# **Reply to the reviewers' comments: Large-scale features and evaluation of the PMIP4-CMIP6 midHolocene simulations** (cp-2019-168)

C. M. Brierley *et al.* **Correspondence:** c.brierley@ucl.ac.uk

#### Summary of changes

We have adopted many of the revisions to the text suggested to the reviewers. Two of the reviewers commented on the use of the PaleoCalAdjust software, and we shall rewrite that sub-section to provide better explanation and justification. We intend to work further on the text surrounding the data-model analysis in the light of Reviewer 2's comments (Section 3.5). We shall

5 make the improvements to the figures as requested. Further simulations have since become available from at least 3 models and we plan to incorporate them into our analysis. Finally, We have devoted more effort to applying best practice for CMIP6 data citation and documentation within the manuscript.

Blue text below is our response to the reviewer's comments (reproduced in black).

#### 10 Reviewer 1

In this manuscript, the authors have evaluated the PMIP4-CMIP6 simulations along with PMIP3-CMIP5 simulations. They found that there is no significant difference in the simulated climate between these two sets of simulations. The manuscript is well written and I would like to suggest this manuscript to be accepted with some minor revision.

We would like to thank the reviewer for their kind comments and are happy to make the revisions suggested

#### 15

page 2, line 8, change 'cooler that' to 'cooler than'. Done

page 3, figure 3 caption, is 1981-1900 correct? or it should be 1900-1981 or some other time periods.20 This should have been 1871-1900, and has now been corrected.

For figure 2, it might be a good idea to add two panels to represent the difference between PMIP4 and PMIP3 as shown in Figure 6.

We did initially include these panels during the paper drafts. However, the ensemble differences are pretty consistent in both

25 seasons - and they both look similar to Fig 1e. We now explicitly state "The pattern of cooling in both seasons is very similar to the annual mean ensemble difference in Fig. 1e (not shown)"

It might be a good idea to test the significance for Figure 6e,f instead of just use Figure 10 to show the multi-variables test. Given that the more-sophisticated multi-variate analysis demonstrates that the PMIP3 & PMIP4 ensembles can be treated as

30 a single super-ensemble, we feel such analysis risks giving a false positive. However, we now shout forward to the formal significance testing at the relevant location: "(testing the significance of the differences between the ensembles is discussed in sec. 3.5)"

For the standard deviation in Figures 2 and 6, it is not clear whether it is calculated from each individual model first, then averaged across models or it is the across model standard deviation.

The standard deviation presented is the across-model standard deviation and is an attempt to show the intermodel spread. We have specified this at several places in the revised text - including the figure captions.

Also, it seems that the standard deviation is not discussed in any detail in the manuscript. If so, those panels could be removed 40 from Figures 2 and 6.

Whilst we have not explicitly mentioned these fields, they are often alluded to in our discussion of what changes are 'consistent' or 'robust'. We feel they are important to provide to future researchers using the ensemble, so hope to keep these figure panels.

#### **Reviewer 2**

35

45 This manuscript presents the very recent simulations performed within the PMIP4-CMIP6 project and aiming at reproducing the climate of the mi-Holocene (6000 years ago). This climate was characterized by specific external forcing (greenhouse gases concentrations) and orbital parameters, which have been prescribed in a consistent way in the different models, following a previous study (Otto-Bliesner et al. 2017). At the time of the publication, twelve models have performed these simulations and are analyzed here, which is equivalent to the number of models having performed the previous coordinated exercise. The

50 manuscript proposes a first analysis of four specific features of the midHolocene climate: surface temperature, monsoons, extratropical hydrological responses and the ocean circulation. A final section discusses general assessments of the midHolocene climate.

The paper is in general very well written and presented, and figures are of high quality. References are thoroughly cited. This paper is clearly an important milestone for the PMIP4-CMIP6 intercomparison process, benchmarking the publication of

55 the simulations and probably calling for subsequent more original analysis. I only have a few major comments and a series of minor ones listed below. I want to indicate to the authors that I am not a paleo-expert, I come from the climate modeling community, typically rather focusing and the historical and future periods.

We thank you for your diligent review and hope that the revisions we've made resolve your concerns.

#### 60 Main comments

On the calendar adjustment (section 2.2), I am surprised that this calendar adjustment has to be performed offline. Is there a fundamental reason that the protocol does not include a change of the online calendar? It is now obvious that our discussion of the calendar impact was unclear, so we have rephrased several sentences in this subsection. The insolation changes resulting from the altered orbital configuration are a key part of the experiment protocol. The problem here is to do with the aggregating

65 of output data during run time up to monthly resolution. To fix this online can require substantial modification of a model's output processing code, which would act as hurdle to participation.

Could you provide a comparison of the effect of the calendar adjustment for a variable for which it is proven useful (temperature?). Here, you only give numbers for the monsoon system, which you claim is not suited for this adjustment.

- 70 This form of calendar adjustment has never previously been implemented in a multi-model study, despite several calls for it (e.g. Kutzbach and Gallimore, 1988; Joussaume and Braconnot, 1997; Bartlein and Shafer, 2019). It is generally considered a minor error. The creation of easy-to-use software by Bartlein and Shafer (2019) has meant it has been possible to include it for the first time here. Justifications and use cases are provided by Bartlein and Shafer (2019).
- 75 It sounds illogical to me to assess (in this case to reject) the reliability of the calendar adjustment for the monsoon system using a fixed domain (lines 95-98) and just after (lines 99 and following) explain that the spatial extent of the monsoon varies so that the latter has to be adjusted for future assessments. Could you justify or modify? Thank you for pointing out the apparent contradiction raised by this juxtaposition. We have altered the order of these two
- subsections. Unfortunately the number of gridboxes meeting the criteria to be considered within the *midHolocene* monsoon
  domain is different between the original and calendar-adjusted monthly resolution output. And neither are completely identical to the number of gridboxes meeting the criteria when using the daily-resolution data. We therefore had little option but to only consider include in our average gridboxes that are within the domain in all three instances.

#### Section 3.5 is relatively confusing.

85 We agree that this section was not sufficiently clear and have rewritten it. We have substantially revised the first paragraph in this subsection, which hopefully makes its clearer.

L. 304-307: I have to say that I don't understand the point that is made here. First, grammatically, I don't understand the "they" 1. 305. And logically, I don't understand the link of what is said here with the beginning of the paragraph.

#### 90 We hope that the substantially revised and expanded paragraphs more understandable.

L. 309 "noisy" would be more accurate than "chaotic" I think.

We agree that chaotic has a more specific meaning in the context of climate. We think it is better to describe the distribution of significant points as "random" and we have modified the sentence accordingly. Note we have modified the paragraph slightly

95 to remove the implied double negative, and now say that "There are hardly any locations that exceed the false discovery rate"

Paragraph beginning 1. 320: I find this paragraph difficult to follow and perhaps not so well organized.

We have rewritten this and split it into two paragraphs, one dealing with the change in sensitivity between model generations, and one exploring the possible link between sensitivity and seasonality. Please see specific comments below.

100

105

L. 322-323: it is not clear to me how Fig. 1 shows that the protocol is responsible for the detected differences.

Our argument rests on two pieces of evidence. The first is that although some models show higher sensitivity, there is no real difference in the range of sensitivities shown by the two ensembles. The second point is that the change in GHG forcing, and specifically the lowering of  $CO_2$  by 20 ppm in the CMIP6/PMIP4 simulations is consistent with the observed cooling. There is a strong correlation between the simulated cooling shown in panel (e) and the implied change due to this change in forcing

shown in panel (f). We now are more explicit about these arguments in the text.

The fact that "there is no inherent relationship between climate sensitivity and seasonality is repeated twice (l. 324 and l. 326) with no obvious demonstration in-between.

- 110 Our intention here was to state that there is no reason to expect a relationship over the oceans but that we might expect a relationship over land because of the feedbacks. Our, as yet limited analyses, do not support the idea of a relationship over land but we do not want to rule out the possibility that such a relationship might be found and we do want to encourage people to investigate this. We have rewritten this paragraph to make the argument clearer.
- 115 It could/should be said more explicitly that Fig. 11 is just an example and that checks have been performed for other features of the miHolocene climate features (have they?)

We now state that this is an example situation, although a telling one. We have performed similar checks on the other features listed in the Tab. S1 without showing significant correlations. However, this is only a limited subset of regions that have typically been used for data-model comparison, so it is still possible that such a relationship exists. We have rewritten the

120 paragraph to make it clearer.

I don't fully understand the last sentence of the paragraph: in what sense could it provide a constrain then? We want to say that our analysis of a limited number of examples should not dissuade future work on the role of climate sensitivity in the mid-Holocene simulations. We have rephrased the sentence.

125

#### Minor comments

1. 129: I don't know what recrine (212) is, please explain.

This was a typo for marine and we have corrected this in the revised text. Our intention here is to indicate the number of marine versus terrestrial records available

130

lines 35-38: the increased climate sensitivity in the CMIP6 generation of climate models is still under investigation and I think it is worth specifying: add "The reasons for this increase is still under investigation, but it may influence also the sensitivity of the models to the midHolocene external forcing."

We agree that it is worth adding something on this, but we have additionally clarified that some models have increased sensitivity and some decreased sensitivity.

End of section 2.1: on the specificities of the protocol and the models differences, I would like to advertise the site https://esdoc.org/ that provides a detailed documentation of all these points, with a full documentation of each model.

A shout out to ES-DOC has now been inserted. We have also included a new supplementary table (S1) to provide the doi for each of the simulations (as recommended by CMIP/ESGF).

L. 138: as 'as compared to piControl conditions' after 'Europe' and specify Fig. 1a. We have made this clarification.

L. 155-156: From Fig. 1, I am not completely sure that I can state that "the change in MAT with respect to the piControl in the PIMP4-CMIP6 ensemble is less than in the PMIP3-CMIP5: I see indeed a weaker warming at high latitudes but a stronger cooling in the tropics. Please correct or clarify.

We had intended this sentence to incorporate the direction, rather than just magnitude of the changes. We have altered 'less' to 'generally cooler' to remove this ambiguity.

150

L. 166: I would suggest naming explicitly the C20 reanalysis in the brackets. This has been done

L. 170-171: as shown in Fig. 1(e), the PIMP4-CMIP6 models are generally colder than the PMIP3-CMIP5 ones. The reduced
warm bias along the EBUS could simply be a consequence of this. Thorough analysis of the evolution of this bias would require the use of relative temperature (e.g. Hourdin et al. 2015, GRL).

We agree about this observation about the piControl biases. We have amended the sentence to point out it that it doesn't

necessarily mean the models are better. We note that Fig 1e shows the difference in the midHolocene signal though, so actually the evidence behind this statement is not presented anywhere in the manuscript. It will be available in IPCC AR6.

160

L. 175: I don't understand how this correlation coefficient is computed.

It was a correlation coefficient between the Arctic dots in panels A and B. In fact, we could make our point just as easily without such quantification. The sentence now reads: "There is no simple relationship between a model's representation of the preindustrial temperature and the magnitude of its simulated mid-Holocene temperature response (Fig. 4)"

165

L. 190: I would rather use the term "intermodel spread", as in several other places of the manuscript. Standard deviation typically relates to temporality in my view. Changes in the legend of Fig. 1, 2 and 6 are also needed. This is a useful suggestion and we have implemented it throughout the manuscript.

L. 233-235: the term "the changes of precipitation" is sometimes confusing: I think you speak of the changes of each PMIP exercise with respect to piControl, but one could also read as the changes from PMIP3 to PMIP4. I suggest clarifying. We have rephrased two sentences here. They now read "However, there is little relationship between the *piControl* precipitation biases and the simulated midHolocene changes in precipitation (Fig. S1). The variations in the midHolocene rainfall signal appear to be more related to monsoon dynamics rather than orbitally-induced local temperature variations."

175

L. 239: remove the word "indeed"

It has been removed.

L. 250 this sentence does not mean much to me. Large spread with respect to what?

180 We have expanded this phrase. It now reads "There are large differences in the simulated change in mid-Holocene precipitation between different models, as shown by the standard deviation around the ensemble mean, in both the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles (Fig. 6 & 8). Unsurprisingly, the largest differences between models occurs where the simulated change in precipitation is also largest (Fig. 6)."

185 L. 277 I would rather write "a decline of up to  $\sim 15\%$ " This has been implemented.

L. 280: It should be made clear that the magnitude of the AMOC computed at 50°N in z-coordinates is a little bit misleading, since in fact intense transport of water mass occurs at similar depths, but with very different densities (e.g. Zhang GRL 2010). We recognise this possible confounding factor, but feel such a technical caveat is not helpful in this situation. Our purpose was

190

to not rely only on 30°N - which seemed justified given Reviewer's 3 comments about its specification. Instead we have added a sentence acknowledging potential issues, but without specificyig what they are in detail. The manuscript now reads "Using a single metric to categorise AMOC is awkward – that two measures, both with their own foibles, support show the same result 195 L. 286: I don't clearly understand the link that is made with the magnitude of low frequency internal variability in AMOC. Clarify.

We had meant to suggest that the changes in the AMOC shown may be within the bounds of natural variability. As the sentence caused confusion and did not add much to the narrative, it has been removed.

200 Conclusion section will have to be changed accordingly to previous remarks (on section 3.5 in particular) We have revised several sentences in the conclusions section. As neither the methodological revisions nor the increase in ensemble size fundamentally revised our findings, these are relatively minor.

L. 365: I have no proof at this stage that the changes in the implementation of aerosols in CMIP6 is an improvement. Wedid not either, and this sentence was written in anticipation that some relevant work would be submitted at the same time as this manuscript. We have changed 'improvement' to 'advances', which does not specify that CMIP6 is better than CMIP5 so forthrightly. We have also weakened the phrasing around the aerosols.

#### **Reviewer 3**

210 The manuscript presents the recent simulations of mid-Holocene performed within the PMIP4-CMIP6. It is a valuable and interesting work as it will be one of the main references for the future studies. The text is written and structured well, and the storyline is what would be expected from such a paper but I have few main critical and some major/minor comments as are listed below. Overall, I think it should be published in Climate of the Past, after a minor-moderate revision.

We thank the reviewer for their kind comments, and hope that the revisions we have implemented are sufficient to satisfy their expectations.

#### Main comments

As one of the important components of the climate, I wonder why the sea ice story was not included except the short sentence at line 193. I suggest to add the 2-D maps of sea ice concentration in the Arctic (for both summer and winter) which would also be relevant to the AMOC story.

```
also be relevant to the AMOC story.
```

There are two main reasons that we did not include sea ice in this manuscript. Firstly, we wanted to constrain the scope of this manuscript to a manageable amount of analyses as sea ice analyses can easily become stand alone manuscripts (e.g. Berger et al., 2013; Kageyama et al., 2020). Secondly, there were technical issues around the calendar adjustment using the Bartlein and Shafer (2019) software. It had been developed and evaluated on surface temperature and precipitation. The software

225 development required for it to adjust fields on rotated grids has been attempted, but not been scientifically validated for sea ice coverage. Using one of the three of the models that have provided daily fields, we have performed that.

We have now performed a substantial amount of analysis into the sea ice coverage changes at the midHolocene, which have culminated in the addition of a new figure. Despite our efforts, we have not been able to find a visualisation method using the 2-D maps suggested by the reviewer that adds to narrative. We therefore include two summary scatter plots, inspired by Berger et al. (2013), that try to provide a concise summary of the results.

230

240

For instance, when there is sea ice covering part of the deep water formation (DWF) region in the Labrador Sea (due to model bias), DWF is reduced. If the model has too much sea ice over the Labrador Sea in both PI and midHolocene, then that can partially explain why the AMOC does not differ much between the two periods, as any freshening or cooling cannot influence the DWF.

We feel that this particular aspect needs to be addressed first by authors evaluating the performance of the sea ice models against observations in the historical simulations. Several works may have been submitted on the topic, but there is disappointingly little openly accessibly at the moment for us to base our derived analysis upon. We do now highlight PMIP3 previous work on the topic (Găinuşă-Bogdan et al., 2020) in a hope to inspire future researchers.

With regards to AMOC, it would be nice to say something (few sentences) about the regions where deep water formation happens based on mixed layer depth values in March. It could be that if there is sea ice over the Labrador (recalling my previous comment), then the region of deep water formation might shift.

245 Detailed, subsequent analysis of the mid-Holocene AMOC and deep water formation has been initiated, but is the focus of a future paper.

Again about AMOC, I know that normally 30° N (models) or 26.5° N (observations) are chosen for calculating the maximum AMOC value. Since this latitude of maximum AMOC can differ between models as well as the two periods of study, I would

not pick one latitudinal point. Instead, I will define a range e.g., 25-35° N to calculate the maximum AMOC.
 We understand the reviewers position, but would prefer to stick with the definition adopted by the IPCC'S 6th Assessment Report.

I also suggest to add the observations to your plot (RAPID-MOCHA array observations, Smeed et al., 2017).

255 This is a sensible suggestion. We have now added the RAPID-MOCHA array to compare with 30°N, and also the OSNAP array to provide a more northerly benchmark for the 50°N metric.

#### Specific comments regarding the text and analysis

Methods: The mean values, are they averaged over the entire simulated years mentioned in Table 1?

260 Yes. We have now added a footnote to the table to clarify this.

Line 86: Can you say in one sentence how is PaleoCalAdjust performing in general?

Bartlein and Shafer (2019) provide convincing evidence that it is performing well on PMIP3 mid-Holocene simulations for monthly surface temperature and precipitation. We have now mentioned this in the revised text through an additional sentence: "This software was developed and been favorably evaluated for monthly temperature and precipitation variations with both

PMIP3-CMIP5 and transient simulations (Bartlein and Shafer, 2019)."

Lines 88-98: For the annual mean I understand you do not need calendar adjustment. But if you use your daily values from PaleoCalAdjust and make the annual mean, how much would it differ from the main annual mean? This can give you some

270 ideas about the potential interpolation errors (if there is no original daily data). Thank you for your suggestion. A subset of the models have actually provided simulation output at a daily resolution, so this step is not necessary. We provided an example of this kind of analysis for Arctic sea ice extent as a supplement in our earlier Author Comment. We do not feel this manuscript is the correct place to provide additional evaluation of the methodlogical technique. Your comments provide further motivation for a further detail analysis to be published.

275

265

Line 98: I do not understand "we have therefore..." so you use the method when you think it is good?

Fundamentally, yes. There is a balance to navigate between the original size of the error due to the calendar misalignment and errors introduced by PaleoCalAdjust from the interpolation step. The analysis in the previous few sentences demonstrates that the interpolation errors are greater than the misalignment, so there is little advantage using PaleoCalAdjust for this diagnostic.

280

Line 107: in this line and any other lines (line 219) please change "interannual variability" to internal climate variability because the variability is not only interannual.

This new terminology has been adopted in the text. It is however retained in the caption, because it concisely conveys the message that the standard deviations is measured across a time series that is only resolved annually.

285

Lines 154-161: move these lines to after line 140.

This restructuring of the paragraphs has been adopted, although the one sentence about seasonal changes is moved instead to the end of the paragraph describing Fig. 2.

