To whom it may concern,

Re: Response letter to Reviewers, after submission of manuscript "*The UK contribution to CMIP6/PMIP4: mid-Holocene and Last Interglacial experiments with HadGEM3, and comparison to the pre-industrial era and proxy data*" by CJR Williams *et al*. to Climate of the Past (PMIP4 Special Edition).

Thank you very much for informing me that the discussion period for the above manuscript is now over. Thank you also for the opportunity to resubmit a revised manuscript, according to the reviewer's comments, and for the extension granted to me.

I extend my sincere appreciation to the reviewers for their thorough examination of my manuscript, and their detailed and highly constructive comments. I propose to address all of their concerns, both minor and major, so please see attached for a revised manuscript, still with the Track Changes included, to show my proposed modifications. I also attach a tidy version of this manuscript. Please note that the line numbers shown here relate to the Track Changes version.

Here, I address the reviewers' suggestions, comment-by-comment. In the following, the reviewers' comments are italicised and in a smaller font, and my corresponding response follows in a standard font.

I very much hope that my responses will satisfy the reviewers and meet your expectations, and therefore request you to consider our revised manuscript for publication in your Journal.

Yours faithfully,

Dr Charles JR Williams, and co-authors

REVIEWER 1

MAIN CONCERNS:

As mentioned above, it appears to me that the manuscript presents various data sets and experiments that in its current form lack relevance or seem somewhat out of place. This makes that the manuscript lacks a clear common thread, making the manuscript difficult to read. Clarify the relevance of all the presented results and consistently mention all of them in the abstract, introduction, results and conclusion sections. Clarify why you present: a comparison to previous versions of the model; a comparison to CMIP/PMIP results; a comparison with proxy-data. A change to the overall structure of the manuscript as detailed in the following could also improve the flow of the manuscript.

The manuscript has now been restructured. In short, the section on the spin-up results has been reduced and moved into the Methodology section (with the figures included in the Supplementary Material), and the Results section has been restructured to focus primarily on two measures: firstly a simulation comparison to assess changes relative to the same model's PI simulation, and secondly a model-data comparison in which the most recent version of the simulations are compared to both previous generations of the same model and newly-available proxy data. Importantly, all versions of the same model are now compared against all available proxy data. The section focusing on the Saharan greening has been reduced and is now discussed only briefly at the end, with the figure being moved into the Supplementary Material. Following the reviewer's comment above, this new structure is consistently mentioned in the abstract, introduction, results and conclusion sections.

The analysis of temperature on the one hand and precipitation on the other hand are very different. Temperatures are looked at globally, while the analysis of precipitation is solely on Africa. Clarify to the reader why this choice was made. The structure of the manuscript would potentially be improved if the analysis is firstly on global features (mostly temperature, but perhaps also precipitation considering the newly developed Last Interglacial precipitation reconstruction), and secondly zooms in on African precipitation changes.

This structure has now been modified, with an additional figure showing global precipitation changes (Figure 3) as well as appropriate discussion (lines 521-534) so that, when discussing the most recent simulations, both temperature and precipitation are considered at the global scale, before zooming into Africa as an example of monsoon changes.

An extensive description of the spin-up results is given in the manuscript. While in general I appreciate it that such potentially important modelling details are given, the relevance to the rest of the manuscript is not clear to me. Wouldn't it be sufficient to provide the numbers of the spin-up results in a table and refer to that table in the method section? Is a whole results section for the spin-up results needed, given also that they are not referred to anymore in the remainder of the analysis?

This section has now been significantly reduced, and moved to a subsection of the Methodology section (lines 394-429). The results are summarised in Table 2 and referred to here, and the figures (showing examples of timeseries for atmospheric and oceanic equilibrium) have been moved into the Supplementary Material (SM2 and SM4).

The model-data comparison is limited to temperature proxies while recently a new compilation of precipitation reconstructions for the Last Interglacial has become available (Scussolini et al., 2019) which should be used here as well.

This compilation has now been included in the manuscript, with a new figure (Figure 10) as well as new text (lines 769-790), which we feel further adds to the novelty of the manuscript.

