dr Charles JR Williams, BA DPhil FRGS
Research Fellow
School of Geographical Sciences
University Road, Bristol, BS8 1SS

Wednesday 18 August 2021

To whom it may concern,

Re: Response letter to Reviewers 1 and 2, after submission of manuscript “*Simulation of the mid-Pliocene Warm Period using HadGEM3: Experimental design and results from model-model and model-data comparison*” by CJR Williams *et al.* to *Climate of the Past*.

I extend my sincere appreciation to both Reviewers for a thorough examination of my manuscript, and your detailed and highly constructive comments. I have addressed all of your concerns, so please see attached for a revised manuscript, still with the Track Changes included, to show my modifications.

Here, I address your suggestions, comment-by-comment. In the following, your comments are blue and in a smaller font, and my corresponding response follows in a standard font. Line numbers in the following refer to the Tracked Changes version of the manuscript.

I very much hope that my responses will meet your expectations, and satisfy you that it is now acceptable for publication.

Yours faithfully,



Dr Charles JR Williams, and co-authors

REVIEWER 1

Williams et al. described the experimental setup and model results from the HadGEM3 mid-Pliocene simulation. Modeled temperature responses are compared with proxy, previous generations of HadGEM models, and the other climate models that participated in the PlioMIP2. Consistent with the high climate sensitivity of HadGEM3, its Pliocene simulation is warmer than the other CMIP6/PMIP4 models and the previous generations of HadGEM models. The simulated Pliocene warmth is also larger than proxy reconstructions.

The paper is in general well written and easy to read. The authors did an excellent job describing details of the simulation setup. Simulation results are also presented clearly. In my view, major drawbacks of the manuscript are its short simulation length and the lack of in-depth analysis of simulation results. The simulation length is certainly limited by computing resources but the implications on results presented in the manuscript should be more clearly stated. The lack of in-depth mechanistic understanding of model results may be due to the descriptive nature of the manuscript but clearly makes the manuscript scientifically less interesting. Please see my detailed comments and suggestions below.

1. The short length of the simulation. The fully coupled simulation was carried out for a total of 576 years. At the end of the simulation, model temperatures and radiation still exhibit a large trend (Table 1). In particular, Figure S5 shows that the net energy imbalance at top-of-the-atmosphere (TOA) is close to 1.0 W m^{-2} , which is not a small number. This is an important piece of information and should be better presented in the manuscript. (1) I suggest the authors add a “Gregory-like plot” by simply plotting the TOA energy imbalance against the global mean surface temperature for the production simulation. This new figure will provide information on the equilibrium state of the production run and perhaps provide some estimates of the magnitude of the final warming if the simulation is run to equilibrium state. (2) When discussing results on model-model and model-data comparison, this caveat should be more clearly stated. For example, the global mean SAT/SST anomaly in the model and the root mean squared error from proxy data are very likely greater than values listed in the manuscript. In that sense, the performance of HadGEM3 in the simulation of mid-Pliocene could be poorer than indicated in the manuscript. (3) If possible, please also document the equilibrium state (e.g. TOA energy imbalance) of the other HadGEM model versions and take it into consideration in model-model comparison. I think it is critically important to state that this simulation is far from equilibrium in terms of the surface climate, as the simulation data has been uploaded to the ESGF and future users may easily overlook this information.

We thank the reviewer for this comment, and agree entirely that with unlimited computational resources a longer simulation would be preferable; using this version of the model, however, it is simply not feasible to run for the thousands of years required to reach full equilibrium. This point has been made more clear throughout the manuscript and in particular a new row has been added into Table 1 to include mean TOA radiation, with appropriate discussion (lines 448-451). To respond to the comments individually:

- (1) A Gregory plot has been added into the Supplementary Material (Figure S8) and is discussed in the manuscript (lines 456-466).
 - (2) The caveat of the relatively short run length has been made more clear, added into the abstract (lines 68-72), results (lines 456-466) and conclusions (lines 832-834 and 846-849).
 - (3) The equilibrium state (i.e. the TOA radiation balance) has been calculated and presented (for all of the Hadley Centre models used here) in a new table (Table S2) in the Supplementary Material, as well as appropriate discussion (also in the Supplementary Material, Section 2) and a reference to this in the main manuscript (lines 451-454).
2. The lack of in-depth analysis. The current manuscript is highly descriptive and lacks in-depth analysis to provide any mechanistic understanding of simulation results. The Abstract basically states that the HadGEM3 Pliocene simulation is warmer than the previous generation of models, the other PlioMIP2 models, and likely the proxy reconstruction. In the main text, the authors have not provided sufficient analysis to probe why the simulation is so warm and which aspects of the model simulation may be responsible for the excessive warmth. I suggest the authors add scatter plots of model ECS against the simulated Pliocene warming for both PlioMIP2 models and also the HadGEM family models. Moreover, the authors can show multi-model results on the ratio of Earth system sensitivity over the equilibrium climate sensitivity. These simple plots are not as sophisticated as a comprehensive forcing-feedback analysis but they can provide important insights on mechanisms for the warmth in the Pliocene simulations.

A scatterplot of model ECS, with ESS as a separate y-axis, against simulated Pliocene warming for all available models has been included (Fig. 16), with appropriate discussion in the manuscript (lines 771-784). As an example of this discussion, it is concluded that *“There is a clear linear relationship between ECS and global mean warming (or ESS), with the two models showing the highest ECS also having the highest Pliocene warming or ESS (HadGEM3 and CESM2). Despite some outliers... this would suggest that for most models Pliocene temperature anomalies (and ESS) are increasing in line with ECS.”*

3. The “Short title” could be improved. The current one describes “research activities” but not “findings”. I think it is better to describe the most important findings, especially when the main title is already on “research activities”. A revised one could be “HadGEM3 simulates a warmer Pliocene than proxy data and other models”.

We thank the reviewer for this suggestion, and have incorporated the suggested short title, *“HadGEM3 simulates a warmer Pliocene than proxy data and other climate models”* (line 33).

4. Line 340: Which boundary layers are you referring to, atmosphere or ocean?

This has now been clarified, to insert the word “atmospheric” (line 358).

REVIEWER 2

Williams et al describe a HadGEM3 simulation run using the PlioMIP2 protocol. They document many technical details associated with this simulation that may be important for subsequent researchers. They provide some headline results from the simulation and place it in context within the literature. This is a manuscript that is certainly worthy of publication in *Climate of the Past*. Naturally, I have some comments about the manuscript and propose some edits that would make the manuscript more appealing to readers. However, I see them as potential improvements rather than obstacles.

Content-related comments

- I was surprised that there was no mention of “hydrological sensitivity” within this manuscript. Given the focus on both the global mean precipitation change and the global mean temperature changes, this seems like an oversight. This approach in effect normalizes precip changes by the warming, resulting in units of %/oC. It is the most common metric to intercompare global mean precipitation changes.

Thank you for this comment. A new figure, in the form of a scatterplot showing normalised precipitation changes versus temperature changes (Fig. 14), has been added to the main manuscript, with appropriate accompanying discussion (lines 699-706). As an example of this discussion, it is concluded that *“... there is a clear linear relationship shown by most of the models, with Pliocene increases in precipitation increasing in line with SAT increases (Fig. 14). This relationship is not entirely linear, however, with the aforementioned result being shown again here i.e. although the HadGEM3 mPWP simulation is the 2nd warmest of all models in PlioMIP2, it is not the wettest, suggesting that although the model is highly sensitive to the Pliocene forcings in terms of its temperature response, it may be less sensitive in terms of its hydrological response.”*

- At no point is there any discussion as to whether the climate of modified control simulation is the same as the standard control run. This is potentially important for future researchers, as the manuscript only describes you changes w.r.t. piControl_mod, whilst the data available on the ESFG only allows a calculation of changes w.r.t. piControl.

We agree with this comment so thank the reviewer - the impact of the model input parameter change is discussed very briefly in Section 2.3.3, where it is stated that *“... testing of those changes for GC4 has found that they have no detectable impact on model climatology”*, but we did not provide a

comparison of the *piControl* and *piControl_mod*, other than including both of them in the timeseries in Figure 6. We have therefore added some discussion about these just after Figure 6 (lines 419-427), and have also included a new figure (Fig. S7) in the Supplementary Material in the form of spatial maps showing temperature and precipitation climatologies from both simulations. As an example of this discussion, it is concluded that “... *there is little or no difference between the two PI simulations... using temperature as an example, over the last 50 years of the simulations there is a mean of 13.79°C and 13.97°C for the piControl and piControl_mod respectively, and a standard deviation of 0.13°C for both, further confirming the negligible impact of the model parameter change in the model climatology.*”

- The discussion of the sea ice changes does not explain to the reader some key features – such as the fact that the model is seasonally ice-free in both hemispheres. Replacing Fig. 8 with something more pertinent would go a long way to improve the discussion. Please consider polar stereographic plots (perhaps without the annual mean), along with some time-series of sea ice area/extent in both simulations.