290 Line 171: "...colder conditions over the Labrador current..." which figure you are referring to? And I assume you meant Labrador Sea and not current?

#### We have corrected 'current' to 'Sea' and added a reference to Fig. 3b.

Line 174: not only in the tropics but over the oceans in general there is a better match

295 We have now added this insight.

Line 234: "... change in precipitation" change between what? This has actually been altered to 'rainfall changes', to stop the paragraph feeling too repetitive (as per next comment).

300 Line 234-236: you used "change" three times in one sentence, modify please and combine it with the previous sentence. This sentence has now been modified to only include 1 'change'. We have not combined with the previous sentence though, as it's already rather long.

Line 324: role

305 We have implemented this change.

Line 374: "... need for improved physics and processes..." This change has been implemented.

Figure 3 caption: check the years where observation was used e.g., you wrote "1981-1900". Also check the rest of caption. This should have been 1871-1900, and has now been corrected.

Figure 5: would it be possible to make the similar figure for the observation/reanalysis?

It is possible, and was included in the supplement to our earlier Author Comment. However, given that the boundary of the 315 domain in the observations/reanalysis is marked both panels already, we feel it is superfluous to include an additional panel in the manuscript.

#### References

Bartlein, P. J. and Shafer, S. L.: Paleo calendar-effect adjustments in time-slice and transient climate-model simulations (PaleoCalAdjust v1.

- 320 0): impact and strategies for data analysis, Geoscientific Model Development, 12, 3889–3913, https://doi.org/10.5194/gmd-12-3889-2019, 2019.
  - Berger, M., Brandefelt, J., and Nilsson, J.: The sensitivity of the Arctic sea ice to orbitally induced insolation changes: a study of the mid-Holocene Paleoclimate Modelling Intercomparison Project 2 and 3 simulations, Climate of the Past, 9, 969–982. https://doi.org/10.5194/cp-9-969-2013, 2013.
- 325 Găinusă-Bogdan, A., Swingedouw, D., Yiou, P., Cattiaux, J., Codron, F., and Michel, S.: AMOC and summer sea ice as key drivers of the spread in mid-Holocene winter temperature patterns over Europe in PMIP3 models, Global and Planetary Change, 184, 103 055, 2020.
  - Joussaume, S. and Braconnot, P.: Sensitivity of paleoclimate simulation results to season definitions, Journal of Geophysical Research: Atmospheres, 102, 1943–1956, https://doi.org/10.1029/96JD01989, 1997.
  - Kageyama, M., Sime, L. C., Sicard, M., Guarino, M.-V., de Vernal, A., Schroeder, D., Stein, R., Malmierca-Vallet, I., Abe-Ouchi, A., Bitz, C.,
- 330 Braconnot, P., Brady, E., Chamberlain, M. A., Feltham, D., Guo, C., Lohmann, G., Meissner, K., Menviel, L., Morozova, P., Nisancioglu, K. H., Otto-Bliesner, B., O'ishi, R., Sherriff-Tadano, S., Stroeve, J., Shi, X., Sun, B., Volodin, E., Yeung, N., Zhang, Q., Zhang, Z., and Ziehn, T.: A multi-model CMIP6 study of Arctic sea ice at 127 ka: Sea ice data compilation and model differences, Climate of the Past Discussions, 2020, 1-33, https://doi.org/10.5194/cp-2019-165, https://www.clim-past-discuss.net/cp-2019-165/, 2020.

Kutzbach, J. and Gallimore, R.: Sensitivity of a coupled atmosphere/mixed layer ocean model to changes in orbital forcing at 9000 years BP, Journal of Geophysical Research: Atmospheres, 93, 803–821, 1988.

335

## Large-scale features and evaluation of the PMIP4-CMIP6 *midHolocene* simulations

Chris M. Brierley<sup>1</sup>, Anni Zhao<sup>1</sup>, Sandy P. Harrison<sup>2</sup>, Pascale Braconnot<sup>3</sup>, Charles J. R. Williams<sup>4,5</sup>, David J. R. Thornalley<sup>1</sup>, Xiaoxu Shi<sup>6</sup>, Jean-Yves Peterschmitt<sup>3</sup>, Rumi Ohgaito<sup>7</sup>, Darrell S. Kaufman<sup>8</sup>, Masa Kageyama<sup>3</sup>, Julia C. Hargreaves<sup>9</sup>, Michael P. Erb<sup>8</sup>, Julien Emile-Geay<sup>10</sup>, Roberta D'Agostino<sup>11</sup>, Deepak Chandan<sup>12</sup>, Matthieu Carré<sup>13,14</sup>, Partrick J. Bartlein<sup>15</sup>, Weipeng Zheng<sup>16</sup>, Zhongshi Zhang<sup>17</sup>, Qiong Zhang<sup>18</sup>, Hu Yang<sup>6</sup>, Evgeny M. Volodin<sup>19</sup>, Robert A. Tomas<sup>20</sup>, Cody Routson<sup>8</sup>, W. Richard Peltier<sup>12</sup>, Bette Otto-Bliesner<sup>20</sup>, Polina A. Morozova<sup>21</sup>, Nicholas P. McKay<sup>8</sup>, Gerrit Lohmann<sup>6</sup>, Allegra N. Legrande<sup>22</sup>, Chuncheng Guo<sup>17</sup>, Jian Cao<sup>23</sup>, Esther Brady<sup>20</sup>, James D. Annan<sup>9</sup>, and Ayako Abe-Ouchi<sup>7,24</sup>

<sup>1</sup>Department of Geography, University College London, London, WC1E 6BT, UK

<sup>2</sup>Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AB, UK

<sup>3</sup>Laboratoire des Sciences du Climat et de l'Environnement-IPSL, Unité Mixte CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, Gif-sur-Yvette, France

<sup>4</sup>Department of Meteorology, University of Reading, Reading, RG6 6BB, UK

<sup>5</sup>School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK

<sup>6</sup>Alfred-Wegener-Institut Helmholtz-Zentrum f ur Polar- und Meeresforschung, Bremerhaven, Germany

<sup>7</sup>Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

<sup>8</sup>School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA.

<sup>9</sup>Blue Skies Research Ltd, The Old Chapel, Albert Hill, Settle, BD24 9HE, UK

<sup>10</sup>Department of Earth Sciences, University of Southern California, Los Angeles, California, USA

<sup>11</sup>Max Planck Institute for Meteorology, Hamburg, Germany

<sup>12</sup>Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario, M5S1A7, Canada

<sup>13</sup>LOCEAN Laboratory, Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Paris, France

<sup>14</sup>CIDIS-LID-Facultad de Ciencias y Filosofía-Universidad Peruana Cayetano Heredia, Lima, Peru

<sup>15</sup>Department of Geography, University of Oregon, Eugene, OR 97403, USA

<sup>16</sup>LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

<sup>17</sup>NORCE Norwegian Research Centre, Bjerknes Center for Climate Research, Bergen, Norway

<sup>18</sup>Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, 10691, Stockholm, Sweden

<sup>19</sup>Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119333, Russia

<sup>20</sup>Climate and Global Dynamics Laboratory, National Center for Atmospheric Research (NCAR), Boulder, CO 80305, USA

<sup>21</sup>Institute of Geography, Russian Academy of Sciences, Staromonetny L. 29, Moscow, 119017, Russia

<sup>22</sup>NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

<sup>23</sup>School of Atmospheric Sciences, Nanjing University of Information Science & Technology Nanjing, 210044, China

<sup>24</sup>Atmospheric and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

Correspondence: c.brierley@ucl.ac.uk

Abstract. The mid-Holocene (6,000 years ago) is a standard time period for the evaluation of the simulated response of global climate models using paleoclimate reconstructions. The latest mid-Holocene simulations are a paleoclimate entry card for the Palaeoclimate Model Intercomparison Project (PMIP4) component of the current phase of the Coupled Model Intercomparison

Project (CMIP6). Here we provide an initial analysis and evaluation of the results of the experiment for the mid-Holocene. We

- 5 show that state-of-the-art models produce climate changes that are broadly consistent with theory and observations, including increased summer warming of the northern hemisphere and associated shifts in tropical rainfall. Many features of the PMIP4-CMIP6 simulations were present in the previous generation (PMIP3-CMIP5) of simulations. The PMIP4-CMIP6 ensemble for the mid-Holocene has a global mean temperature change of -0.3 K, which is -0.2 K cooler than the PMIP3-CMIP5 simulations predominantly as a result of the prescription of realistic greenhouse gas concentrations in PMIP4-CMIP6. Biases in the
- 10 magnitude and the sign of regional responses identified in PMIP3-CMIP5, such as the amplification of the northern African monsoon, precipitation changes over Europe and simulated aridity in mid-Eurasia, are still present in the PMIP4-CMIP6 simulations. Despite these issues, PMIP4-CMIP6 and the mid-Holocene provide an opportunity both for quantitative evaluation and derivation of emergent constraints on the hydrological cycle, feedback strength and potentially climate sensitivity.

#### 1 Introduction

- 15 Future climate changes pose a major challenge for Human civilisation, yet uncertainty remains about the nature of those changes. This arises from societal decisions about future emissions, internal variability, and also uncertainty stemming from differences between the models used to make the projections (Hawkins and Sutton, 2011; Collins et al., 2013). Coupled general circulation models (GCMs) can be used to simulate past changes in climate as well as those of the future. Palaeoclimate simulations allow us to test the theoretical response of such models to various external forcings and provide an independent
- 20 evaluation of them. The Coupled Model Intercomparison Project (CMIP; Eyring et al., 2016), which coordinates efforts to compare climate model simulations, includes simulations designed to test model performance under past climate regimes. Evaluation of these palaeoclimate simulations against palaeoclimate reconstructions, coordinated through the Palaeoclimate Modelling Intercomparion Project (PMIP; Kageyama et al., 2018), provides an independent test of the ability of state-of-the-art models to simulate climate change.
- The mid-Holocene (6000 years ago, 6ka) is one of the palaeoclimate simulations included in the current phase of CMIP (PMIP4-CMIP6; Otto-Bliesner et al., 2017). This period is characterised by an altered seasonal and latitudinal distribution of incoming solar radiation, because of larger obliquity and orbital precession, meaning that the Earth was closest to the Sun in austral spring boreal autumn (rather than in austral summer boreal winter as today) and that the northern latitudes received more solar radiation than today. The mid-Holocene has been a baseline experiment for PMIP since its inception (Joussaume et al.,
- 30 1999; Braconnot et al., 2007, 2012). As such, it has been a focus for synthesis of palaeoenvironmental data (see summary in Harrison et al., 2016) and for the reconstruction of palaeoclimate variables from these data (e.g. Kohfeld and Harrison, 2000; Bartlein et al., 2011), facilitating systematic model evaluation (e.g. Hargreaves et al., 2013; Jiang et al., 2013; Prado et al., 2013; Harrison et al., 2014; Mauri et al., 2014; Perez-Sanz et al., 2014; Harrison et al., 2015; Bartlein et al., 2017).
- The PMIP4-CMIP6 simulations differ from previous palaeoclimate simulations in two ways. Firstly, they represent a new generation of climate models with greater complexity, improved parameterisations and often run at higher resolution. Changes to the model configuration have, in some cases (e.g. CCSM4/CESM2, HadGEM2/HadGEM3, IPSL-CM5A/IPSL-CM6A),

resulted in substantially higher climate sensitivity than the previous PMIP3-CMIP5 version of the same model, although this is not a feature of all of the models (Tab. 1,2). The reasons for this change are still under investigation, but it may also influence the model sensitivity to the mid-Holocene external forcing. Secondly, the protocol for the PMIP4-CMIP6 mid-

- 40 Holocene experiment (called *midHolocene* on the Earth System Grid Federation, and henceforth herein) simulations was designed to better represent the observed conditions than previous forcings better than in previous mid-Holocene simulations (Otto-Bliesner et al., 2017). In addition to the change in orbital configuration, which was the only change imposed in the PMIP3-CMIP5 experiments, the current experiments include a realistic specification of changes in atmospheric greenhouse gas concentrations. Because of these small changes the lower values of greenhouse gas concentrations, the PMIP4-CMIP6
- 45 simulations are expected to be slightly colder that in the previous PMIP phase than the PMIP3-CMIP5 experiments (Otto-Bliesner et al., 2017). The model configuration and all other forcings are the same as in the pre-industrial control simulation (*piControl*, 1850 CE). This means that models with dynamic vegetation in the *piControl* are run with dynamic vegetation in the *midHolocene* experiment, so the PMIP4-CMIP6 ensemble includes a mixture of simulations with prescribed or interactive vegetation. Although some of the PMIP3-CMIP5-models were run with an interactive carbon cycle, none included fully-
- 50 dynamic vegetation.

Here, we provide a preliminary analysis of the PMIP4-CMIP6 *midHolocene* simulations, focusing on surface temperature changes (sec. 3.1), hydrological changes (sec. 3.2 & 3.3) and the deep ocean circulation (sec. 3.4). We examine the impact of changes in model configuration and <u>experiment experimental</u> protocol on these simulations, specifically how far these changes improve known biases in the simulated changes. We draw on <u>a more an</u> extended set of observation-derived benchmarks to

55 evaluate these simulations. Finally we discuss the implications of this evaluation for future climate changes, for example by investigating whether the different including investigating whether changes in climate sensitivities between the CMIP6 and CMIP5 generation models has an impact on the new results simulations.

#### 2 Methods

#### 2.1 Experimental Setup and Models

- 60 The protocol and experimental design for the PMIP4-CMIP6 *midHolocene* simulations are described by Otto-Bliesner et al. (2017) and Earth System Documentation (2019). The *midHolocene* simulations are run with known orbital parameters for 6000 yr BP and atmospheric trace greenhouse gas concentrations (GHGs) derived from ice-core records (as described by Otto-Bliesner et al., 2017). Eccentricity is increased by 0.001918 in the *midHolocene* simulations relative to the *piControl*, obliquity is increased by 0.646<sup>20</sup>, and perihelion ( $\omega 180^{\circ}$ ) is changed from 100.33<sup>o</sup>-<sup>o</sup>in the *piControl* (in January) to 0.87<sup>o</sup>-<sup>o</sup>in the
- 65 midHolocene (near the boreal autumn equinox). The result of these astronomical changes is a difference in the seasonal and latitudinal distribution of top-of-atmosphere (TOA) insolation. During boreal summer, anomalies-insolation between 40-50°N are °N was 25 W/m<sup>2</sup> higher in the *midHolocene* simulations than in the *piControl* (Otto-Bliesner et al., 2017). The long-lived greenhouse gases are specified at their observed concentrations. Carbon dioxide is specified at 264.4 ppm (vs 284.3 ppm

during the pre-industrial) and methane at 597 ppb (vs 808 ppb) and N<sub>2</sub>O at 262 ppb (versus 273 ppb). These changes in GHG concentrations lead to an effective radiative forcing of  $-0.3 \text{ W/m}^2$  (Otto-Bliesner et al., 2017).

Twelve-Sixteen models (Tab. 1) have performed the PMIP4-CMIP6 *midHolocene* simulations. A similar number of models have performed the equivalent PMIP3-CMIP5 *midHolocene* simulation (Tab. 2). The PMIP4-CMIP6 simulations are either available from the Earth System Grid Federation (from which they are freely downloadable) or will be lodged there in the near future. We evaluate these simulations as part of an ensemble and <u>do not always</u> only sometimes identify individual models. Most

- 75 of the models included in the PMIP4-CMIP6 ensemble are state-of-the-art climate models, but we also include some results from models that are either lower resolution or less complex (and therefore faster). Even though all models have the same orbital parameters and trace gases in the *midholocene* experiment, there may be other differences compared to the *piControl* which differences in the specification of other boundary conditions can mean that the forcing is not exactly the sameidentical in every model. For example, the different models have different models may have slightly varying solar constants (see notes in
- 80 Table 1), reflecting choices made by the different groups for the *piControl* simulations. Similarly, the orbital parameters used by some groups for the *piControl* are the same as for the historical simulation and the trace gases are slightly different from the PMIP4-CMIP6 recommendations. Differences in the pre-industrial planetary albedo, resulting from surface albedo and clouds, may also mean the effective solar forcing is different between models (Braconnot et al., 2012). Experimental setup and spin-up procedure shall be are documented for each *midHolocene* simulation individually elsewhere (Otto-Bliesner et al., 2017).

#### 85 2.2 Calendar adjustments and analysis Analysis techniques and calendar adjustments

70

Model outputs for the Fixed monsoon domains are often used when investigating variability and future changes in monsoon rainfall (e.g. Christensen et al., 2013). However, this is not appropriate in the mid-Holocene when the monsoons were greatly extended. Following Jiang et al. (2015), we adopt the definition of Wang et al. (2011) for analysis of monsoon regions: a grid point is considered to be affected by the monsoon if the rainfall predominantly falls in the summer (MJJAS in the Northern

- 90 Hemisphere, NDJFM in the Southern Hemisphere; assessed using summer rainfall forming at least 55% of the annual total) and the average rain rate difference between summer and winter (called monsoon intensity) is at least 2 mm/day or more. The ensemble mean global domain is determined by applying both these criteria to the ensemble mean summer rainfall and monsoon intensity. We calculate annual (November-October) times series of the areal extent for 7 land-based monsoon systems (Christensen et al., 2013), as well as determining the average precipitation rate within each system. Internal climate variability
- 95 is characterised by the standard deviation of these annual time series. The integral of these values is the total monsoon rainfall for each regional monsoon.