I find the whole analysis of African precipitation changes in the different simulations incomplete and too simplistic, leading to figures and results that are misleading.
To improve this situation I suggest the following:
-> Remove the ocean grid cells from the domain over which the analysis is performed. Presenting zonal-mean figures is not appropriate if the results are clearly not zonally homogeneous as is the case around the equator in this analysis.
-> Why is the focus on JJA precipitation over Africa? The results in figure 12 are again annual mean. Be consistent and clarify your reasoning.
-> The authors seem to use the words 'monsoon' and 'ITCZ' interchangeably. While indeed they cannot easily be separated based solely on the analysis of precipitation, they are driven by fundamentally different processes and an attempt should be made to separate the two. An interesting read on this topic is by Nicholson (2009).
-> The results for African precipitation are rather different for the different periods that are considered and for the various versions of the climate model. Provide some analysis as to what drives these differences.

All of the above points have now been addressed:
-> The zonal mean figure has now had all ocean grid cells removed

-> An explanation is given for why JJA precipitation is considered over Africa (i.e. because it is the primary wet season) versus annual precipitation in figure 12 (which has now been moved to the Supplementary Material) i.e. to be consistent with the proxy data-based threshold (lines 546-551) and (lines 926-927)
-> The confusion between the words "monsoon" and "ITCZ" has been clarified, by sticking to terms such as "rainfall maxima" or "rainbelt" (lines 555-565)
-> The reason for the difference in African precipitation between the different periods has been made more clear (lines 488-603), and a possible reason for one of the differences between versions of the model has been added (lines 817-820).

In many figures no measure of the robustness of the results is given. Provide measures to determine if your results are significant or if we are looking at internal variability of the climate system. A possibility would be to use long-term variability within the different simulations (PI, 6k, 127k) to deduce whether the depicted anomalies are outside of the range of this variability.

This has now been addressed, with the global temperature and precipitation figures of the most recent simulations (Figure 2 and Figure 3) now showing the results of a Students t-test (at the 99% confidence level). The same measure of statistical significance was not included when comparing the most recent simulations with previous versions of the same model and proxy data because stippling would make the proxy data locations harder to visualise.

MINOR COMMENTS:

Line 48: are surface (0 meter) temperatures discussed in the manuscript? Or do the authors mean SSTs here?

This line has now been addressed, to differentiate between 1.5m air temperature and SST, both of which are used in different places (lines 49-69)

Line 82: these periods are not 'warm' everywhere and all the time. Please clarify.

This has now been clarified, emphasising that they are two interglacial simulations representing the most recent warm periods (particularly in the Northern Hemisphere) (lines 106)

Line 83: 129-116 is not the period that is discussed in this manuscript. Why not simply 127ka?

This has now been clarified (lines 107)

Lines 108-112: Clarify that this warming is mainly located at high latitudes.

This has now been corrected (lines 169)

Line 126: convection in the atmosphere or ocean or both?

This has now been clarified, to confirm that it should be atmospheric convection (lines 190)

Line 127: are the ocean and sea ice models completely new or have parts been updated?

This has now been corrected, to confirm that these models are not completely new, but rather updated (lines 191)

Line 140-141: This division in two subsections (3.1 and 3.2) suggest to me that the two topics are of similar importance while in reality this is certainly not the case, with the results on the spin-up phase being only a small side topic. Consider changing this structure to better represent the importance of the different topics.

This is now been corrected as per the comment above, with significantly less emphasis being given to the spin up section (which is now in the Methodology section and summarised in Table 2, when the figures in the Supplementary Material (see above comment) (lines 394-429)

Line 141: For me the term productions runs is a little strange, perhaps it is CMIP kind of language, but in the context of a manuscript is doesn't mean much to me.

This has been removed at this line, and has been clarified when used elsewhere by a new Terminology subsection (section 2.1.1) (lines 216-226)

Line 214: Isn't precipitation impacted by ENSO?

This reference to ENSO has been removed

Lines 248-252: I don't think such details (number of output variables) are relevant for a manuscript.

These have been removed

Line 273: what is your definition of 'summer'?

This has been clarified, to confirm 'summer' here = July September (lines 362)

Line 276: You constructed this 127 ka time-slice of the Hoffman et al data? Do you provide this data for future work?

Yes, that is correct. A sentence has been added into the Data availability section, detailing access (lines 1014-1016)

Line 306: Is a trend of 0.16 degrees per century small? Sounds significant to me. Please clarify.

The reviewer is correct, so the ambiguous word "small" has been removed

Line 335: "the current two warm climate", what does that mean or refer to?

This term has now been clarified in the new Terminology subsection (section 2.1.1), see above comment

Line 335: Which newly-available proxy data are you referring to? Did you gather new data? Or do you mean the 127 ka time slice based on the Hoffman et al data?

This has now been clarified (lines 477)

Line 342: HadGEM3 warm climate simulations?