Thank you for pointing out this omission. We have replaced Figure 8 with seasonal polar stereographic plots, and added a timeseries of sea ice area, with appropriate discussion (lines 521-528).

- Remove the table which is masquerading as a panel in Fig. 14

This has been done, with the RMSE values instead being included inside brackets next to each model name (Fig. 15).

- L761. “excess warming”. You have not provided any evidence that the warming in HadGEM3 is excessive. Your focus solely on RMSE during your data-model comparison does not provide a direction. If you want to make this kind of statement (which would enhance the overall reach of the publication), then you need to look at other metrics as well.

This phrase has been changed (line 841 and 864).

Technical comments

- Sect 2.3.2.2: please include some text about how you are specifying vegetation over West Antarctica. Some stuff about this emerges in later sections, but I was expecting something here. Is there a reason you didn't specify it all as lake?

This has been changed, to add some text about how vegetation over West Antarctica was specified (lines 271-273). To respond directly to the reviewer's question: we agree that one option is to prescribe the West Antarctic ice sheet as a lake tile, but we chose not to do this as extensive lake tiles lead to potential non-conservation of the global water budget. As explained in the manuscript, instead we chose to make this region a mix of bare soil and shrubs (line 273).

- L306. Here we reach something about Antarctica, but it was unclear at this point whether you meant dominant vegetation in the piControl or mPWP. how had you determined the vegetation in mPWP?

This has been clarified (lines 317-318).

- Sect 2.3.2.4 seems to only discuss the initial conditions for the land model. I presume that the initial conditions for the atmosphere and sea ice models don't really matter. However, the initial conditions for the ocean must play a large role in the distance from equilibrium discussed later. Please be explicit in what initial conditions are being used.

This section has been clarified (lines 326-330), to explain that “*Oceanic initial conditions, such as ocean temperature and salinity, were derived from the mean equilibrium state of the piControl simulation. Some atmospheric initial conditions, such as those relating to the land surface (e.g. soil moisture and soil temperature at four levels of depth), used the same method as that applied to soil properties.*”

- Sect 3.1.2 I feel that there are criteria specified for equilibria in either the CMIP or PMIP protocols (and possibly both). It would be better to refer to those, than solely compare with a single preindustrial control simulation value.

The authors are not aware of any specific criteria listed within either the CMIP or PMIP websites, nor in any of the PMIP GMD papers; the only exception is Otto-Bliesner *et al.* (2017), where it says the simulation "... *should be long enough to minimize at least surface climate trends*". Likewise the Pliocene experimental design paper (Haywood *et al.* 2016) only says "*integration length is to be set to at least 500 years...*". To the authors knowledge, the only PMIP-related paper that specifically gives equilibrium criteria is the Eocene experimental design paper (Lunt *et al.* 2017). This is why the preindustrial control simulation value from Menary *et al.* (2018) was used. We request that this remains the same in the revised version, although we have modified this sentence to make it clearer (line 444).

- L447. Why are only the extratropical temperatures considered to make assessments about the whole hemisphere?

A sentence has been inserted into the manuscript to explain that these metrics were used to be consistent with Haywood *et al.* (2016), which we discuss in section 4.2 (lines 500-501).

- L611 What are the “warmest PI anomalies”? The manuscript shows no assessment of the piControl_mod simulation.

This has been clarified (line 674).

- L766. Do you really mean ‘climate sensitivity’ here – or should it be Earth System sensitivity?

This sentence has been rewritten (line 869).

- L770. “such as” is inappropriate as only the 3 named variables are available.

This has been removed (line 874).

- L775. I get “authors of the appropriate publication” for the models in section 4.1. However, it is not clear who readers should be contacting for output from models included in section 4.2. If this is Alan Haywood (because of H20) then please state explicitly.