The midHolocene simulation are currently mostly available as monthly values. The experiment involves redistributing the incoming insolation spatially and through the year (Otto-Bliesner et al., 2017). This altered orbital configuration in during the mid-Holocene resulted in a change in the Earth's transit speed through the seasonal cycle along different parts of its orbit such

100 that, when considered as angular fractions of the Earth's orbit, the month lengths differed during the mid-Holocene (Joussaume and Braconnot, 1997; Bartlein and Shafer, 2019). Northern Hemisphere winter (December, January, February, DJF) was longer and summer (June, July, August, JJA) correspondingly shorter from an insolation perspective than in the present day and the *piControl* simulation. However simulation output by CMIP6 models is restricted to modern calendars (Juckes et al., 2019). This is not an problem for annual or daily diagnostics, but summarizing model output using only the modern calendar prohibits

- 105 straightforward adjustment of the numbers of days over which the aggregation of monthly simulation output takes place. To take account of these differences in calculating monthly or seasonal variables, we use the PaleoCalAdjust software (Bartlein and Shafer, 2019), which interpolates from non-adjusted monthly averages to daily values and pseudo-daily values and then calculates the average values for adjusted months defined as angular fractions of the orbit. However, since calendar adjustment is not necessary for annual measures of climate, we use the original outputs to calculate annual variables (as This software)
- 110 was developed and been favourably evaluated for monthly temperature and precipitation variations with both PMIP3-CMIP5 and transient simulations (Bartlein and Shafer, 2019). Given the experimental protocol fixes the date of the automated scripts Phillips et al. (2014) weight each month evenlyvernal equinox as 21st March (Otto-Bliesner et al., 2017), the largest impact of the calendar adjustment occurs in September (a key month for Arctic sea ice coverage). The PaleoCalAdjust software computes adjusted monthly variables from original monthly means, a computation which could impact the accuracy of variables
- 115 that change abruptly throughout the year, rather than gradually, such as the sudden increase in precipitation in monsoon regions (Pollard and Reusch, 2002). To explore whether potential interpolation errors from PaleoCalAdjust are justified in such situations, we analysed the averaged rain rate during the monsoon season over the South American monsoon domain in the IPSL-CM6A-LR *midHolocene*, for which daily-resolution data is also provided on the Earth System Grid Federation. The Since the areal extent of South American monsoon domain varies slightly when using different temporal data, so comparisons
- 120 were only made over the common grid points we make this comparison based on the grid points that always fall within the monsoon domain to provide the most robust assessment of the impact of the change in calendar. The average monsoon rain rate from the daily-resolution data is 7.0 mm/day: compared to 6.7 mm/day from calendar-adjusted monthly data and 7.1 mm/day using monthly data without this adjustment. The average monsoon rain rate in the *piControl* is 7.5 mm/day. We have therefore not applied the calendar adjustment when analysing monsoon variables.
- 125 Although fixed monsoon domains are often used when investigating variability and future changes in monsoon rainfall (e.g. Christensen et al., 2013), this is not appropriate in the mid-Holocene when the monsoons were greatly extended. Following Jiang et al. (2015), we adopt the definition of Wang et al. (2011) for analysis of monsoon regions: a grid point is considered to be affected by the monsoon if the rainfall predominantly falls in the summer (MJJAS in the Northern Hemisphere, NDJFM in the Southern Hemisphere; assessed using summer rainfall forming at least 55% of the annual total) and the average rain rate
- 130 difference between summer and winter (called monsoon intensity) is at least 2 mm/day or more. The ensemble mean global domain is determined by applying both these criteria to the ensemble mean summer rainfall and monsoon intensity. We also calculate the areal extent of 7 land-based monsoon systems annually (November-October), as well as determining the average precipitation rate within each domain. Interannual variability is characterised by the standard deviation of these two quantities. The integral of these values is the total monsoon rainfall.
- 135 The analysis presented here mainly uses generalised evaluation software tools derived from the Climate Variability Diagnostics Package (Phillips et al., 2014), which has been adapted for palaeoclimate purposes (Brierley and Wainer, 2018). It uses the surface air temperature and precipitation rate variables ('tas' and 'pr' respectively in the ESGF controlled vocabulary; Juckes

et al., 2019), as well as several different ocean overturning mass streamfunction variables. The software and routines used to ereate the figures presented here are available to download (see *code and data availability* statement). and sea ice concentration variables.

140

#### 2.3 Palaeoclimate reconstructions and model evaluation

We provide only a preliminary quantitative evaluation of the realism of the PMIP4-CMIP6 simulations, drawing attention to obvious similarities and mismatches between the simulations and observational evidence of past climates. Some of this evidence is qualitative (e.g. changes in surface hydrology evidenced by lakes and vegetation records; Kohfeld and Harrison, 2000; Prentice et al., 200

- 145 , but we also use We concentrate our evaluation on two compilations of quantitative reconstructions from a number of sources. We use temperature reconstructions from the recent 'Temperature 12k' database (Kaufman et al., 2020). We extracted anomalies for the mid-Holocene compared to the last millennium interval  $(6.0 \pm 0.5 \text{ ka} 0.6 \pm 0.5 \text{ ka})$  for site-level comparison with the PMIP4-CMIP6 simulations. This database has 1319 time series reconstructions of temperature (mean annual, summer and winter temperature) based on a variety of different ecological, geochemical and biophysical marine (209) and terrestrial
- 150 (470) sites (Kaufman et al., 2020). Additionally, area-averaged temperature anomalies (w.r.t. 1800-1900) over 30° latitudinal bands have been generated using five different methods (Kaufman et al., in press) to yield a single composite value with confidence intervals. Bartlein et al. (2011) provide pollen-based reconstructions of land climate , including (mean annual temperature, mean temperature of the coldest month, growing season temperature(indexed by growing degree days above a baseline of 0), mean annual precipitation and an index of soil moisture (alpha, the ratio of actual to potential evaporation),
- 155 <u>although we mainly focus on mean annual temperature and precipitation here</u>. They combined the reconstructions at individual pollen sites to produce an estimate for a 2°x2°grid , (a resolution comparable with the climate models;-) reconstruction uncertainties are estimated as a pooled estimate of the standard errors of the original reconstructions for all sites in each grid cell(Bartlein et al., 2011). This data set was used to evaluate the PMIP3-CMIP5 simulations (Harrison et al., 2014) and has... There is good coverage of northern hemisphere terrestrial sites, although there are gaps in the coverage especially
- 160 for the tropics and southern hemisphere . We also use temperature reconstructions from the 'Temperature 12k' database (Kaufman et al., 2020). We extracted anomalies for the mid-Holocene compared to (Bartlein et al., 2011). The Bartlein et al. (2011) data set was extended with some speleothem and ice core temperature reconstructions and used to evaluate the PMIP3-CMIP5 simulations (Harrison et al., 2014). In this study we use the pollen-only data set from Bartlein et al. (2011) and the last millennium interval ( $6.0 \pm 0.5$ ka -  $0.6 \pm 0.5$ ka) for site-level comparison with the PMIP4-CMIP6 simulations . This database has 1332
- 165 time series reconstructions of temperature (mean annual, summer and winter temperature) based on a variety of different ecological, geochemical and biophysical recrine (212) and terrestrial (472) archives (Kaufman et al., 2020). Additionally area-averaged temperature anomalies (w.r.t. 1800-1900) over 30° latitudinal bands have been generated using five different methods (Kaufman et al., in press) to yield a single composite value with confidence intervals. Differences in multi-proxy data set (Kaufman et al., 2020) to provide a measure of the uncertainties in reconstructed climates, although differences in method-
- 170 ology and coverage preclude direct comparison between the Bartlein et al. (2011) and Kaufman et al. (2020) two data sets. We use both data sets to provide a measure of the uncertainties in reconstructed climates incorporate an additional data set

to facilitate comparisons of the northern African monsoon between the CMIP6-PMIP4 simulations and previous generations of simulations, namely water-balance estimates of the quantitative change in precipitation required to support the observed mid-Holocene vegetation change at each latitude compared to present (Joussaume et al., 1999).

#### 175 3 Simulated mid-Holocene Climates

#### 3.1 Temperature Response

As expected from the insolation forcing, the PMIP4-CMIP6 ensemble shows an increase in mean annual temperature (MAT) as compared to piControl conditions in the high northern and southern latitudes and over Europe (Fig. 1a). Yet there is a decrease in MAT elsewhere, which is especially large over northern Africa and India. The ensemble produces a global cool-

- 180 ing of -0.3°C compared to the *piControl* simulation (Tab. S1S2). The relatively small change in MAT is consistent with the fact that the *midHolocene* changes are largely driven by seasonal changes in insolation<del>, yet are of a different sign than the +0.5°Creconstruction (Kaufman et al., in press) derived from the Temperature 12k compilation (Kaufman et al., 2020). The geographic patterns of temperature changes in the PMIP4-CMIP6 ensemble are very similar to those seen in the PMIP3-CMIP5 ensemble. However, the change in MAT with respect to the *piControl* in the PMIP4-CMIP6 ensemble is generally cooler than</del>
- in the PMIP3-CMIP5 (Fig. 1). The difference in the experimental protocol between the two sets of simulations would be expected to cause a slight cooling, since the difference in GHG concentrations would result in an effective radiative forcing of  $-0.3 W/m^{-2}$  (Otto-Bliesner et al., 2017). To evaluate this, we estimate the ensemble-mean forced response (Fig. 1) based on the climate sensitivity of each model (Tab. 1) and pattern scaling. The estimated global mean pattern-scaled anomaly is  $-0.28^{\circ}$ C, similar to the difference between the two model generations (Fig. 1).
- As might be expected, the higher insolation in northern hemisphere (NH) summer results in a pronounced summer (JJA) warming, particularly over land (Fig. 2). The increase in summer temperature over land in the NH high latitudes in the ensemble mean is 1.1°C (Tab. S1Table S2). Increased NH summer insolation leads to a northward shift and intensification of the monsoons (sec. 3.2), with an acompanying accompanying JJA cooling in the monsoon-affected regions of northern Africa and and South Asia. Reduced insolation in the NH winter (DJF) results in cooling over the northern continents and this cooling
- 195 extends into the northern tropical regions, although the Arctic is warmer than in the *piControl* simulation (Fig. 2). Although the Southern Ocean shows warmer temperatures in the *midHolocene* than the *piControl* simulations in austral summer (DJF) as a result of increased obliquity, this warming does not persist into the winter to the same extent as seen in the Arctic. The damped insolation seasonality, together with the large effective heat capacity of the ocean heavily damps seasonal variations in surface air temperature in the Southern Ocean. The enhanced NH seasonality and the preponderance of land in the NH therefore results
- 200 in large cause seasonal variations of the interhemispheric temperature gradient, which translate into a small increase in favour results in a small warming of the northern hemisphere in the annual, ensemble mean.

The geographic and seasonal patterns of temperature changes in the PMIP4-CMIP6 ensemble are very similar to those seen in the PMIP3-CMIP5 ensemble. However, the change in MAT with respect to the *piControl* in the PMIP4-CMIP6 ensemble is less than in the PMIP3-CMIP5 (Fig. 1). at the expense of the southern hemisphere in the annual, ensemble mean. The

- 205 PMIP4-CMIP6 ensemble is cooler than the PMIP3-CMIP5 ensemble in both summer and winter (Fig. 2). The difference in the experimental protocol between the two sets of simulations would be expected to cause a slight cooling , since the difference in GHG concentrations would result in an effective radiative forcing of  $-0.3 W/m^{-2}$  (Otto-Bliesner et al., 2017). To evaluate this, we estimate the ensemble-mean forced response (pattern of cooling in both seasons is very similar to the annual mean ensemble difference in Fig. 1) based on the climate sensitivity of each model (Tab. 1)and pattern scaling (Brierley et al., 2019)
- 210 . The estimated global mean pattern-scaled anomaly is -0.28°C, similar to the difference between the two model generations (Fig. 1) e (not shown), further supporting the lower greenhouse gas concentrations in the experimental protocol (sect. 2.1) as the cause of the cooling.

Biases in the control simulation may influence the response to mid-Holocene forcing (Braconnot et al., 2012; Ohgaito and Abe-Ouchi, 2009; Harrison et al., 2014; Braconnot and Kageyama, 2015) and certainly affect the pattern and magnitude

- of simulated changes. There is some difficulty in diagnosing biases in the *piControl*, because there are few spatially-explicit observations for the pre-industrial <u>period</u>, especially for precipitation. We therefore evaluate these simulations using reanalysed climatological temperatures (between 1871-1900 CE; Compo et al., 2011) for the spatial pattern (Fig. 3) and zonal averages of observed temperature (Fig. 4) for the period 1850-1900 CE from the HadCRUT4 dataset (Morice et al., 2012; Ilyas et al., 2017). We compare these with the mean difference between the pre-industrial climatology of each model (i.e. the ensemble
- 220 mean bias). The PMIP4-CMIP6 models are generally cooler than the observations, most noticeably in polar regions, over land and over the NH oceans (Fig. 4). The models are still too warm over the eastern boundary upwelling currents, even though this bias has been reduced in some of the models although it remains to be seen whether this indicates improved representation of the relevant physical processes compared to PMIP3-CMIP5. The colder conditions over the Labrador Current also indicate a Sea (Fig. 3b) also indicate difficulty with resolving the regional ocean circulation features sufficiently. The polar regions are
- 225 noticeably too cold in the ensemble mean (Fig. 3& 4), though the match between the models ), but there is considerable spread between individual models (Fig. 4). There is no simple relationship between a model's representation of the preindustrial temperature and the temperature observations/reanalysis appears satisfactory in the tropics. The magnitude of the magnitude of its simulated mid-Holocene temperature response in the Arctic is not significantly correlated with the bias in the *piControl* simulation (r = -0.28, (Fig. 4). Other factors such as ice albedo and ocean temperature advection affect the affect the regional
- 230 direct and indirect response to mid-Holocene forcingin these regions, such as ice albedo and ocean temperature advection into the Arctic. PMIP4-CMIP6 also includes simulations with dynamic vegetation, for example. The associated vegetationsnow albedo feedback would tend to reduce the simulated cooling (e.g. O'ishi and Abe-Ouchi, 2011), but can introduce a larger cooling bias in the *piControl* simulation (Braconnot et al., 2019). However, changes in the treatment of aerosols in the PMIP4-CMIP6 ensemble could enhance the simulated cooling (Pausata et al., 2016; Hopcroft and Valdes, 2019).
- 235 The reconstructed zonaltemperature changes Kaufman et al. (in press) suggest that zonal, annual mean temperatures during the mid-Holocene suggest a warming at all latitudes were warmer at most latitudes (Fig. 4), with maximum warming in the Arctic, using the reconstructions in the Temperature 12k compilation (Kaufman et al., 2020). Individual records in the Bartlein et al. (2011) compilation demonstrate the heterogeneity within these estimates (Fig. 4). This feature is robust between the Bartlein et al. (2011) and Kaufman et al. (2020) reconstructions. The PMIP4-CMIP6 ensemble is equivocal about whether

- 240 the polar regions were warmer or cooler on the annual mean. Furthermore, the PMIP4-CMIP6 models show a consistent cooling in the tropics. Tropical cooling was present but less pronounced in the PMIP3-CMIP5 ensemble (Fig. 4). Tropical cooling is not consistent with the Temperature 12k area-averages (Kaufman et al., 2020). (Kaufman et al., in press) (although the Bartlein et al. (2011) compilation does not discount it, the majority of the reconstructions are solely from Africa). Interestingly comparisons over Europe and North America, both well-sampled by the Bartlein et al. (2011) compilation, the models appear
- 245 to show too much warming in both summer and winter (Fig. S3). Further work is required to determine whether the discrepancies between the temperature reconstructions and PMIP4-CMIP6 simulations indicate model deficiencies, or tell a more nuanced story (e.g. Liu et al., 2014b; Marsicek et al., 2018)arise from systematic model error, sampling biases in the data compilation (e.g. Liu et al., 2014b; Marsicek et al., 2018; Rodriguez et al., 2019) or a contribution from both sources.

There is substantial disagreement within the PMIP4-CMIP6 ensemble about the magnitude of the surface temperature

- 250 changes . The standard deviation at the regional scale. The intermodel spread of the temperature response across the PMIP4-CMIP6 ensemble is of the same magnitude as the ensemble mean for both annual (Fig. 1) and seasonal (Fig. 2) temperature changes. There is a very large spread in the high-latitude oceans and adjacent land areas in the winter hemisphere, where the spread originates from inter-model differences in the extent of the simulated sea ice -(sect. 3.4). Ice-albedo feedback would enhance inter-model temperature differences (Berger et al., 2013). The second region characterised by large inter-model dif-
- 255 ferences is where there are large changes in precipitation in the tropics. This suggests that the spread originates in inter-model differences in simulated large scale water advection, evaporative cooling, cloud cover and precipitation changes. There is no systematic reduction in the spread of temperature responses within PMIP4-CMIP6 ensemble compared to the PMIP3-CMIP5 ensemble (Fig. 1, Fig. 2). Each of the ensembles include models of different complexity, and the lack of a systematic difference suggest that complexity and model tuning has a larger impact on the responses than differences in the protocol. Thus,
- even with the though there is a protocol-forced cooling of PMIP4-CMIP6 relative to PMIP3-CMIP5, it may still be possible to consider them both these simulations can be considered as subsets of the same combined ensemble (Harrison et al., 2014), a single ensemble (see sect. 3.5; Harrison et al., 2014). However, new approaches to classify given the large inter-model range in temperature changes in both subsets of this ensemble, it may be that classifying the models to highlight the impact of model complexity or of model biases on the response climate response would be useful. This would also allow selection of subsets of the models for specific analyses, following a fit-for-purpose approach, are clearly needed.

#### 3.2 Monsoonal Response

redistribution of moisture (see e.g. Braconnot, 2004).

The enhancement of the global monsoon is the most important consequence of the mid-Holocene changes in seasonal insolation for the hydrological cycle (Jiang et al., 2015). The global monsoon domain is expanded in the PMIP4-CMIP6 *midHolocene* simulations: this occurs because of changes in both the summer rain rate and the monsoon intensity (Fig. 5). The weakening of the annual range of precipitation over the ocean and the strengthening over the continents indicates the changes reflect a

270

The most pronounced and robust changes in the monsoon occur over northern Africa and the Indian subcontinent (Fig. 6). The areal extent of the northern African monsoon is 20-50% larger than in the *piControl* simulations, but the average

rain rate only increases by 10% (Fig. 7). The intensification of precipitation on the southern flank of the Himalayas (Tab.

- 275 S1Table S2) in the *midHolocene* simulations is offset by a reduction in the Philippines and Southeast Asia (Fig. 6), so the area-averaged reduction in rain rate is reduced over the South Asian monsoon domain (Fig. 7). There is an extension and intensification of the East Asian monsoon that is consistent across the PMIP4-CMIP6 ensemble, but the change is <10% (Fig. 7). This is a region where previous analyses have shown that simulated changes in monsoon rainfall reflect the competition between enhanced contrast in moist static energy between land and ocean and increased local evaporation over the warmer</p>
- 280 oceans (Ohgaito et al., 2013). Ensemble mean changes in the North American Monsoon System, and the Southern Hemisphere monsoons are also small (Fig. 6), and less consistent across the ensemble although most of the models show a weakening and contraction of the Southern American Monsoon System and Southern African monsoon (Fig. 7). Changes in interannual internal climate variability within the monsoon systems (characterised by standard deviations in of the annual time series of both the areal extent and area-averaged rain rate; Fig. 7) are not consistent across the PMIP4-CMIP6 ensemble. Furthermore,
- those models that have the largest change in variability in one region are not necessarily the models that have large changes in other regions, which suggests that this variability is linked with regional feedbacks, rather than being an inherent characteristic of a model.