This has now been clarified (lines 488)

Line 364: 30 degrees east doesn't sound like west African to me. Please clarify why this domain was chosen, also in light of my main concern on this topic.

The domain used to calculate this zonal mean has now been changed, to become 20°W-15°E, with appropriate clarification text (lines 573)

Line 369: The wind patterns to me show an increase at nearly all latitudes, is that typical for an ITCZ shift?

Reference to the ITCZ has now been removed, instead referring to the regions of rainfall maxima, and an appropriate sentence has been added to clarify that the enhanced wind patterns do indeed occur at all latitudes, but especially over regional rainfall maxima (lines 558)

Lines 374-377: Do we see the same kind of pattern to the south of the equator, so the South African region?

This sentence has now been removed, so is no longer problematic

Line 407: Do proxies suggest a global annual mean warming during the MH?

This has now been addressed, to clarify that we don't see a global annual mean warming from proxies, but rather do see warming in many locations (lines 628)

Line 439: "within the average uncertainty range'? Please clarify this statement.

As part of other changes, this sentence has now been removed, as it was ambiguous

Line 466: The model is seasonally dependent? What does that mean? Do you mean the comparison of models and data?

This sentence has now been clarified, to confirm that it should read "the accuracy of the model is seasonally dependent" (lines 799)

Lines 488-492: Why would you compare your results to results from previous model version to see if you get sufficient precipitation over the Sahara to promote vegetation growth?

This entire paragraph has now been removed as part of other changes, so is no longer problematic

Lines 515-529: These kind of detailed (small) differences make me wonder whether we are really discussing forced differences or if we are discussing internal variability of the system. Please show statistics to argue either way.

This entire paragraph has now been removed as part of other changes, so is no longer problematic. However, as detailed above, significance testing has now been carried out on the most recent warm climate simulations

Lines 546-553: When you are talking about an 'improvement' this suggest that we know what 'good' means. What kind of data or proxies do you use to determine 'good' and what is the uncertainty of these estimates?

This entire paragraph has now been removed as part of other changes, so is no longer problematic. However, as detailed above, all model versions have now be compared against all available proxy data, allowing a quantitative and qualitative determination of 'good'

Line 555: Aren't the paragraphs before already discussing "rainfall across the Sahara"?

Yes, they are. This is therefore now been changed to a more appropriate title (lines 909)

Lines 570-583: What is the relationship between vegetation in the Sahara (the topic of this paragraph) and the state of the equatorial Atlantic ('drying'?)? Please clarify.

As part of other changes, this paragraph has been rewritten and shortened (and indeed the figure has been moved into the Supplementary Material), and therefore this sentence no longer exists

Lines 570-583: Not only is a vegetation model missing to directly determine whether or not vegetation would grow with the simulated amount of precipitation, but also all vegetation related feedbacks on the climate are missing. Discuss the possible impact of these missing feedbacks on you results.

As part of other changes, this paragraph has been rewritten and shortened. However, an additional sentence has been added at the end of this paragraph, briefly discussing the current lack of vegetation-related feedbacks (lines 933-935)

Lines 616-619: meltwater does not only yield a warming, it usually results in a spatially varying pattern with regions of warming and regions of cooling. Please clarify.

This has been clarified, to reflect the accurate comments of the reviewer (lines 986-987)

Lines 619-621: Is the length of the spin-up really a potentially important caveat? Do you have evidence to support this?

As part of other changes, this sentence has now been removed

Line 628: Only MH or both MH and 127ka?

This has been clarified (lines 997)

Table 2: If some values are for the full ocean depth and others for the top 1054 meter, can we still compare them? Isn't it comparing apples and oranges?

This has been corrected, such that all the values are for the full ocean depth.

Table 4: I appreciate the attempt to provide a lot of information, but I find this table very confusing. Perhaps it can be split or rearranged?

This has now been corrected, with the table being split into Table 4a and b

Figure 1: Have calendar effects been taken into account when making this figure? Please apply corrections, following for instance the methodology outlined by Bartlein et al. (2019).

Calendar adjustments, both this figure and all subsequent figures involving monthly or seasonal data, have now been applied, following the methodology of Pollard & Reusch (2002) and Marzocchi et al. (2015). This is briefly discussed in the introduction (lines 118-121), with examples of the data on the modern calendar (for comparative purposes) included in the Supplementary Material (SM1).

Figure 2: There seems to be a gap between the control data and the start of the 127k simulation, is this a real data gap or an error in the figure?