This has been changed so that the readers are pointed to Haywood *et al.* (2020), where the appropriate references are given (lines 879-880).

Text/presentation comments

- Remove section headers 2.2.1 and 2.2.2. Just add the single sentence about other models to the end of the previous paragraph

These have been removed (line 152 and 178).

- L232. replace comma before JULES with a bracket.

This has been done (line 240).

- L257. preindustrial

This has been corrected (line 264).

- Fig. 4. It is hard to read the text in this image, or in fact really grasp the details of the upper 5 panels. Is there a way to condense this down, so it is more visible? Consider only showing a single PFT and increasing the panel size.

This has been done, replacing this figure with one example PFT only (Fig. 4).

- L315/316. Such as is repeated.

As part of rewriting this section according to the comment above, this error has been removed (lines 326-329).

- Fig. 5. It is nigh-on-impossible to read the text in this figure when it is printed out. Please consider whether all panels are necessary. I believe the snow depth at least should be removed as the values shown may not have much physical meaning (1000 snow-water-equivalent in kgm-2 is roughly equal to 3m depth of fresh-ish snow).

This figure has been changed, removing the soil dust properties and snow depth and making the others (especially the colour bars) larger and easier to read (Fig. 5).

- L359. “the thousands of years ideally needed” -> “thousands of years”. Let the reader judge whether this is appropriate.

This sentence has been changed (lines 376-377).

- L361/362. Please confirm that there is no typo here with the number of years. They both contain the same digits, but in a different order.

There is no typo here with the number of years: 576 is the total number of our simulation, with 50 years being the final climatology therefore leaving 526 years as the spin-up; in contrast, 652 years corresponds to the spin-up used in the *piControl* (lines 379-380).

- Table 1. Some, if not all, the units stated in this table are incorrect. They should include “/century”.

This error has been corrected (Table 1).

- Table 1. Please revise the caption of this table. Firstly, it does not show “trends in ... measures of climate equilibrium”. Secondly, please write your TOA statement as a full sentence, and be more explicit about its relation to warming. For example, “A positive TOA imbalance indicates a net loss of energy from the Earth System”.

This has been corrected (Table 1) and the suggested statement has been used in the caption (lines 436-439).

- L425. Little is gained by abbreviating OceTemp and OceSal? Please remove.

These abbreviations have been removed (lines 469-473).

- L439-468. This paragraph is both long and dense. Consider splitting the sea ice discussion into its paragraph.

This paragraph has been separated (lines 514-528).

- L444-449. This sentence is rather long. Consider subdividing.

This sentence has been divided (line 499).

- L477-482. This sentence is rather long with many subclauses. Consider subdividing.

This sentence has been divided (line 544).

- L480. The sudden reappearance of H₂O confused me (in part as it looks like the chemical symbol for water). I suggest removing the abbreviation.

This abbreviation has been removed (line 551).

- L517. Can you move “for comparative purposes” to the end of the sentence to improve readability?

This has been moved (line 581).

- L522. Remove hanging data availability sentence.

This has been removed (line 585).

- Fig. 11 Do you mean “annual mean” instead of “climatology”?

This has been corrected (line 590).

- L548 and onwards. Consider replacing your PA abbreviation with just “amplification” if polar amplification is really too long.

This change has been made throughout the manuscript, but especially lines 728-747.

- L550. “from” seems the wrong preposition

This has been changed (line 614).

- Table 4. Can you put the ECS into this table as well, please?

ECS values have been added (Table 4).

- Fig 12. Label all the panels, not just over half them.

All the panels have now been labelled (Fig. 12).

- L608. Clause about being at top-end of range felt overly repetitive.

This sentence has been changed (lines 671-672).

- L627. Be consistent in your terminology. In the previous sentence “wettest” related to absolute rainfall. Here it appears to relate to a change in rainfall.

This sentence has been changed (line 690).

- L653-4. Please rephrase this sentence.

This sentence has been rephrased (lines 728-729).

- L663-671. This additional discussion got actually me more confused. Is it possible to rephrase it?

This paragraph has been rewritten (lines 745-747).

- L676. The RMSE is given in Table 3, please cite it.

Table 3 has been cited (line 754).

- L676-686. Personally, I would chop everything after “halfway amongst them”. The subsequent description provides list insight into the reason.

This latter half of the sentence has been removed (line 754-756).