The broad scale changes in the PMIP4-CMIP6 simulations, with weaker southern and stronger and wider northern hemisphere monsoons, were present in the PMIP3-CMIP5 simulations (Fig. 6; testing the significance of the differences between

- 290 the ensembles is discussed in sec. 3.5). The response is robust across model results, indicating that all models produce the same large scale redistribution of moisture by the atmospheric circulation in response to the interhemispheric and land-sea gradients induced by the insolation and trace gas forcing. At a regional scale, however, there are differences between the two ensembles. The PMIP4-CMIP6 *midHolocene* ensemble shows wetter conditions over the Indian Ocean, a larger northward shift of the ITCZ in the Atlantic and a widening of the Pacific rain belt compared to the PMIP3-CMIP5 models (Fig. 6). The expansion
- 295 of the summer (JJA) monsoon in northern Africa is also greater in the PMIP4-CMIP6 than PMIP3-CMIP5 ensemble (Tab. S1Table S2) and the location of the northern boundary is more consistent between models. This is associated with a better representation of the northern edge of the rainbelt for the *piControl* simulation in the PMIP4-CMIP6 ensemble compared with previous generations (Fig. S1S3). However, there is no little relationship between the amount of precipitation in the *piControl* simulations and the change in precipitation. The changes in precipitation precipitation biases and the simulated midHolocene
- 300 changes in precipitation (Fig. S3). The variations in the midHolocene rainfall signal appear to be more related to local-monsoon dynamics rather than orbitally-induced insolation changes (D'Agostino et al., 2019); some of the changes may be related to the inclusion of new land surface models, or dynamic vegetation in some local temperature variations (D'Agostino et al., 2019). The modulation of this dynamical response by the land surface and vegetation components of the PMIP4-CMIP6 models should be investigated.
- Although the PMIP4-CMIP6 models show the expected expansion of the monsoons, this expansion is weaker than indicated by palaeoclimate reconstructions (Fig. 8 & S1S3). This was a feature of the PMIP3-CMIP5 simulations (Braconnot et al., 2012; Perez-Sanz et al., 2014) and indeed-previous generations of climate models (Joussaume et al., 1999; Braconnot et al., 2007). It has been suggested that this persistent mismatch between simulations and reconstructions arises from biases in the

piControl (Harrison et al., 2015). Indeed, the ensemble mean global monsoon domain in the PMIP4-CMIP6 ensemble is

- 310 more equatorward in the *piControl* compared to the observations, particularly over the ocean (Fig. 5). In northern Africa, the expansion of the monsoon domain in the *midHolocene* simulations merely removes the underestimation of its poleward extent in the *piControl* simulations (Fig. 5). Furthermore, evaluation of the *piControl* simulations using climatological precipitation data for the period between 1970 and the present day (Adler et al., 2003) shows the models fail to capture the magnitude of rainfall in the Intertropical Convergence Zone (ITCZ) and simulate a the southern portion of the South Pacific Convergence
- 315 Zone (SPCZ). The SPCZ is too zonal because of the poor representation of the SST gradient between the equator and 10°S in the west Pacific (Fig. 3; Brown et al., 2013)(Fig. 3; Brown et al., 2013; Saint-Lu et al., 2015). The PMIP4-CMIP6 models exhibit a dry bias over tropical and high northern latitude land areas, although the mid-latitude storm tracks are captured with varying levels of fidelity (Fig. 3).

There is large spread in the are large differences in the simulated change in mid-Holocene precipitation response across

320 between different models, as shown by the standard deviation around the ensemble mean, in both the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles (Fig. 6 & 8). Unsurprisingly, the ensembles exhibits the largest spread in its simulated mid-Holocene response where that response has the highest magnitude largest differences between models occurs where the simulated change in precipitation is also largest (Fig. 6).

#### 3.3 Extratropical hydrological responses

- Hydrological changes in the extratropics are comparatively muted in the PMIP4-CMIP6 ensemble, and closely resemble features seen in the PMIP3-CMIP5 ensemble. There is a reduction in rainfall at the equatorward edge of the mid-latitude storm tracks, most noticeable over the ocean (Fig. 6). The NH extratropics are generally drier in the *midHolocene* simulations than in the *piControl*. There is a large inter-model spread in the summer rainfall changes over eastern North America and central Europe (Fig. 8). The spread in summer rainfall in both regions is clearly linked to the large inter-model spread in summer
  temperature (Fig. c.f. Figs 2 & 6). Reconstructions from eastern North America suggest slightly drier conditions while reconstructions for central Europe show somewhat wetter conditions, but in neither case are these incompatible with the simulations. There are regions, however, where there is a substantial mismatch between the PMIP4-CMIP6 simulations and the pollen-
- based reconstructions. There is a simulated reduction in summer rainfall in mid-continental Eurasia (Fig. 6). This reduction is somewhat larger in the PMIP4-CMIP6 ensemble than in the PMIP3-CMIP5 ensemble, although this difference is likely not significant (Fig. 8). However, this reduction in precipitation and the consequent increase in mid-continental temperatures is incon-
- sistent with palaeoenvironmental evidence (and climate reconstructions), which show that this region was characterised by wetter and cooler conditions than today in the mid-Holocene (Fig. 8; Bartlein et al., 2017, Tab. S1)(Fig. 8; Bartlein et al., 2017, Table S2) . This indicates that model improvements have not resolved this persistent mismatch between simulated and observed mid-Holocene climate. Bartlein et al. (2017) pinpointed poor biases in the simulation of the extratropical atmospheric circulation
- 340 as the underlying cause of this mismatch. The higher resolution of most PMIP4-CMIP6 models does not seem to improve the representation of <u>these aspects of</u> the circulation. <u>Poor-Imperfect</u> simulation of the extratropical circulation could also explain the failure to capture precipitation changes over Europe accurately (Mauri et al., 2014). The PMIP4-CMIP6 ensemble shows

little change in mean annual precipitation over Europe (Fig. 6)and fails to capture the north-south gradient of changes in ... Reconstructions of mid-Holocene precipitation shown by reconstructions: with suggest modest increases in northern Europe.

345 increases in Central Europe, and much wetter conditions in the Mediterranean , compared to modest increases in northern Europe – something which is not captured by the PMIP4-CMIP6 ensemble (Fig. 8, Fig. S3).

#### 3.4 Ocean Circulation and Cryospheric Changes

The AMOC is an important factor affecting the Northern Hemisphere climate system and is a major source of decadal and multidecadal climate variability (e.g. Rahmstorf, 2002; Lynch-Stieglitz, 2017; Jackson et al., 2015). Recent studies have reported a decline of up to ~15% decline in AMOC strength from the pre-industrial period to the present day (Rahmstorf et al., 2015; Dima and Lohmann, 2010; Caesar et al., 2018; Thornalley et al., 2018), at least partly in response to anthropogenic forcing. Reproducing the AMOC of the mid-Holocene is important for understanding the climate responses to external forcing at millennial timescales. The members of both the PMIP4-CMIP6 and PMIP3-CMIP5 ensemble have different AMOC strengths in their *piControl* simulations (Fig. 9), although all models correctly predict that it is stronger at 30°N than at 50°N. There is a strong correlation (r=0.99 at 30°N) between the simulated strength of the AMOC in the *midHolocene* and the *piControl*. Fur-

- thermore, there is little change in the overall strength of the AMOC between the *midHolocene* and *piControl* experiments (Fig. 9) in either the PMIP4-CMIP6 or the PMIP3-CMIP5 simulations, and no consistency in whether this comparatively small (and probably non-significant) change is positive or negative. The small difference between the *midHolocene* and *piControl* states is suprising given the magnitude of low frequency internal variability in AMOCUsing a single metric to categorise changes in the
- 360 AMOC is awkward that two measures, both with their own uncertainties, indicate the same result increases our confidence that the overall changes were small. Shi and Lohmann (2016) detect large differences in simulated AMOC anomalies between models with coarse and higher resolutions. They suggest ocean and atmospheric processes affecting ocean salinity close to the sites of deep convection mean that higher resolution models tending tend to produce stronger *midHolocene* AMOC and lower resolution simulations a weaker AMOC than the *piControl*. The comparatively small changes in the AMOC strength be-
- 365 tween the PMIP4-CMIP6 *piControl* and *midHolocene* simulations are consistent with these earlier results, where the simulated changes are generally of less than 2 Sv (Fig. 9).

It is difficult to reconstruct past changes in the AMOC, especially its depth-integrated strength. Previous analyses have focussed on examining individual components of the AMOC, for example by using sediment grain size (Hoogakker et al., 2011; Thornalley et al., 2013; Moffa-Sanchez et al., 2015). The overall strength of the AMOC may be constrained by using

- 370 the sedimentary Pa/Th proxy-(e.g. McManus et al., 2004), although geochemical observations show that several additional factors influence Pa and Th distribution (Hayes et al 2013). The available Pa/Th records indicate no significant change in the AMOC between the mid-Holocene and the pre-industrial period (McManus et al., 2004; Ng et al., 2018; Lippold et al., 2019). Reconstruction of changes in the upper limb of the AMOC, based on geostrophic estimates of the Florida Straits surface flow, also indicate little change over the past 8000 years (Lynch-Stieglitz et al., 2009). Thus, overall, the palaeo-reconstructions are
- 375 consistent with the simulated results (Fig. 9).

The altered distribution of incoming solar radiation at the mid-Holocene would be expected to alter the seasonal cycle of sea ice concentration. Analysis of simulations from previous generations of PMIP found a consistent reduction in Arctic summer sea ice extent at the mid-Holocene, and that the amount of sea ice reduction was related to the magnitude of warming in the region (Berger et al., 2013; Park et al., 2018). These findings hold for the PMIP4 models (Fig. 10). The CMIP6-PMIP4 models

- have slightly more realistic sensitivities of Arctic sea ice to warming and greenhouse gas forcing than CMIP5-PMIP3 models, but their simulated sea ice extents cover the same large spread easily encompassing the observations (SIMIP Community, 2020)
   There is little Arctic-wide relationship between the preindustrial sea ice extent and its reduction at the mid-Holocene (Fig. 10). Local relationships may hold for key regions, such as the North Atlantic, where connections between preindustrial sea ice coverage and mid-Holocene AMOC and summer sea ice reductions have been observed (Găinuşă-Bogdan et al., 2020).
- 385 Prior statistical analysis (Berger et al., 2013) supported by recent process-based understanding (Yoshimori and Suzuki, 2019) suggests that further analysis of *midHolocene* sea ice changes would be informative for future Arctic projections (Yoshimori and Suzuki, 20 .

#### 3.5 Evaluation of mid-Holocene climate features

405

Comparisons of the PMIP4-CMIP6 simulations with either palaeoenvironmental observations or palaeoclimate reconstructions have highlighted a number of regions where there are mismatches either in magnitude or sign of the simulated response. Whilst it is possible to attempt assessment of overall The combination of the mismatches in, for example, simulated mean annual temperature or temperature seasonality results in an extremely poor overall assessment of the performance of each model (e.g. Taylor, 2001, Fig. S2) or for individual regional features (e.g. Fig. S3), their utility is unclear. Substantial further research is required before (Fig. S2). This global assessment also provides little basis for discriminating between

- 395 models, a necessary step in using the quality of *midHolocene* simulations can be used specific midHolocene simulations operationally to enhance future projections for climate services – although Schmidt et al. (2014a) provide most of the necessary groundwork(Schmidt et al., 2014a). At a regional scale (Fig. 4; Fig. 8; Fig. S2) it is clearly possible to identify models that are unable to reproduce the observations satisfactorily. Thus, there would be utility in making quantitative assessment of model performance at a regional scale. Combining regional benchmarking of model performance with process diagnosis-to ensure
- 400 that a model is correct because it captures the right processes—would therefore provide a firmer basis for using the *midHolocene* simulations to enhance our confidence in future projections.

Analyses of key features of the *midHolocene* simulations, such as the monsoon amplification or the strength of the AMOC, suggest that the PMIP4-CMIP6 simulations are should regarded as from the same population as the PMIP3-CMIP5 simulations. We formally test this by calculating Hotelling's  $T^2$  statistic (Wilks, 2011), a multivariate generalization of the ordinary *t*-statistic that is often used to examine differences in climate-model simulations (Chervin and Schenider, 1976), at each grid point of a common 1° grid for different combinations of climate variables. The patterns of "significant" (i.e. p < 0.05) tests (where one would reject the null hypothesis that the PMIP4-CMIP6 and PMIP3-CMIP6 ensemble means are equalbetween groups) are quite chaotic are random (Fig. 11) and show little relation to the largest climate anomalies (Fig. 1 & 6). There are few locations that do not fall below. The total number of "significant" grid cells does not exceed the false discovery rate

- (Wilks, 2006). Consequently there is little support for the idea that the PMIP4-CMIP6 generation of simulations differ from 410 the PMIP3-CMIP5 simulations, which were themselves not significantly different from the PMIP3-CMIP5-PMIP2-CMIP3 simulations (Harrison et al., 2015). This suggests, that all of these simulations could be considered as a single ensemble for process-based analysis (e.g. D'Agostino et al., 2019) or for the investigation of emergent constraints (Yoshimori and Suzuki, 2019), which would. Combining models from multiple ensembles could considerably enhance the statistical power of such
- 415 analyses.

Many-Several of the PMIP4-CMIP6 models have a higher climate sensitivity, defined as the response of global temperature to a doubling of CO<sub>2</sub> (Gregory et al., 2004)), than earlier versions of the same model (Tab. 1, Tab. 2), Although This increased sensitivity could contribute to the PMIP4-CMIP6 ensemble being somewhat cooler than the PMIP3-CMIP5 ensemble, the. However, two of the PMIP4-CMIP6 models have lower sensitivity and there is no real difference in the range of sensitivities of

- the two ensembles. This suggests that the change in the experimental protol appears to be the dominant protocol, specifically 420 the fact that the specified atmospheric  $CO_2$  concentration is ca 20 ppm lower in the PMIP4-CMIP6 experiments than in the PMIP3-CMIP5 experiments, is a more likely explanation for this change. This is borne out by comparison of the implied forcing as a result of the change in  $CO_2$  (Fig. 1f) and the difference in temperature between the two ensembles (Fig. 1e) (Fig. 1).
- 425 There is no inherent relationship between climate sensitivity and seasonality, because of differences in the rôle the influence of the ocean is different on seasonal compared to multi-annual timescales. Nevertheless, since the change However, changes in climate sensitivity arises from differences in basic climate feedbacks, such as can arise from water vapour or elouds, cloud feedbacks, and thus it is feasible that the change changes in climate sensitivity could affect the simulated changes in seasonality. However, we find no inherent relationship between climate sensitivity and temperature seasonality, here shown for seasonality
- 430 This is not borne out by analyses of seasonality changes in central Asia (Fig. 12). Although : although four of the individual models that have higher sensitivity in PMIP4-CMIP6 than the corresponding version of that model in PMIP3-CMIP5 show an increase in the seasonality (Fig. 12), others show a decrease in seasonality with increased sensitivity. Even if there is no ubiquitous relationship, the The fact that changes in climate sensitivity can be detected in the thermodynamic response to orbital forcing, even though the relationship in this example is not constant, raises the possibility that the changes in seasonality
- 435 shown in the *midHolocene* simulations could provide a constraint on climate sensitivity. Although we have not identified such a relationship in any region used to make model evaluations, analyses of other regions would help to verify this.

Circum-Pacific paleoclimate records document marked fluctuations in ENSO activity throughout the Holocene (Tudhope et al., 2001; McGregor and Gagan, 2004; Koutavas and Joanides, 2012; McGregor et al., 2013; Cobb et al., 2013; Carré et al., 2014; Chen et al., 2016; Grothe et al., 2019). In the central and eastern Pacific, the deepest reduction (around 2/3 in terms of

440 2-7yr variance) are observed in the 3-5 ka BP interval, rather than around the canonical ENSO variability was reduced at 6 ka midpoint compared to present (Emile-Geay et al., 2016). This reduction has been simulated by models of various complexity (e.g. Clement et al., 2000; Liu et al., 2000; Zheng et al., 2008; Chiang et al., 2009; An and Choi, 2014; Liu et al., 2014a) and is a feature of the PMIP4-CMIP6 midHolocene simulations (Tab. S1 Brown et al., submitted)(Table S3 Brown et al., submitted) . Analyses of simulated and reconstructed changes in tropical Pacific climate variability (Emile-Geay et al., 2016) showed

- that the PMIP3-CMIP5 models rarely produced a reduction in ENSO as large an ENSO as quiescent as shown by the paleoclimate observations, though the imposition of mid-Holocene boundary conditions did increase those odds. This is also true for most of the PMIP4-CMIP6 models (Table S1S2). With the exception of MIROC-ES2L, the models produce a reduction in ENSO variability but this is much smaller than the reduction implied by the palaeoclimate records. A key result of Emile-Geay et al. (2016) was that while CMIP5-PMIP3 models showed an inverse relationship between ENSO variance (inferred
- 450 from 2-7yr bandpass filtered metrics of ENSO) and seasonality (defined as the range of the monthly-mean annual cycle), the observations showed either no relationship, or a weakly positive relationship. Proxy evidence also points to one. The analysis of the PMIP4-CMIP6 ensemble of Brown et al. (submitted) shows little to no relationship as well, in accordance with this set of paleoclimate observations.

Palaeoenvironmental evidence also hints at an increased zonal SST gradient in the equatorial Pacific during the mid Holocene (Koutavas et al., 2002; Linsley et al., 2010; Carré et al., 2014), whilst the PMIP4-CMIP6 ensemble yields a slight decrease in the gradient (Table \$1\$2). Analysis of equatorial Pacific climate change and variability finds little evidence for simulated relationship between either the seasonality or SST gradient and ENSO variance in the PMIP4-CMIP6 ensemble (Brown et al., submitted).

#### 4 Conclusions

- 460 The PMIP4-CMIP6 *midHolocene* simulations show changes in seasonal temperatures and precipitation that are consistent with the expected in-line with theoretical response to changes in insolation forcing. The broadscale broad-scale patterns of change are similar to those seen in previous generations of models, most particularly the PMIP3-CMIP5 ensemble. Both ensembles show increased temperature seasonality , but with enhanced warming year-round at high northern and southern latitudes in the Northern Hemisphere resulting from higher obliquity and feedbacks from sea ice and snow cover. These
- 465 contrasting seasonal responses result in a muted annual-mean temperature changes. Both show an enhancement of the Northern Hemisphere monsoons and a weakening of the southern hemisphere monsoons. Neither the PMIP4-CMIP6 nor the PMIP3-CMIP5 models show a significant change in the AMOC during the mid-Holocene. This suggests that the changes in wind forcing, temperature gradients, seasonality of sea-ice and precipitation are not sufficient to alter the overall AMOC strength, although investigations into its various components may deliver greater insight.
- 470 Although the geographic and seasonal patterns of temperature changes in the PMIP4-CMIP6 ensemble are very similar to those seen in the PMIP3-CMIP5 ensemble, the PMIP4-CMIP6 ensemble is cooler than the PMIP3-CMIP5 ensemble in both summer and winter. This difference is consistent with the change in radiative forcing induced by using realistic GHG concentrations in the PMIP4-CMIP6. Improvements Advances in the models themselves could also contribute to this difference, in particular changes in the for example through their implementation of aerosols. There is a considerable spread in simulated
- 475 regional *midHolocene* responses climate between the PMIP4-CMIP6 models. In some cases, for example in the strength of the AMOC, this spread is clearly related to the spread in the *piControl* simulations. Biases in the *piControl* may also help to explain

the underestimation of the northward expansion of the NH monsoons, since the global monsoon domain is underestimated by both CMIP/PMIP ensembles in the *piControl* compared to observations.

This preliminary analysis of the PMIP4-CMIP6 *midHolocene* simulations already demonstrates the utility of running palaeoclimate simulations to evaluate the ability of state-of-the-art models to simulate climate change and thus to simulate the likely trajectory of future climate changes realistically. It showed that relationships between the quality of models representations of the present day and its ability to correctly simulate mid-Holocene climate changes are not straightforward: a finding that holds even for higher resolution models. Although it is disappointing that the PMIP4-CMIP6 simulations are not significantly better than the PMIP3-CMIP5 models in capturing important features of the mid-Holocene climate, analyses of the mechanisms

- 485 giving rise to these failures should shed light on the need for improved process representation physics and processes in future versions of the CMIP climate models. The examination of the how biases in the *piControl* simulations impact on the simulation of past climates is directly relevant to understanding how modern biases are propagated into future projections. Furthermore, the similarities between the PMIP4-CMIP6 and PMIP3-CMIP5 simulations provides an argument for combining these to create a single ensemble, which will considerably enhance the statistical skill-power of future analyses. Sensitivity tests, already
- 490 planned within the framework of PMIP4-CMIP6 (Otto-Bliesner et al., 2017), should help to disentangle the impacts of specific feedbacks on simulated climate changes. Finally, the-

The PMIP4-CMIP6 *midHolocene* simulations provide an opportunity both for quantitative evaluation and derivation of emergent constraints on sensitivity and feedback strength. of different aspects of model performance at both global and regional scales. They can be used in process-based analyses to assess the plausibility of future climate change mechanisms

- 495 (Braconnot and Kageyama, 2015; D'Agostino et al., 2019; Yoshimori and Suzuki, 2019). Palaeoclimate evaluations can then be used to weight models when creating fit-for-purpose ensembles to investigate climate impacts on environmental processes
  both in the past and in future projections (Schmidt et al., 2014a). Accurate representation of mid-Holocene climate, through the creation of a best-estimate climate from the PMIP ensembles, would allow us to examine e.g. the role of climate changes on the spread of early agriculture (d'Alpoim Guedes and Bocinsky, 2018; Petraglia et al., 2020). In a similar way, by constraining
- 500 the choice of future projections to models that can simulate past climate changes well, it would be possible to construct more realistic best-estimates of the impacts of projected climate changes on food security and ecosystem services (Firdaus et al., 2019; Malhi et a , or on extreme events such as flooding (Boelee et al., 2019).

*Code and data availability.* The necessary output variables from both the *midHolocene* and *piControl* simulations shall be freely available from the Earth System Grid Federation at https://esgf-node.llnl.gov/search/cmip6/. (HadGEM3-GC31-LL, AWI-ESM-1-1-LR and UofT-

- 505 CCSM-4 have committed to lodge their data as soon as practical). A GitHub repository is available at https://github.com/chrisbrierley/ PMIP4-midHolocene with the code used for this analysis. The Temperature 12k database, along with latitude-zone and global temperature reconstructions using multiple statistical methods, is available through the World Data Service (NOAA) Paleoclimatology (www.ncdc.noaa. gov/paleo/study/27330). The Bartlein et al. (2011) reconstructions are downloadable as an Electronic Supplementary Material of that article. The Compo et al. (2011) Reanalysis can be foudn at www.esrl.noaa.gov/psd/data/gridded/data.20thC\_ReanV2c.html. The precipitation
- 510 observations of Adler et al. (2003) and Xie and Arkin (1997) are archived at https://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html

and https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html respectively. The preindustrial latitudinal average temperatures were created using anomalies of Ilyas et al. (2017) from https://oasishub.co/dataset/global-monthly-temperature-ensemble-1850-to-2016 combined with the HadCRUT4 (Morice et al., 2012) absolute climatological temperatures from https://crudata.uea.ac.uk/cru/data/temperature/.

*Author contributions.* There are three tiers of authorship for this research, with the latter two in reverse alphabetical order. C.M.B., A.Z.,
515 S.P.H. and P.Br. performed the bulk of the writing and analysis. C.J.R.W., D.J.R.T., X.S., J-Y.P., R.O., D.S.K., M.K., J.C.H., M.P.E., J.E-G.,
R.D'A., D.C., M.C. and P.Ba. contributed text and analysis to the research. The third tier of authors contributed data for the manuscript.

Competing interests. The authors declare no competing interests

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the groups developing the climate models (listed in Tab. 1 & 2 of this paper) for producing and making available their model output. C.M.B., S.P.H., P.Br., C.J.R.W., X.S., R.D'A., M.C. and G.L. received funded by JPI-Belmont Forum project entitled Palaeoclimate Constraints on Monsoon Evolution and Dynamics (PaCMEDy). C.M.B., C.J.R.W. and S.P.H. were funded in part by NERC (NE/P006752/1). D.J.R.T. and C.M.B. were funded in part by NERC (NE/P006752/1). D.J.R.T. and C.M.B. were funded in part by NERC (NE/S009736/1). S.P.H. acknowledges the ERC-funded project GC2.0 (Global Change 2.0: Unlocking the past for a clearer future, grant number 694481). R.O. acknowledges support from the Integrated Research Program for Advancing Climate Models (TOUGOU programme) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. The simulations using MIROC models were conducted on the Earth Simulator of JAMSTEC. The NorESM

simulations were performed on resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway. D.S.K., C.R. and N.M. were funded by US-NSF-AGS-1602105. P.M. was supported by the state assignment project 0148-2019-0009. E.V. was supported by RSF grant 20-17-00190. We thank R. Eyles (UCL) for some invaluable database management and preprocessing.

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present), Journal of hydrometeorology, 4, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 2003.
- An, S.-I. and Choi, J.: Mid-Holocene tropical Pacific climate state, annual cycle, and ENSO in PMIP2 and PMIP3, Climate Dynamics, 43, 957–970. https://doi.org/10.1007/s00382-013-1880-z, 2014.
  - Bao, Q., Lin, P., Zhou, T., Liu, Y., Yu, Y., Wu, G., He, B., He, J., Li, L., Li, J., et al.: The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2, Advances in Atmospheric Sciences, 30, 561–576, https://doi.org/10.1007/s00376-012-2113-9, 2013.

Bartlein, P. J. and Shafer, S. L.: Paleo calendar-effect adjustments in time-slice and transient climate-model simulations (PaleoCalAdjust v1.
0): impact and strategies for data analysis, Geoscientific Model Development, 12, 3889–3913, https://doi.org/10.5194/gmd-12-3889-2019,

540 2019.

Bartlein, P. J., Harrison, S., Brewer, S., Connor, S., Davis, B., Gajewski, K., Guiot, J., Harrison-Prentice, T., Henderson, A., Peyron, O., et al.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, Climate Dynamics, 37, 775–802, https://doi.org/10.1007/s00382-010-0904-1, 2011.

Bartlein, P. J., Harrison, S. P., and Izumi, K.: Underlying causes of Eurasian midcontinental aridity in simulations of mid-Holocene climate,

#### 545 Geophysical research letters, 44, 9020–9028, https://doi.org/10.1002/2017GL074476, 2017.

Bauer, S. E. and Tsigardis, K.: Description of the GISS-E2-1-G, Journal of Advances in Modeling Earth Systems, 2020.

Berger, M., Brandefelt, J., and Nilsson, J.: The sensitivity of the Arctic sea ice to orbitally induced insolation changes: a study of the mid-Holocene Paleoclimate Modelling Intercomparison Project 2 and 3 simulations, Climate of the Past, 9, 969–982, https://doi.org/10.5194/cp-9-969-2013, 2013.

550 Boelee, L., Lumbroso, D. M., Samuels, P. G., and Cloke, H. L.: Estimation of uncertainty in flood forecasts–A comparison of methods, Journal of Flood Risk Management, 12, e12 516, 2019.

Boucher, et al.: Description of the IPSL-CM6A model, Journal of Advances in Modeling Earth Systems, 2020.

Braconnot, P.: Modéliser le dernier maximum glaciaire et l'Holocène moyen, Comptes Rendus Geoscience, 336, 711–719, https://doi.org/10.1016/j.crte.2003.12.023, 2004.

- 555 Braconnot, P. and Kageyama, M.: Shortwave forcing and feedbacks in Last Glacial Maximum and Mid-Holocene PMIP3 simulations, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 20140424, https://doi.org/10.1098/rsta.2014.0424, 2015.
  - Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C., et al.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and

560 large-scale features, Climate of the Past, 3, 261–277, https://doi.org/10.5194/cp-3-261-2007, 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, Nature Climate Change, 2, 417, https://doi.org/10.1038/nclimate1456, 2012.

Braconnot, P., Zhu, D., Marti, O., and Servonnat, J.: Strengths and challenges for transient Mid-to Late Holocene simulations with dynamical vegetation, Climate of the Past, 15, 997–1024, https://doi.org/10.5194/cp-15-997-2019, 2019.

565 Brierley, C. and Wainer, I.: Inter-annual variability in the tropical Atlantic from the Last Glacial Maximum into future climate projections simulated by CMIP5/PMIP3, Climate of the Past, 14, 1377–1390, https://doi.org/10.5194/cp-14-1377-2018, 2018.

- Brierley, C., Koch, A., Ilyas, M., Wennyk, N., and Kikstra, J.: Half the world's population already experiences years 1.5 C warmer than preindustrial, https://doi.org/10.31223/osf.io/sbc3f, 2019.
- Brown, J., Brierley, C. M., An, S.-I., Guarino, M.-V., Stevenson, S., Williams, C. J. R., et al.: Comparison of past and future simulations of ENSO in CMIP5/PMIP3 and CMIP6/PMIP4 models, Climate of the Past, ?, ???-???, submitted.
- Brown, J. R., Moise, A. F., and Colman, R. A.: The South Pacific Convergence Zone in CMIP5 simulations of historical and future climate, Climate dynamics, 41, 2179–2197, https://doi.org/10.1007/s00382-012-1591-x, 2013.

570

600

- Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed fingerprint of a weakening Atlantic Ocean overturning circulation, Nature, 556, 191, https://doi.org/10.1038/s41586-018-0006-5, 2018.
- 575 Cao, J., Wang, B., Young-Min, Y., Ma, L., Li, J., Sun, B., Bao, Y., He, J., Zhou, X., and Wu, L.: The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation, Geoscientific Model Development, 11, 2975–2993, https://doi.org/10.5194/gmd-11-2975-2018, 2018.
  - Carré, M., Sachs, J. P., Purca, S., Schauer, A. J., Braconnot, P., Falcón, R. A., Julien, M., and Lavallée, D.: Holocene history of ENSO variance and asymmetry in the eastern tropical Pacific, Science, 345, 1045–1048, https://doi.org/10.1126/science.1252220, 2014.
- 580 Chandan, D. and Peltier, W. R.: Regional and global climate for the mid-Pliocene using the University of Toronto version of CCSM4 and PlioMIP2 boundary conditions, Climate of the Past, 13, 919, https://doi.org/10.5194/cp-13-919-2017, 2017.
  - Chen, S., Hoffmann, S. S., Lund, D. C., Cobb, K. M., Emile-Geay, J., and Adkins, J. F.: A high-resolution speleothem record of western equatorial Pacific rainfall: Implications for Holocene ENSO evolution, Earth and Planetary Science Letters, 442, 61–71, https://doi.org/10.1016/j.epsl.2016.02.050, http://www.sciencedirect.com/science/article/pii/S0012821X16300759, 2016.
- 585 Chervin, R. M. and Schenider, S. H.: On determining the statistical significance of climate experiments with general circulation models, Journal of the Atmospheric Sciences, 33, 405–412, https://doi.org/10.1175/1520-0469(1976)033<0405:ODTSSO>2.0.CO;2, 1976.
  - Chiang, J. C. H., Fang, Y., and Chang, P.: Pacific Climate Change and ENSO Activity in the Mid-Holocene, Journal of Climate, 22, 923–939, https://doi.org/10.1175/2008JCLI2644.1, 2009.

Christensen et al.: Climate phenomena and their relevance for future regional climate change [IPCC WG1 AR5 Chap14], 2013.

- 590 Clement, A. C., Seager, R., and Cane, M. A.: Suppression of El Niño during the mid-Holocene by changes in the Earth's orbit, Paleoceanography, 15, 731–737, https://doi.org/10.1029/1999PA000466, 2000.
  - Cobb, K. M., Westphal, N., Sayani, H. R., Watson, J. T., Di Lorenzo, E., Cheng, H., Edwards, R. L., and Charles, C. D.: Highly Variable El Niño–Southern Oscillation Throughout the Holocene, Science, 339, 67, https://doi.org/10.1126/science.1228246, 2013.
  - Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G.,
- et al.: Long-term climate change: projections, commitments and irreversibility, in: Climate Change 2013-The Physical Science Basis:
   Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1029–1136,
   Cambridge University Press, 2013.
  - Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., et al.: Development and evaluation of an Earth-System model–HadGEM2, Geoscientific Model Development, 4, 1051–1075, https://doi.org/10.5194/gmd-4-1051-2011, 2011.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

D'Agostino, R., Bader, J., Bordoni, S., Ferreira, D., and Jungclaus, J.: Northern Hemisphere Monsoon Response to Mid-

- 605 Holocene Orbital Forcing and Greenhouse Gas-Induced Global Warming, Geophysical Research Letters, 46, 1591–1601, https://doi.org/10.1029/2018GL081589, 2019.
  - Dima, M. and Lohmann, G.: Evidence for two distinct modes of large-scale ocean circulation changes over the last century, Journal of Climate, 23, 5–16, https://doi.org/10.1175/2009JCLI2867.1, 2010.
  - Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., et al.:
- 610 Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.
  - d'Alpoim Guedes, J. and Bocinsky, R. K.: Climate change stimulated agricultural innovation and exchange across Asia, Science advances, 4, eaar4491, 2018.
- Earth System Documentation: ES-DOC, https://view.es-doc.org/index.html?renderMethod=id&project=cmip6&id= 8c42ab00-1ef2-4d5b-ade1-8bf8803cb6d4, accessed: 2020-06-22, 2019.
  - Emile-Geay, J., Cobb, K. M., Carre, M., Braconnot, P., Leloup, J., Zhou, Y., Harrison, S. P., Correge, T., McGregor, H. V., Collins, M., Driscoll, R., Elliot, M., Schneider, B., and Tudhope, A.: Links between tropical Pacific seasonal, interannual and orbital variability during the Holocene, Nature Geosci, 9, 168–173, https://doi.org/10.1038/ngeo2608, 2016.
    - Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercompar-
- ison Project Phase 6 (CMIP6) experimental design and organization., Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-1937-2016, 2016.
  - Firdaus, R. R., Gunaratne, M. S., Rahmat, S. R., and Kamsi, N. S.: Does climate change only affect food availability? What else matters?, Cogent Food & Agriculture, 5, 1707 607, 2019.
  - Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation
- 625 of climate models, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 741–866, Cambridge University Press, 2013.
  - Găinuşă-Bogdan, A., Swingedouw, D., Yiou, P., Cattiaux, J., Codron, F., and Michel, S.: AMOC and summer sea ice as key drivers of the spread in mid-Holocene winter temperature patterns over Europe in PMIP3 models, Global and Planetary Change, 184, 103 055, 2020.
  - Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B.,
  - 630 Rasch, P. J., Vertenstein, M., et al.: The community climate system model version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.
    - Gettelman, A., Hannay, C., Bacmeister, J., Neale, R., Pendergrass, A., Danabasoglu, G., Lamarque, J.-F., Fasullo, J., Bailey, D., Lawrence, D., et al.: High climate sensitivity in the Community Earth System Model Version 2 (CESM2), Geophysical Research Letters, https://doi.org/10.1029/2019GL083978, 2019.
  - 635 Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, 2013.
    - Gregory, J. M., Ingram, W., Palmer, M., Jones, G., Stott, P., Thorpe, R., Lowe, J., Johns, T., and Williams, K.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, https://doi.org/10.1029/2003GL018747, 2004.
  - 640 Grothe, P., Cobb, K. M., Liguori, G., Di Lorenzo, E., Capotondi, A., Lu, Y., Cheng, H., Edwards, R. L., Southon, J., Santos, G., Deocampo, D., Lynch-Stieglitz, J., Chen, T., Sayani, H. R., Townsend, K., Hagos, M., O'Connor, G., Thompson, D. M., Toth, L., Conroy,

J. L., and Moore, A.: Evidence for intensification of El Niño - Southern Oscillation over the late 20th century, Geophys. Res. Lett., https://doi.org/10.1029/2019GL083906, 2019.

Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: Description and evaluation

- 645 of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM), Geoscientific Model Development, 12, 343–362, https://doi.org/10.5194/gmd-12-343-2019, https://www.geosci-model-dev.net/12/343/2019/, 2019.
  - Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Description of the MIROC-ES2L Earth system model and evaluation of its climate–biogeochemical processes and feedbacks, Geoscientific Model Development, 13, 2197–2244, https://doi.org/10.5194/gmd-13-
- 650 2197-2020, https://doi.org/10.5194/gmd-13-2197-2020, 2020.
  - Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A.: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene, Climate of the Past, 9, 811–823, https://doi.org/10.5194/cp-9-811-2013, 2013.
    - Harrison, S., Bartlein, P., Brewer, S., Prentice, I., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Climate Dynamics, 43, 671–688, https://doi.org/10.1007/s00382-013-1922-6, 2014.
- 655 Harrison, S. P., Bartlein, P., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P., and Kageyama, M.: Evaluation of CMIP5 palaeosimulations to improve climate projections, Nature Climate Change, 5, 735–743, https://doi.org/10.1038/nclimate2649, 2015.
  - Harrison, S. P., Bartlein, P. J., and Prentice, I. C.: What have we learnt from palaeoclimate simulations?, Journal of Quaternary Science, 31, 363–385, https://doi.org/10.1002/jqs.2842, 2016.
- Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.
- Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al.: EC-Earth V2. 2: description and validation of a new seamless earth system prediction model, Climate dynamics, 39, 2611–2629, https://doi.org/10.1007/s00382-011-1228-5, 2012.

He et al.: CAS FGOALS-f3-L Model datasets for CMIP6 DECK experiments, to be submitted, 2020.

- 665 Hoogakker, B. A., Chapman, M. R., McCave, I. N., Hillaire-Marcel, C., Ellison, C. R., Hall, I. R., and Telford, R. J.: Dynamics of North Atlantic deep water masses during the Holocene, Paleoceanography, 26, https://doi.org/10.1029/2011PA002155, 2011.
  - Hopcroft, P. O. and Valdes, P. J.: On the Role of Dust-Climate Feedbacks During the Mid-Holocene, Geophysical Research Letters, 46, 1612–1621, https://doi.org/10.1029/2018GL080483, 2019.

Ilyas, M., Brierley, C. M., and Guillas, S.: Uncertainty in regional temperatures inferred from sparse global observations: Application to a

- probabilistic classification of El Niño, Geophysical Research Letters, 44, 9068–9074, https://doi.org/10.1002/2017GL074596, 2017.
- Jackson, L., Kahana, R., Graham, T., Ringer, M., Woollings, T., Mecking, J., and Wood, R.: Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM, Climate dynamics, 45, 3299–3316, https://doi.org/10.1007/s00382-015-2540-2, 2015.
  Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., Wong, K., and Syktus, J.: Australia's CMIP5 submission using the CSIRO Mk3. 6 model, Aust. Meteor. Oceanogr. J, 63, 1–13, https://doi.org/10.22499/2.6301.001, 2013.
- 675 Jiang, D., Tian, Z., and Lang, X.: Mid-Holocene net precipitation changes over China: model–data comparison, Quaternary Science Reviews, 82, 104–120, https://doi.org/10.1016/j.quascirev.2013.10.017, 2013.
  - Jiang, D., Tian, Z., and Lang, X.: Mid-Holocene global monsoon area and precipitation from PMIP simulations, Climate Dynamics, 44, 2493–2512, https://doi.org/10.1007/s00382-014-2175-8, 2015.

Joussaume, S. and Braconnot, P.: Sensitivity of paleoclimate simulation results to season definitions, Journal of Geophysical Research:

- 680 Atmospheres, 102, 1943–1956, https://doi.org/10.1029/96JD01989, 1997.
  - Joussaume, S., Taylor, K., Braconnot, P., Mitchell, J., Kutzbach, J., Harrison, S., Prentice, I., Broccoli, A., Abe-Ouchi, A., Bartlein, P., et al.: Monsoon changes for 6000 years ago: results of 18 simulations from the Paleoclimate Modeling Intercomparison Project (PMIP), Geophysical Research Letters, 26, 859–862, https://doi.org/10.1029/1999GL900126, 1999.
- Juckes, M., Taylor, K. E., Durack, P., Lawrence, B., Mizielinski, M., Pamment, A., Peterschmitt, J.-Y., Rixen, M., and Sénésis, S.: The CMIP6 Data Request (version 01.00.31), Geoscientific Model Development Discussions, 2019, 1–35, https://doi.org/10.5194/gmd-2019-
- 219, https://www.geosci-model-dev-discuss.net/gmd-2019-219/, 2019.
  - Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., et al.: The PMIP4 contribution to CMIP6-Part 1: Overview and over-arching analysis plan, Geoscientific Model Development, 11, 1033–1057, https://doi.org/10.5194/gmd-11-1033-2018, 2018.
- 690 Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., et al.: A global database of Holocene paleo-temperature records, Scientific Data, 7, https://doi.org/10.1038/s41597-020-0445-3, 2020.
  - Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P., Heiri, O., and Davis, B.: Holocene global mean surface temperature: a multi-method reconstruction approach, Scientific Data, in press.

Kohfeld, K. E. and Harrison, S. P.: How well can we simulate past climates? Evaluating the models using global palaeoenvironmental

- datasets, Quaternary Science Reviews, 19, 321–346, https://doi.org/10.1016/S0277-3791(99)00068-2, 2000.
   Koutavas, A. and Joanides, S.: El Niño-Southern Oscillation extrema in the Holocene and Last Glacial Maximum, Paleoceanography, 27, PA4208, https://doi.org/10.1029/2012PA002378, 2012.
  - Koutavas, A., Lynch-Stieglitz, J., Marchitto, T. M., and Sachs, J. P.: El Nino-like pattern in ice age tropical Pacific sea surface temperature, Science, 297, 226–230, https://doi.org/10.1126/science.1072376, 2002.
- 700 Li, L., Lin, P., Yu, Y., Wang, B., Zhou, T., Liu, L., Liu, J., Bao, Q., Xu, S., Huang, W., et al.: The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2, Advances in Atmospheric Sciences, 30, 543–560, https://doi.org/10.1007/s00376-012-2140-6, 2013.
  - Linsley, B. K., Rosenthal, Y., and Oppo, D. W.: Holocene evolution of the Indonesian throughflow and the western Pacific warm pool, Nature Geoscience, 3, 578–583, https://doi.org/10.1038/ngeo920, 2010.
- 705 Lippold, J., Pöppelmeier, F., Süfke, F., Gutjahr, M., Goepfert, T. J., Blaser, P., Friedrich, O., Link, J. M., Wacker, L., Rheinberger, S., et al.: Constraining the variability of the atlantic meridional overturning circulation during the holocene, Geophysical Research Letters, https://doi.org/10.1029/2019GL084988, 2019.
  - Liu, Z., Kutzbach, J., and Wu, L.: Modeling climate shift of El Niño variability in the Holocene, Geophysical Research Letters, 27, 2265–2268, https://doi.org/10.1029/2000GL011452, 2000.
- 710 Liu, Z., Lu, Z., Wen, X., Otto-Bliesner, B. L., Timmermann, A., and Cobb, K. M.: Evolution and forcing mechanisms of El Niño over the past 21,000 years, Nature, 515, 550–553, https://doi.org/10.1038/nature13963, 2014a.
  - Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Timm, O. E.: The Holocene temperature conundrum, Proceedings of the National Academy of Sciences, 111, E3501–E3505, https://doi.org/10.1073/pnas.1407229111, 2014b.
- 715 Lozier, M., Li, F., Bacon, S., Bahr, F., Bower, A., Cunningham, S., De Jong, M., De Steur, L., Deyoung, B., Fischer, J., et al.: A sea change in our view of overturning in the subpolar North Atlantic, Science, 363, 516–521, 2019.

Lynch-Stieglitz, J.: The Atlantic meridional overturning circulation and abrupt climate change, Annual review of marine science, 9, 83–104, https://doi.org/10.1146/annurev-marine-010816-060415, 2017.

Lynch-Stieglitz, J., Curry, W. B., and Lund, D. C.: Florida Straits density structure and transport over the last 8000 years, Paleoceanography, 24. https://doi.org/10.1029/2008PA001717, 2009.

- Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M. G., Field, C. B., and Knowlton, N.: Climate change and ecosystems: threats, opportunities and solutions, Philosophical Transactions of the Royal Society B: Biological Sciences, 375, 20190104, https://doi.org/10.1098/rstb.2019.0104, 2020.
- Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., and Brewer, S.: Reconciling divergent trends and millennial variations in Holocene temperatures, Nature, 554, 92, https://doi.org/10.1038/nature25464, 2018.
  - Mauri, A., Davis, B., Collins, P., and Kaplan, J.: The influence of atmospheric circulation on the mid-Holocene climate of Europe: a datamodel comparison, Climate of the Past, 10, 1925–1938, https://doi.org/10.5194/cp-10-1925-2014, 2014.
    - Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., et al.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO2, Journal of Advances in Modeling

#### 730 Earth Systems, 11, 998–1038, 2019.

720

750

- McGregor, H. V. and Gagan, M. K.: Western Pacific coral  $\delta^{18}O$  records of anomalous Holocene variability in the El Niño–Southern Oscillation, Geophysical Research Letters, 31, L11 204, https://doi.org/10.1029/2004GL019972, 2004.
- McGregor, H. V., Fischer, M. J., Gagan, M. K., Fink, D., Phipps, S. J., Wong, H., and Woodroffe, C. D.: A weak El Niño-Southern Oscillation with delayed seasonal growth around 4,300 years ago, Nature Geoscience, 6, 949–953, https://doi.org/10.1038/ngeo1936, 2013.
- 735 McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D., and Brown-Leger, S.: Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes, Nature, 428, 834, https://doi.org/10.1038/nature02494, 2004.
  - Moffa-Sanchez, P., Hall, I. R., Thornalley, D. J., Barker, S., and Stewart, C.: Changes in the strength of the Nordic Seas Overflows over the past 3000 years, Quaternary Science Reviews, 123, 134–143, https://doi.org/10.1016/j.quascirev.2015.06.007, 2015.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change

- vusing an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.
  - Ng, H. C., Robinson, L. F., McManus, J. F., Mohamed, K. J., Jacobel, A. W., Ivanovic, R. F., Gregoire, L. J., and Chen, T.: Coherent deglacial changes in western Atlantic Ocean circulation, Nature communications, 9, https://doi.org/10.1038/s41467-018-05312-3, 2018.

Ohgaito, R. and Abe-Ouchi, A.: The effect of sea surface temperature bias in the PMIP2 AOGCMs on mid-Holocene Asian monsoon enhancement, Climate dynamics, 33, 975–983, https://doi.org/10.1007/s00382-009-0533-8, 2009.

Ohgaito, R., Sueyoshi, T., Abe-Ouchi, A., Hajima, T., Watanabe, S., Kim, H.-J., Yamamoto, A., and Kawamiya, M.: Can an Earth System Model simulate better climate change at mid-Holocene than an AOGCM? A comparison study of MIROC-ESM and MIROC3, Climate of the Past, 9, 1519–1542, https://doi.org/10.5194/cp-9-1519-2013, 2013.

O'ishi, R. and Abe-Ouchi, A.: Polar amplification in the mid-Holocene derived from dynamical vegetation change with a GCM, Geophysical Research Letters, 38, https://doi.org/10.1029/2011GL048001, 2011.

Otto-Bliesner, B. L., Braconnot, P., Harrison, S. P., Lunt, D. J., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Capron, E., Carlson, A. E., Dutton, A., et al.: The PMIP4 contribution to CMIP6–Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations, Geoscientific Model Development, 10, 3979–4003, https://doi.org/10.5194/gmd-10-3979-2017, 2017. Otto-Bliesner, B. L., Brady, E. C., Tomas, R. A., Albani, S., Bartlein, P. J., Mahowald, N. M., Shafer, S. L., Kluzek, E., Lawrence, P. J.,

- 755 Leguy, G., Rothstein, M., and Sommers, A.: A comparison of the CMIP6 midHolocene and lig127k simulations in CESM2, *submitted*, 2020.
  - Park, H.-S., Kim, S.-J., Seo, K.-H., Stewart, A. L., Kim, S.-Y., and Son, S.-W.: The impact of Arctic sea ice loss on mid-Holocene climate, Nature communications, 9, 1–9, 2018.
- Pausata, F. S., Messori, G., and Zhang, Q.: Impacts of dust reduction on the northward expansion of the African monsoon during the Green Sahara period, Earth and Planetary Science Letters, 434, 298–307, https://doi.org/10.1016/j.epsl.2015.11.049, 2016.
- Perez-Sanz, A., Li, G., González-Sampériz, P., and Harrison, S. P.: Evaluation of modern and mid-Holocene seasonal precipitation of the Mediterranean and northern Africa in the CMIP5 simulations, Climate of the Past, 10, 551–568, https://doi.org/10.5194/cp-10-551-2014, 2014.
- Petraglia, M. D., Groucutt, H. S., Guagnin, M., Breeze, P. S., and Boivin, N.: Human responses to climate and ecosystem change in ancient
   Arabia, Proceedings of the National Academy of Sciences, 117, 8263–8270, 2020.
  - Phillips, A. S., Deser, C., and Fasullo, J.: Evaluating modes of variability in climate models, Eos, Transactions American Geophysical Union, 95, 453–455, https://doi.org/10.1002/2014EO490002, 2014.
    - Phipps, S., Rotstayn, L., Gordon, H., Roberts, J., Hirst, A., and Budd, W.: The CSIRO Mk3L climate system model version 1.0–Part 2: Response to external forcings, Geoscientific Model Development, 5, 649–682, https://doi.org/10.5194/gmd-5-649-2012, 2012.
- 770 Pollard, D. and Reusch, D. B.: A calendar conversion method for monthly mean paleoclimate model output with orbital forcing, Journal of Geophysical Research: Atmospheres, 107, ACL–3, 2002.
  - Prado, L. F., Wainer, I., and Chiessi, C. M.: Mid-Holocene PMIP3/CMIP5 model results: Intercomparison for the South American monsoon system, The Holocene, 23, 1915–1920, https://doi.org/10.1177/0959683613505336, 2013.
- Prentice, I. C., Jolly, D., and Participants, B. .: Mid-Holocene and glacial-maximum vegetation geography of the northern continents and
   Africa, Journal of biogeography, 27, 507–519, https://doi.org/10.1046/j.1365-2699.2000.00425.x, 2000.
- Rahmstorf, S.: Ocean circulation and climate during the past 120,000 years, Nature, 419, 207, https://doi.org/10.1038/nature01090, 2002.
  Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J.: Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation, Nature climate change, 5, 475, https://doi.org/10.1038/nclimate2554, 2015.
- Rodriguez, L. G., Cohen, A. L., Ramirez, W., Oppo, D. W., Pourmand, A., Edwards, R. L., Alpert, A. E., and Mollica, N.: Mid-Holocene,
  coral-based sea surface temperatures in the western tropical Atlantic, Paleoceanography and Paleoclimatology, 34, 1234–1245, 2019.
- Saint-Lu, M., Braconnot, P., Leloup, J., Lengaigne, M., and Marti, O.: Changes in the ENSO/SPCZ relationship from past to future climates, Earth and Planetary Science Letters, 412, 18–24, 2015.
  - Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., 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., Tremblay, L.-B., and Yiou, P.:
- 785 Using palaeo-climate comparisons to constrain future projections in CMIP5, Climate of the Past, 10, 221–250, https://doi.org/10.5194/cp-10-221-2014, https://www.clim-past.net/10/221/2014/, 2014a.
  - Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., et al.:
     Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 6, 141–184, https://doi.org/10.1002/2013MS000265, 2014b.
- 790 Seland, Ø., Bentsen, M., Seland Graff, L., Olivié, D., Toniazzo, T., Gjermundsen, A., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Schancke Aas, K., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Hafsahl Karset, I. H.,

Landgren, O., Liakka, J., Onsum Moseid, K., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iverson, T., and Schulz, M.: The Norwegian Earth System Model, NorESM2 – Evaluation of theCMIP6 DECK and historical simulations, Geoscientific Model Development Discussions, 2020, 1–68, https://doi.org/10.5194/gmd-2019-378, https://www.geosci-model-dev-discuss.net/gmd-2019-378/, 2020.

795

- Shi, X. and Lohmann, G.: Simulated response of the mid-Holocene Atlantic meridional overturning circulation in ECHAM6-FESOM/MPIOM, Journal of Geophysical Research: Oceans, 121, 6444–6469, https://doi.org/10.1002/2015JC011584, 2016.
- Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Gößling, H. F., et al.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Climate Dynamics, 44, 757–780, https://doi.org/10.1007/s00382-014-2290-6, 2015.
  - SIMIP Community: Arctic Sea Ice in CMIP6, Geophysical Research Letters, 47, e2019GL086749, https://doi.org/10.1029/2019GL086749, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL086749, 2020.
    - Smeed, D., Moat, B., , Rayner, D., Johns, W., Baringer, M., Volkov, D., and Frajka-Williams, E.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional overturning circulation and heatflux array-Western boundary time
- 805 series) array at 26N from 2004 to 2018, https://doi.org/10.5285/8cd7e7bb-9a20-05d8-e053-6c86abc012c2, https://dx.doi.org/10.5285/ 8cd7e7bb-9a20-05d8-e053-6c86abc012c2, 2019.
  - Sueyoshi, T., Ohgaito, R., Yamamoto, A., Chikamoto, M., Hajima, T., Okajima, H., Yoshimori, M., Abe, M., O'ishi, R., Saito, F., et al.: Setup of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM, Geoscientific Model Development, 6, 819–836, https://doi.org/10.5194/gmd-6-819-2013, 2013.
- 810 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/ 10.1029/2000JD900719, 2001.
  - Thornalley, D. J., Blaschek, M., Davies, F. J., Praetorius, S., Oppo, D. W., McManus, J. F., Hall, I. R., Kleiven, H., Renssen, H., and McCave, I. N.: Long-term variations in Iceland–Scotland overflow strength during the Holocene, Climate of the Past, 9, 2073–2084, https://doi.org/10.5194/cp-9-2073-2013, 2013.
- 815 Thornalley, D. J., Oppo, D. W., Ortega, P., Robson, J. I., Brierley, C. M., Davis, R., Hall, I. R., Moffa-Sanchez, P., Rose, N. L., Spooner, P. T., et al.: Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years, Nature, 556, 227, https://doi.org/10.1038/s41586-018-0007-4, 2018.
  - Tudhope, A. W., Chilcott, C. P., McCulloch, M. T., Cook, E. R., Chappell, J., Ellam, R. M., Lea, D. W., Lough, J. M., and Shimmield, G. B.: Variability in the El Niño-Southern Oscillation through a glacial-interglacial cycle, Science, 291, 1511–1517,
- 820 https://doi.org/10.1126/science.1057969, 2001.
- Voldoire, A., Sanchez-Gomez, E., y Mélia, D. S., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., et al.: The CNRM-CM5. 1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y, 2013.
  - Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., Iakovlev, N. G.,
- 825 Shestakova, A. A., et al.: Simulation of the modern climate using the INM-CM48 climate model, Russian Journal of Numerical Analysis and Mathematical Modelling, 33, 367–374, https://doi.org/10.1515/rnam-2018-0032, 2018.
  - Walsh, J. E., Fetterer, F., Stewart, J. S., and Chapman, W. L.: A database for depicting Arctic sea ice variations back to 1850, Geographical Review, 1, 89–107, 2016.

Wang, B., Kim, H.-J., Kikuchi, K., and Kitoh, A.: Diagnostic metrics for evaluation of annual and diurnal cycles, Climate dynamics, 37,

830 941–955, https://doi.org/10.1007/s00382-010-0877-0, 2011.

Wang, H., Li, L. J., Chen, X. L., and Wang, B.: Comparison of Climate sensitivities and feedbacks between FGOALS-g3 and FGOALS-g2, to be submitted, 2020.

Wilks, D.: On "field significance" and the false discovery rate, Journal of applied meteorology and climatology, 45, 1181–1189, https://doi.org/10.1175/JAM2404.1, 2006.

- 835 Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, 2011.
  - Williams, C. J. R., Guarino, M.-V., Capron, E., Malmierca-Vallet, I., Singarayer, J. S., Sime, L. C., Lunt, D. J., and Valdes, P. J.: The UK contribution to CMIP6/PMIP4: mid-Holocene and Last Interglacial experiments with HadGEM3, and comparison to the pre-industrial era and proxy data, Climate of the Past Discussions, 2020, 1–42, https://doi.org/10.5194/cp-2019-160, https://www.clim-past-discuss.net/ cp-2019-160/, 2020.
- 840 Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H., Hill, R., et al.: The Met Office global coupled model 3.0 and 3.1 (GC3. 0 and GC3. 1) configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, 2018.
  - Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., and Döscher, R.: On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geoscientific Model Development Discussions, 2019, 1–13, https://doi.org/10.5194/gmd-2019-282,
- https://www.geosci-model-dev-discuss.net/gmd-2019-282/, 2019.
  - Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, https://doi.org/10.1175/2008JAMC1921.1, 1997.
  - Xin, X.-G., Wu, T.-W., Jiang-Long, L., Wang, Z.-Z., Li, W.-P., and Wu, F.-H.: How well does BCC\_CSM1.1 reproduce the 20th century climate change over China?, Atmospheric and Oceanic Science Letters, 6, 21–26, https://doi.org/10.1080/16742834.2013.11447053, 2013.
- 850 Yoshimori, M. and Suzuki, M.: The relevance of mid-Holocene Arctic warming to the future, Climate of the Past, 15, 1375–1394, https://doi.org/10.5194/cp-15-1375-2019, 2019.
  - Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., et al.: A new global climate model of the Meteorological Research Institute: MRI-CGCM3—model description and basic performance—, Journal of the Meteorological Society of Japan. Ser. II, 90, 23–64, https://doi.org/10.2151/jmsj.2012-A02, 2012.
- 855 Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., et al.: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2. 0: Description and Basic Evaluation of the Physical Component, Journal of the Meteorological Society of Japan. Ser. II, https://doi.org/10.2151/jmsj.2019-051, 2019.
  - Zheng, W., Braconnot, P., Guilyardi, E., Merkel, U., and Yu, Y.: ENSO at 6ka and 21ka from ocean–atmosphere coupled model simulations, Clim. Dyn., 30, 745–762, https://doi.org/10.1007/s00382-007-0320-3, 2008.



Figure 1. Annual mean surface temperature change in the *midHolocene* simulations (°C). (a) The ensemble mean, annual mean temperature changes in PMIP4-CMIP6 (*midHolocene - piControl*) and (b) its the intermodel spread (defined as the across ensemble standard devationdeviation). (c) The ensemble mean, annual mean temperature change in PMIP3-CMIP5 and (d) its standard devationdeviation. (e) The difference in temperature between the two ensembles. (f) The estimated response to the greenhouse gas concentration reductions in the experimental protocol.



**Figure 2. Seasonal surface temperature changes in the** *midHolocene* **simulations** (°**C**). (a,b) The ensemble mean temperature changes in PMIP4-CMIP6 (*midHolocene - piControl*) in DJF and JJA. (c,d) The ensemble mean temperature changes in PMIP3-CMIP5 in DJF and JJA. The intermodel spread (defined as the across ensemble standard devations deviation) in seasonal temperature changes seen across the ensembles: (e) DJF in PMIP4-CMIP6, (f) JJA in PMIP4-CMIP6, (g) DJF in PMIP3-CMIP5 and (h) JJA in PMIP3-CMIP6.



**Figure 3. Comparison of the CMIP6 ensemble to observations**. (a) The annual mean surface temperatures in the C20 Reanalysis (Compo et al., 2011) between <u>1981-19001881-1900</u>. (b) The ensemble mean difference in annual surface air temperature from the C20 Reanalysis within the *piControl* simulations. Ability of the ensemble to simulate the seasonal cycle of precipitation for the present-day. (c,e) The precipitation climatology seen in the GPCP (Adler et al., 2003) observational dataset between 1971-2000 for DJF and JJA respectively. (d,f) The ensemble mean difference in seasonal precipitation from GPCP within the *piControl* simulations for DJF and JJA respectively. Stippling indicates that two-thirds of the models agree on the sign of the bias.



**Figure 4. Zonal averaged temperatures within the PMIP4-CMIP6 ensemble**. (a) Comparison of the *piControl* zonal mean temperature profile of individual climate models to the 1850-1900 observations. The area-averaged, annual mean surface air temperature for 30° latitude bands in the CMIP6 models (identified), CMIP5 models (blue circles) and a spatially complete compilation of instrumental observations over 1850-1900 (black, Ilyas et al., 2017; Morice et al., 2012). (b) The simulated annual mean temperature change averaged over 30° zonal bands for each of the individual CMIP6 models. The equivalent changes estimated from the Temperature 12k compilation (Kaufman et al., 2020) via a multi-method approach are shown along with their 80% confidence interval. The distribution of Bartlein et al. (2011) reconstructed temperatures within each latitude bands are shown in the NH, because the tropical and southern hemisphere latitudes are only represented by sites in Africa. *The data points for all models, as well as the equivalents over land or ocean, are provided in Table S4*.



**Figure 5. PMIP4-CMIP6 ensemble mean global monsoon domain (mm/day)**. The monsoon domain for each simulation is identified by applying the definitions of Wang et al. (2011) and in sect. 2.2 to the PMIP4-CMIP6 ensemble mean of both (a) the *midHolocene* and (b) the *piControl* simulations. The black contour in (a,b) shows the boundary of the domain derived from present-day observations (Adler et al., 2003). The simulated changes in the monsoon domain are determined by changes in both (c) the monsoon intensity <u>– average rain rate</u> difference between summer and winter <u>– and (d)</u> the summer rain rate. In (c,d) the red and blue contours show the boundary of *midHolocene* and *piControl* global monsoon domains respectively.



**Figure 6.** *midHolocene* **seasonal changes in precipitation (mm/day).** (a,b) The ensemble mean precipitation changes in PMIP4-CMIP6 (*midHolocene - piControl*) in DJF and JJA. (c,d) The ensemble mean precipitation changes in PMIP3-CMIP5 in DJF and JJA. (e,f) The differences in DJF and JJA precipitation between the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles. The <u>intermodel spread (defined as the across ensemble</u> standard <u>devations-deviation</u>) in seasonal precipitation changes seen across the ensembles: (g) DJF in PMIP4-CMIP6, (h) JJA in PMIP4-CMIP6, (i) DJF in PMIP3-CMIP5 and (j) JJA in PMIP3-CMIP6.



**Figure 7. Relative changes in individual** *midHolocene* **monsoons.** Five different monsoon diagnostics (see sect. 2.2) are computed for each of seven different regional domains (Christensen et al., 2013). (a) The change in area-averaged precipitation rate during the monsoon season (MJJAS) for each individual monsoon system. (b) The change in the areal extent of the regional monsoon domains. (c) The percentage change in the total amount of water precipitated in each monsoon season (computed as the precipitation rate multiplied by the areal extent). (d) Change in the standard deviation of interannual variability in the area-averaged precipitation rate. (e) The change in standard deviation of the year-to-year variations in the areal extent of the regional monsoon domain. The abbreviations used to identify each regional domain are: North America Monsoon System (NAMS), North Africa (NAF), Southern Asia (SAS) and East Asia summer (EAS) in the Northern Hemisphere and South America Monsoon System (SAMS), South Africa (SAF) and Australian-Maritime Continent (AUSMC) in the Southern Hemisphere.



**Figure 8.** Comparison between simulated annual precipitation changes and pollen-based reconstructions (from Bartlein et al., 2011). Seven regions where multiple quantitative reconstructions exist are chosen. Six of them are defined after Christensen et al. (2013), and are Northern Europe (NEU), Central Europe (CEU), the Mediterranean (MED), the Sahara/Sahel (SAH), East Asia (EAS) and Eastern North America (ENA). Mid-continental Eurasia (B17) is specified by Bartlein et al. (2017) as 40–60°N, 30-120°E. The distribution of reconstructions within the region are shown by boxes and whiskers. The area-averaged change in mean annual precipitation simulated by CMIP6 (individually identifiable) and CMIP5 (blue) within each region is shown for comparison. (After Flato et al., 2013)(After Flato et al., 2013)



**Figure 9. Atlantic Meridional Overturning Circulation in the simulations**. The strength of the AMOC is defined as the maximum of the mean meridional mass overturning streamfunction below 500m at 30° and 50°N in the Atlantic. The strength in the *piControl* simulation provides the horizontal axis, whilst the vertical location is given by the strength in the *midHolocene* simulation. Data points lying on the 1:1 line demonstrate no change between the two simulations. Observational estimates of the present-day AMOC strength are shown from both the RAPID-MOCHA array (at 26°N, Smeed et al., 2019) and the OSNAP section (between 53°N and 60°N, Lozier et al., 2019).



Figure 10. Changes in Arctic sea ice minimum extent. The change in the areal extent of the minimum sea ice cover (i.e. gridboxes with greater than 15% concentration) at the mid-Holocene compared to (a) the minimum sea ice extent in the piControl simulations and (b) the Arctic annual mean temperature change. Observational estimates of the preindustrial extent (Walsh et al., 2016) and mid-Holocene Arctic warming (Fig. 4; Kaufman et al., in press) are also shown.



### Hotelling's T<sup>2</sup> p-values -- CMIP6/PMIP4 vs. CMIP5/PMIP3

**Figure 11. Maps of the p-values of Hotelling's**  $T^2$  **test** (Wilks, 2011) comparing the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles. Four different combinations of the key variables analysed here are assessed (given in the top left above the panels). Values less than 0.05 would ordinarily be considered to be significant, but the total number of such values on each individual map does not exceed the false discovery rate. Harrison et al. (2015) presents equivalent analysis comparing PMIP3-CMIP5 with PMIP2-CMIP3 (using the variables in the top left panel).



**Figure 12. The relationship between equilibrium climate sensitivity and increasing seasonality over Central Asia**. The seasonality is computed as the mean temperature of the warmest month minus the mean temperature of the coldest month, averaged over 30–50°N, 60–75°E (Christensen et al., 2013). The shifts between different generations of models are highlighted indicated, and labelled after their modelling group (NCAR developed both CCSM4 & CESM2; NCC developed NorESM1-F & NorESM2-LM; UKMO developed HadGEM2-CC, HadGEM2-ES & HadGEM3-GC31-LL).

model	$\begin{vmatrix} \Delta T_{2xCO_2}^{eq} \\ (\mathbf{K}) \end{vmatrix}$	<i>midHolocene</i> length† (yrs)	piControl length† (yrs)	Model Reference	Expt Ref. & Notes
AWI-ESM-1-1-LR	3.6	100	100	Sidorenko et al. (2015)	Dynamic Vegetation
CESM2	5.3	700	1200	Gettelman et al. (2019)	
					Potential Natural
					Land Cover
					Otto-Bliesner et al. (2020)
EC-Earth3-LR				Wyser et al. (2019)	
~~~~~~~	4.3	200	200		~
FGOALS-f3-L	3.0	500	561	Wang et al. (2020)	-
FGOALS-g3	2.9	500	200	He et al. (2020)	-
GISS-E2-1-G	2.7	100	851	Bauer and Tsigardis (2020)	_
HadGEM3-GC31	5.4	100	100	Williams et al. (2018)	
					- <u>Williams et al. (2020)</u>
INM-CM4-8	2.1	200	531	Volodin et al. (2018)	-
IPSL-CM6A-LR	4.5	550	1200	Boucher, et al. (2020)	TSI of 1361.2 $W/m^2$
MIROC-ES2L	2.7	100	500	Hajima et al. (2020)	-
MPI-ESM1-2-LR				Mauritsen et al. (2019)	
	2.8	500	1000		$\overline{\sim}$
MRI-ESM2	3.1	200	701	Yukimoto et al. (2019)	-
NESM3	3.7	100	100	Cao et al. (2018)	-
NorESM1-F	2.3	200	200	Guo et al. (2019)	-
NorESM2-LM				Seland et al. (2020)	
	2.5	200	200		≂
UofT-CCSM-4	3.2	100	100	Chandan and Peltier (2017)	TSI of 1360.89 $W/m^2$

 Table 1. Models contributing midHolocene simulations under CMIP6. See Table S1 for further information about the individual simulations.

†The lengths given are the number of simulated years used here to compute the diagnostics. These years are taken after the model has been spun-up.

model	$\Delta T^{eq}_{2xCO_2}$	midHolocene	piControl	Reference
	(K)	length† (yrs)	length† (yrs)	
bcc-csm1-1	3.1	100	500	Xin et al. (2013)
CCSM4	2.9	301	1051	Gent et al. (2011)
CNRM-CM5	3.3	200	850	Voldoire et al. (2013)
CSIRO-MK3-6-0	4.1	100	500	Jeffrey et al. (2013)
CSIRO-MK3L-1-2	3.1	500	1000	Phipps et al. (2012)
EC-Earth-2-2	4.2	40	40	Hazeleger et al. (2012)
FGOALS-G2	3.7	680	700	Li et al. (2013)
FGOALS-S2	4.5	100	501	Bao et al. (2013)
GISS-E2-R	2.1	100	500	Schmidt et al. (2014b)
HadGEM2-CC	4.5	35	240	Collins et al. (2011)
HadGEM2-ES	4.6	101	336	Collins et al. (2011)
IPSL-CM5A-LR	4.1	500	1000	Dufresne et al. (2013)
MIROC-ESM	4.7	100	630	Sueyoshi et al. (2013)
MPI-ESM-P	3.5	100	1156	Giorgetta et al. (2013)
MRI-CGCM3	2.6	100	500	Yukimoto et al. (2012)

Table 2. Models contributing midHolocene simulations under CMIP5. See Table S1 for links to each individual simulation.

†The lengths given are the number of simulated years used here to compute the diagnostics. These years are taken afer the model has been spun-up.

# **Supplementary Information for Large-scale features and evaluation of the PMIP4-CMIP6** *midHolocene* **simulations**

Chris M. Brierley<sup>1</sup>, Anni Zhao<sup>1</sup>, Sandy P. Harrison<sup>2</sup>, Pascale Braconnot<sup>3</sup>, Charles J. R. Williams<sup>4,5</sup>, David J. R. Thornalley<sup>1</sup>, Xiaoxu Shi<sup>6</sup>, Jean-Yves Peterschmitt<sup>3</sup>, Rumi Ohgaito<sup>7</sup>, Darrell S. Kaufman<sup>8</sup>, Masa Kageyama<sup>3</sup>, Julia C. Hargreaves<sup>9</sup>, Michael P. Erb<sup>8</sup>, Julien Emile-Geay<sup>10</sup>, Roberta D'Agostino<sup>11</sup>, Deepak Chandan<sup>12</sup>, Matthieu Carré<sup>13,14</sup>, Partrick J. Bartlein<sup>15</sup>, Weipeng Zheng<sup>16</sup>, Zhongshi Zhang<sup>17</sup>, Qiong Zhang<sup>18</sup>, Hu Yang<sup>6</sup>, Evgeny M. Volodin<sup>19</sup>, Robert A. Tomas<sup>20</sup>, Cody Routson<sup>8</sup>, W. Richard Peltier<sup>12</sup>, Bette Otto-Bliesner<sup>20</sup>, Polina A. Morozova<sup>21</sup>, Nicholas P. McKay<sup>8</sup>, Gerrit Lohmann<sup>6</sup>, Allegra N. Legrande<sup>22</sup>, Chuncheng Guo<sup>17</sup>, Jian Cao<sup>23</sup>, Esther Brady<sup>20</sup>, James D. Annan<sup>9</sup>, and Ayako Abe-Ouchi<sup>7,24</sup>

<sup>1</sup>Department of Geography, University College London, London, WC1E 6BT, UK

<sup>2</sup>Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AB, UK

<sup>3</sup>Laboratoire des Sciences du Climat et de l'Environnement-IPSL, Unité Mixte CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, Gif-sur-Yvette, France

<sup>4</sup>Department of Meteorology, University of Reading, Reading, RG6 6BB, UK

<sup>5</sup>School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK

<sup>6</sup>Alfred-Wegener-Institut Helmholtz-Zentrum f ur Polar- und Meeresforschung, Bremerhaven, Germany

<sup>7</sup>Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

<sup>8</sup>School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA.

<sup>9</sup>Blue Skies Research Ltd, The Old Chapel, Albert Hill, Settle, BD24 9HE, UK

<sup>10</sup>Department of Earth Sciences, University of Southern California, Los Angeles, California, USA

<sup>11</sup>Max Planck Institute for Meteorology, Hamburg, Germany

<sup>12</sup>Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario, M5S1A7, Canada

<sup>13</sup>LOCEAN Laboratory, Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Paris, France

<sup>14</sup>CIDIS-LID-Facultad de Ciencias y Filosofía-Universidad Peruana Cayetano Heredia, Lima, Peru

<sup>15</sup>Department of Geography, University of Oregon, Eugene, OR 97403, USA

<sup>16</sup>LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

<sup>17</sup>NORCE Norwegian Research Centre, Bjerknes Center for Climate Research, Bergen, Norway

<sup>18</sup>Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, 10691, Stockholm, Sweden

<sup>19</sup>Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119333, Russia

<sup>20</sup>Climate and Global Dynamics Laboratory, National Center for Atmospheric Research (NCAR), Boulder, CO 80305, USA

<sup>21</sup>Institute of Geography, Russian Academy of Sciences, Staromonetny L. 29, Moscow, 119017, Russia

<sup>22</sup>NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

<sup>23</sup>School of Atmospheric Sciences, Nanjing University of Information Science & Technology Nanjing, 210044, China

<sup>24</sup>Atmospheric and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

Correspondence: c.brierley@ucl.ac.uk

Abstract. The mid-Holocene (6,000 years ago) is a standard time period for the evaluation of the simulated response of global climate models using paleoclimate reconstructions. The latest mid-Holocene simulations are a paleoclimate entry card for the Palaeoclimate Model Intercomparison Project (PMIP4) component of the current phase of the Coupled Model Intercomparison

Project (CMIP6). Here we provide an initial analysis and evaluation of the results of the experiment for the mid-Holocene. We

- 5 show that state-of-the-art models produce climate changes that are broadly consistent with theory and observations, including increased summer warming of the northern hemisphere and associated shifts in tropical rainfall. Many features of the PMIP4-CMIP6 simulations were present in the previous generation (PMIP3-CMIP5) of simulations. The PMIP4-CMIP6 ensemble for the mid-Holocene has a global mean temperature change of -0.3 K, which is -0.2 K cooler than the PMIP3-CMIP5 simulations predominantly as a result of the prescription of realistic greenhouse gas concentrations in PMIP4-CMIP6. Biases in the
- 10 magnitude and the sign of regional responses identified in PMIP3-CMIP5, such as the amplification of the northern African monsoon, precipitation changes over Europe and simulated aridity in mid-Eurasia, are still present in the PMIP4-CMIP6 simulations. Despite these issues, PMIP4-CMIP6 and the mid-Holocene provide an opportunity both for quantitative evaluation and derivation of emergent constraints on the hydrological cycle, feedback strength and potentially climate sensitivity.

#### Table S2

- 15 Key metrics of change in the PMIP4-CMIP6 midHolocene simulations, in either absolute terms or as a percentage of the *piControl* simulations. For comparison with the reconstructions when available, the quoted values are the average simulated at site locations only, otherwise they are area-averages. Northern high-latitude land is defined as any land between 50–80°N. Midcontinental Eurasia is defined as 40–60°N, 30–120°E by Bartlein et al. (2017). Central Asia is defined as 30–50°N, 60–75°E (Christensen et al., 2013) and these values appear in Fig. 11. The northward monsoon expansion is calculated by determining
- 20 the change in latitude where the zonal mean summer (MJJAS) rain rate (Fig. S1) equals 2 mm/day over the North Africa (15°W–30°E). The area-averaged mean annual rainfall changes are computed over 20°S –0°N, 65–45°W for South America, and 25–30°N, 70–85°E for the Indo-Gangetic Plain. ENSO activity is measured by the change in variance of monthly sea surface temperature anomalies in the Niño3.4 region (5°S–5°N, 170–120°W; Brown et al., submitted). The probability of a 50-year record in which pseudocoral ENSO activity is weak as in reconstructions for 3-5ka BP is shown separately for both the
- 25 midHolocene and piControl simulations (Emile-Geay et al., 2016). The zonal sea surface temperature (SST) gradient along the Equatorial Pacific is calculated as difference between the annual mean area average over 5°S–5°N,150–190°E and the annual mean area average over 5°S–5°N,240–270°E after Brown et al. (submitted).

#### Table S3: (Excel Spreadsheet)

Simulated Temperature Changes. The surface air temperature changes averaged in 30°latitude-wide bands are computed for
 every model included in the study. These changes are computed separately over the ocean and land as well. The annual mean SST change should closely track the surface air temperature change presented here, but can vary in regions of sea ice cover. A common land sea mask at 1° × 1° resolution is used for all models (Phillips et al., 2014). The relative weightings of land, sea and combined areas are provided to allow averages over other regions to be determined. This table is provided as a file called PMIP4-midHolocene-latband-tempchange-table.xls.

#### 35 References

45

65

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present), Journal of hydrometeorology, 4, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, 2003.

Bartlein, P. J., Harrison, S., Brewer, S., Connor, S., Davis, B., Gajewski, K., Guiot, J., Harrison-Prentice, T., Henderson, A., Pey-

40 ron, O., et al.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, Climate Dynamics, 37, 775–802, https://doi.org/10.1007/s00382-010-0904-1, 2011.

Bartlein, P. J., Harrison, S. P., and Izumi, K.: Underlying causes of Eurasian midcontinental aridity in simulations of mid-Holocene climate, Geophysical research letters, 44, 9020–9028, https://doi.org/10.1002/2017GL074476, 2017.

Boucher, O., Denvil, S., Caubel, A., and Foujols, M. A.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5229, 2018a.

- Boucher, O., Denvil, S., Caubel, A., and Foujols, M. A.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5251, 2018b.
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C., et al.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and

50 large-scale features, Climate of the Past, 3, 261–277, https://doi.org/10.5194/cp-3-261-2007, 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, Nature Climate Change, 2, 417, https://doi.org/10.1038/nclimate1456, 2012.

Brown, J., Brierley, C. M., An, S.-I., Guarino, M.-V., Stevenson, S., Williams, C. J. R., et al.: Comparison of past and future simulations of ENSO in CMIP5/PMIP3 and CMIP6/PMIP4 models, Climate of the Past, ?, ???–???, submitted.

- 55 Cao, J.: NUIST NESMv3 model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ ESGF/CMIP6.8773, 2019.
  - Cao, J. and Wang, B.: NUIST NESMv3 model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/ 10.22033/ESGF/CMIP6.8776, 2019.

Christensen et al.: Climate phenomena and their relevance for future regional climate change [IPCC WG1 AR5 Chap14], 2013.

- 60 Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10. 22033/ESGF/CMIP6.7674, 2019.
  - Danabasoglu, G., Lawrence, D., Lindsay, K., Lipscomb, W., and Strand, G.: NCAR CESM2 model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.7733, 2019.

Danek, C., Shi, X., Stepanek, C., Yang, H., Barbi, D., Hegewald, J., and Lohmann, G.: AWI AWI-ESM1.1LR model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.9335, 2020.

- EC-Earth Consortium (EC-Earth): EC-Earth-Consortium EC-Earth3-LR model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4847, 2019.
  - EC-Earth Consortium (EC-Earth): EC-Earth-Consortium EC-Earth3-LR model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.4801, 2020.

- 70 Emile-Geay, J., Cobb, K. M., Carre, M., Braconnot, P., Leloup, J., Zhou, Y., Harrison, S. P., Correge, T., McGregor, H. V., Collins, M., Driscoll, R., Elliot, M., Schneider, B., and Tudhope, A.: Links between tropical Pacific seasonal, interannual and orbital variability during the Holocene, Nature Geosci, 9, 168–173, https://doi.org/10.1038/ngeo2608, 2016.
  - Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: NCC NorESM1-F model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.11591, 2019a.
- 75 Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: NCC NorESM1-F model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.11595, 2019b.
  - Hajima, T., Abe, M., Arakawa, O., Suzuki, T., Komuro, Y., Ogura, T., Ogochi, K., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Ohgaito, R., Ito, A., Yamazaki, D., Ito, A., Takata, K., Watanabe, S., Kawamiya, M., and Tachiiri, K.: MIROC MIROC-ES2L model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5710, 2019.
- 80 Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A.: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene, Climate of the Past, 9, 811–823, https://doi.org/10.5194/cp-9-811-2013, 2013.
  - Joussaume, S., Taylor, K., Braconnot, P., Mitchell, J., Kutzbach, J., Harrison, S., Prentice, I., Broccoli, A., Abe-Ouchi, A., Bartlein, P., et al.: Monsoon changes for 6000 years ago: results of 18 simulations from the Paleoclimate Modeling Intercomparison Project (PMIP), Geophysical Research Letters, 26, 859–862, https://doi.org/10.1029/1999GL900126, 1999.
- 85 Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., et al.: A global database of Holocene paleo-temperature records, Scientific Data, 7, https://doi.org/10.1038/s41597-020-0445-3, 2020.
  - Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P., Heiri, O., and Davis, B.: Holocene global mean surface temperature: a multi-method reconstruction approach, Scientific Data, in press.

Li, L.: CAS FGOALS-g3 model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/

90 CMIP6.3448, 2019.

100

- NASA Goddard Institute for Space Studies (NASA/GISS): NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.7380, 2018.
- NASA Goddard Institute for Space Studies (NASA/GISS): NASA-GISS GISS-E2.1G model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.7225, 2019.
- 95 New, M., Hulme, M., and Jones, P.: Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate, Journal of Climate, 13, 2217–2238, https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2, 2000.
  - Ohgaito, R., Abe-Ouchi, A., Abe, M., Arakawa, O., Ogochi, K., Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Ito, A., Yamazaki, D., Ito, A., Takata, K., Watanabe, S., Kawamiya, M., and Tachiiri, K.: MIROC MIROC-ES2L model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.5646, 2019.
- Phillips, A. S., Deser, C., and Fasullo, J.: Evaluating modes of variability in climate models, Eos, Transactions American Geophysical Union, 95, 453–455, https://doi.org/10.1002/2014EO490002, 2014.
  - Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., and Andrews, T.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6294, 2018.
- 105 Seland, o., Bentsen, M., Oliviè, D. J. L., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O. A., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: NCC NorESM2-

LM model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.8217, 2019.

- 110 Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/ 10.1029/2000JD900719, 2001.
  - Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM4-8 model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10. 22033/ESGF/CMIP6.5077, 2019a.
- 115 Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM4-8 model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10. 22033/ESGF/CMIP6.5080, 2019b.
  - Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, https://doi.org/10.1175/2008JAMC1921.1, 1997.
- 120 Yu, Y.: CAS FGOALS-f3-L model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ ESGF/CMIP6.3447, 2019.
  - Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6860, 2019a.
- 125 Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP piControl, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.6900, 2019b.
  - Zhang, Z., Bentsen, M., Oliviè, D. J. L., Seland, o., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H.,
- 130 Landgren, O. A., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Heinze, C., Iversen, T., and Schulz, M.: NCC NorESM2-LM model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/ CMIP6.8079, 2019.
  - Zheng, W. and Dong, L.: CAS FGOALS-g3 model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https://doi.org/10.22033/ESGF/CMIP6.3409, 2019.
- 135 Zheng, W. and He, B.: CAS FGOALS-f3-L model output prepared for CMIP6 PMIP midHolocene, Earth System Grid Federation, https: //doi.org/10.22033/ESGF/CMIP6.12014, 2019.

Table S1. Key metrics of change in the PMIP4-CMIP6 *midHolocene* simulations Digital Object Identifier (doi) for each simulation from CMIP6 and CMIP5. Should the hyperlinks in the table not work, the web address can be created manually by adding https://dx.doi.org/in front of each doi. The full citations are in the References.

model	midHolocene	piControl
AWI-ESM-1-1-LR	N/A	10.22033/ESGF/CMIP6.9335
CESM2	10.22033/ESGF/CMIP6.7674	10.22033/ESGF/CMIP6.7733
EC-Earth3-LR	10.22033/ESGF/CMIP6.4847	10.22033/ESGF/CMIP6.4801
FGOALS-f3-L	10.22033/ESGF/CMIP6.12014	10.22033/ESGF/CMIP6.3447
FGOALS-g3	10.22033/ESGF/CMIP6.3409	10.22033/ESGF/CMIP6.3448
GISS-E2-1-G	10.22033/ESGF/CMIP6.7225	10.22033/ESGF/CMIP6.7380
HadGEM3-GC31-LL	N/A	10.22033/ESGF/CMIP6.6294
INM-CM4-8	10.22033/ESGF/CMIP6.5077	10.22033/ESGF/CMIP6.5080
IPSL-CM6A-LR	10.22033/ESGF/CMIP6.5229	10.22033/ESGF/CMIP6.5251
MIROC-ES2L	10.22033/ESGF/CMIP6.5646	10.22033/ESGF/CMIP6.5710
MRI-ESM2	10.22033/ESGF/CMIP6.6860	10.22033/ESGF/CMIP6.6900
NESM3	10.22033/ESGF/CMIP6.8773	10.22033/ESGF/CMIP6.8776
NorESM1-F	10.22033/ESGF/CMIP6.11591	10.22033/ESGF/CMIP6.11595
NorESM2-LM	10.22033/ESGF/CMIP6.8079	10.22033/ESGF/CMIP6.8217
<u>UofT-CCSM-4</u>	N/A	N/A
bcc-csm1-1	10.1594/WDCC/CMIP5.BCB1mh	10.1594/WDCC/CMIP5.BCB1pc
CCSM4	10.1594/WDCC/CMIP5.NRS4mh	10.1594/WDCC/CMIP5.NRS4pc
CNRM-CM5	10.1594/WDCC/CMIP5.CEC5mh	10.1594/WDCC/CMIP5.CEC5pc
CSIRO-MK3-6-0	10.1594/WDCC/CMIP5.CQMKmh	10.1594/WDCC/CMIP5.CQMKpc
CSIRO-MK3L-1-2	N/A	N/A
EC-Earth-2-2	N/A	N/A
FGOALS-G2	10.1594/WDCC/CMIP5.LSF2mh	10.1594/WDCC/CMIP5.LSF2pc
FGOALS-S2	10.1594/WDCC/CMIP5.LIFSmh	10.1594/WDCC/CMIP5.LIFSpc
GISS-E2-R	10.1594/WDCC/CMIP5.GIGRmh	10.1594/WDCC/CMIP5.GIGRpc
HadGEM2-CC	10.1594/WDCC/CMIP5.MOGCmh	10.1594/WDCC/CMIP5.MOGCpc
HadGEM2-ES	10.1594/WDCC/CMIP5.MOGEmh	10.1594/WDCC/CMIP5.MOGEpc
IPSL-CM5A-LR	10.1594/WDCC/CMIP5.IPILmh	10.1594/WDCC/CMIP5.IPILpc
MIROC-ESM	10.1594/WDCC/CMIP5.MIMEmh	10.1594/WDCC/CMIP5.MIMEpc
MPI-ESM-P	10.1594/WDCC/CMIP5.MXEPmh	10.1594/WDCC/CMIP5.MXEPpc
MRI-CGCM3	10.1594/WDCC/CMIP5.MRMCmh	10.1594/WDCC/CMIP5.MRMCpc

N/A indicates that a doi is not available.

#### Table S2. Key metrics of change in the PMIP4-CMIP6 midHolocene simulations (see above for further details)

	Extratropical						Tropical						
	Global mean temperature (°C)	Summer NH high-lat. land (°C)	Drier Eastern North America (mm/yr)	Midcontinental Eurasia rainfall (mm/yr)	Midcontinental Eurasia Seasonality (°C)	Central Asian Seasonality (°C)	N. African monsoon expansion (°N)	Drier South America (mm/yr)	Indo-Gangetic rainfall (mm/yr)	Niño3.4 Variance‡ (%)	p(suppressed ENSO) in <i>piControl</i> § (%)	p(suppressed ENSO) in <i>midHolocene</i> § (%)	
AWI-ESM-1-1-LR	- <del>0.5</del> -0.4	<del>0.2</del> 0.0	-44-58	<del>-25</del> -21	2.6	2.9	<del>3.2</del> 3.1	<del>140</del> 68	<del>82</del> 115	-41			
CESM2	<del>-0.1</del> -0.2	<del>1.1</del> 0.7	<del>-51</del> -54	-16	2.8	3.1	<del>2.7</del> -0.2	<del>-95</del> -97	<del>92</del> 125	-16	2.4	5.7	
	-0.1	1.8	-28	12	2.3	2.3	-0.5	-29	166	-31	-~	~	
EC-Earth3-LR													
FGOALS-f3-L	-0.4	1.10.6	<b>-6</b> -24	-18-11	3.0	3.0	<del>1.9</del> <u>1.5</u>	<del>-91_85</del>	<del>170<u>165</u></del>	4	-2.8	- <u>1.2</u>	
FGOALS-g3	-0.2	<del>2.31.1</del>	<del>-34-92</del>	- <del>67</del> - <u>58</u>	4.2	4.1	<b>1.11.8</b>	-261-258	4 <u>257</u>	-14	-0.2	-2.5	
GISS-E2-1-G	-0.4	<del>1.10.7</del>	<del>-14</del> - <u>15</u>	<del>-8</del> - <u>9</u>	2.4	2.6	1.6	<del>-61</del> - <u>60</u>	<del>170<u>188</u></del>	2	1.8	5.6	
HadGEM2 CC21HadC	-0.1	1.2	<del>16.9</del>	- <del>21</del> 2	3.0	3.8	<del>3.9</del> 2.3	<del>-103</del> -102	<del>193</del> 207	-8	-0.6	-0.6	
	0.2		24	2	27	2.1	2.1	06	211212	7	1.5	14.2	
INM-CM4-8	-0.5	1.10.5	12.22	-2	2.7	3.1	2.1	-90	159160	12	1.3	14.2	
MIDOC ES2	-0.4		-12-23	-27-32	2.5	3.0	0.01.2	07.111	10977	-15	1./	5.4 82.4	
MIROC-E32E	-0.5	0.50.0	26	-20	2.0	3.4	37	170	180	-49	1.5	7.4	
MPI-ESM1-2-LR	-0.4		~~~~~	~~~~~	~~~~~	~~~~		-1/9		~~~~~		/	
MRI-ESM2-0	-0.2	<del>1.0</del> 0.7	<del>-24</del> -22	<del>-16</del> - <u>15</u>	2.5	2.7	<del>1.3</del> 3.3	<del>-178</del> -179	<del>176</del> 189	-36	-4.8	-34.5	
NESM3	-0.3	1.10.9	<del>61</del> 59	<del>22</del> 24	2.6	2.5	<del>3.7</del> 3.1	<del>-182</del> -155	<del>-161</del> 177	-24	2.1	5.2	
NorESM1-F	-0.4	0.90.4	<del>6</del> -6	<del>-19_8</del>	3.4	3.6	<del>1.5</del> 1.4	<del>-118</del> -116	<del>149</del> 158	-6	_	_	
	-0.2	0.5	137	137	3.3	3.0	-1.9	-85	255	11	-~	~	
NorESM2-LM													
UofT-CCSM-4	-0.2	1. <u>61.1</u>	-10-8	<b>10-3</b>	3.1	2.8	0.41.9	- <del>326</del> -117	27114	-48	-	-	
Reconstructed	0.5†	0.7*	-93¶	121¶	-	-		-	-			-	
PMIP4 Average	-0.3	1.10.8	-4-7	<b>-16-3</b>	3.02.9	3.03.1	<del>1.9</del> <u>1.7</u>	<del>-118_9</del> 9	<del>133<u>162</u></del>	<b>-19-18</b>	-2.4	-14.8	
PMIP3 Average	-0.1	1. <u>31</u> .	-11-10	-5-4	2.6	2.9	<del>1.8</del> <u>3.0</u>	-83	<del>166</del> 175	-11	<b>-</b> <u>3.7§</u>	<b>-</b> 5.8§	
PMIP3 Spread	0.2	0.40.5	<del>1619</del>	15	0.4	0.4	<del>2.14.2</del>	<b>4446</b>	<b>75</b> 81	14	<b>-</b> <u>3.2§</u>	-4.3§	

†Median reconstructed global mean value from Kaufman et al. (in press), with 80% confidence interval of 0.3–0.9 °C. \*average of the difference in summer and winter reconstructions within the re Kaufman et al. (2020) compilation. ¶average of reconstructions within the region from Bartlein et al. (2011) compilation. ‡Values published in Brown et al. (submitted). §Using the analysis approa Emile-Geay et al. (2016) with PMIP3 values directly from it. 7



Figure S1. Simulated North African monsoon through multiple phases of PMIP-CMIP. (top panel) Biome distributions (desert, steppe, xerophytic and savannah/dry tropical forest) as a function of latitude for present (red circles) and 6 ka (green triangles), showing that steppe vegetation replaces desert at 6 ka as far north as 23°N (middle panel) Annual mean precipitation changes (mm/yr) over Africa (20°W–30°E) for the Mid-Holocene climate across multiple PMIP generations. The black hatched lines are estimated upper and lower bounds for the excess additional precipitation required to support grassland, steppe at each latitude during the mid-Holocene based on present water-balance modelling and the modern climatic limits of requirements for desert and grassland taxa in palaeo-ecological recordsplants. (bottom panel) The rainfall distribution in piControl simulations for each model. Three different observationally-based datasets are shown in black: GPCP (Adler et al., 2003), CMAP (Xie and Arkin, 1997), and CRU (New et al., 2000). (Adapted from Joussaume et al., 1999; Braconnot et al., 2007, 2012)



**Figure S2. Statistical description of site-level comparison of simulated mid-Holocene climate changes to reconstructions.** The performance of both the CMIP6 and CMIP5 ensembles are assessed by comparing the annual mean temperature changes and difference between summer mean temperature changes and winter mean temperature changes to multi-proxy Temperature 12k database (red, green; Kaufman et al., 2020) and mean annual precipitation and difference between mean temperature fo temperature of the warmest month (MTWA) changes and mean temperature of the coldest month (MTCO) changes to the pollen-based reconstructions (yellow, blue, purple; Bartlein et al., 2011). The better a model's changes fit with the reconstructions, then closer it should be to the green square (Taylor, 2001). The correlation coefficient is plotted on the azimuth, and the standard error determines the radial distance (Taylor, 2001). The radial distance is adjusted to account for presents the existence ratio of uncertainty the standard deviation in the model and reconstructions (after Hargreaves et al., 2013). (after adjustment to account for the existence of uncertainty in them Hargreaves et al., 2013).



#### Data-model comparison summary CMIP5-PMIP3 vs CMIP6-PMIP4

Figure S3. Alternate presentation of the data-model comparison. Regional comparisons using Monte-Carlo sampling of both the reconstruction uncertainty (Bartlein et al., 2011) and model uncertainty as expressed by interannual variability at individual proxy locations. The regions are defined as Europe  $(35-70^{\circ}N, 10^{\circ}W-30^{\circ}E)$ , West Africa  $(0-30^{\circ}N, 30^{\circ}W-30^{\circ}E)$  and North America  $(20-50^{\circ}N, 140-60^{\circ}W)$ .

Note: Table S3 is provided as an external spreadsheet called PMIP4-midHolocene-latband-tempchange-table.x1s