This has now been moved to the Supplementary Material (SM2). However, to answer the reviewer: yes, there is a purposeful gap between the end of the control data and the start of the LIG simulation, because a number of model crashes caused the first ~50 years of the spin-up to be unstable giving highly varied global mean temperatures. This is briefly noted in the figure caption (SM2)

Figure 2: Are the temperatures in the left-hand figure surface or 1.5 meter temperatures?

This has now been moved to the Supplementary Material (SM2). However, to answer the reviewer: these are 1.5m air temperatures, and this has now been clarified in the figure

Figure 2: Consider: 'b) TOA radiation balance'

This has now been moved to the Supplementary Material (SM2). However, to answer the reviewer: yes, this has now been corrected

Figure 2: This figure gives a good idea of the amount of internal variability in the system, which seems considerable in both the MH and lig127k simulations. Use this information to define which of your results are robust with respect to this internal variability. Is it true that variability is larger in the 'warm climates' than it is in de control?

We now use a Student's test (at the 99% level) as a matter of significant or robustness, which accounts for the interannual variability. Yes, it is true that variability is larger in the warm climate simulations than the PI, and this has been briefly noted in the text (lines 402)

Figure 3: For the control simulation the full depth is used instead of the top 1054 meters according to the main text, please clarify.

This has now been corrected (see comment above)

Figure 4: These figures show some well-know climate change features, including polar amplification. The mechanisms of such spatial temperature anomaly patterns are not discussed. Provide a discussion or refer to previous work on the topic.

This now been addressed, with a short discussion on one of the mechanisms of polar amplification, namely sea-ice interactions, has now been added (lines 515-519), along with an accompanying figure in the Supplementary Material (SM6)

Figure 6: Rainfall anomalies on y-axis?

As part of other changes, this figure has now been removed. However, in subsequent zonal mean figures, the y-axis label has been changed from simply "Rainfall" to "Rainfall anomalies" (relative to the PI), which is what we understand the reviewer to mean here

Figure 6: x-axis values are not easy to read in this format.

As part of the changes, this figure has been removed

Figure 6: consider showing absolute precipitation values because I think those give a much better idea of the width of the wet and dry regions as discussed in the main text.

As part of the changes, this figure has been removed. However, when zonal mean rainfall is shown, both anomalies and absolute values are now shown (Figure 6)

Figure 6: Can't this figure be combined with figure 9?

This has now been done

Figure 8: Remove the ice core data points if the corresponding modeled surface temperature anomalies are not shown.

This has now been done

Figure 9: What does this figure add that is not already depicted in figures 10 and 11? Can't it be removed?

It was decided to remove existing figures 10 and 11 instead, because the same information is shown at the global scale in the new Figure 8 and Figure 10

Figure 12: why are the grey dashes that show required rainfall for grassland growth only start from 16 degrees north?

This has now been moved to the Supplementary Material (SM7), and it has been modified so that the model latitudes begin approximately where the rainfall threshold needed for grassland (the grey dashes) begins. To directly answer the reviewer, the grey dashes in this figure were taken directly from Figure 3a in Joussaume *et al*. (1999), which only has data beginning at 15.5°N. This reference has been added to the figure

Figure 12: Rainfall anomalies on y-axis?

This has now been corrected (see above comment)

Figure 12: Why are anomalies shown? Doesn't the threshold to support grassland depend on the absolute amount of precipitation?

Anomalies, rather than absolute values, are shown because this is following the methodology of Joussaume *et al*. (1999), who also considered annual mean anomalies in their study. Likewise, the region of averaging is larger here than Figure 6 (up to 30°E, as opposed to 15°E), again to be consistent with Joussaume *et al*. (1999). This has been made more clear in the text (lines 926)

TECHNICAL COMMENTS:

Line 41: are similar

This has been corrected

Line 41: period, but

This has been corrected

Line 53: generations of the same

As part of other changes, this sentence has now been removed

Line 121: therefore in the

This has been corrected

Line 146: consider removing "indeed"

This has been corrected

Line 149: double space before "Full"?

This has been corrected

Line 186: tuning of

This has been corrected

Line 201: including a reduction of the temperature bias in many regions

This has been corrected

Line 221: remove comma after 'design'

This has been corrected

Line 244: Too many brackets

This has been corrected

Line 272: annual mean surface

This has been corrected

Line 298: radiation balance?

As part of other changes, this sentence has now been removed

Line 384: in the core monsoon region?

As part of other changes, this has now been removed

Line 395: 'recent', what do you mean?

This has now been corrected

Line 399: what kind of uncertainty in simulated anomalies are you referring to, please clarify.

This has now been corrected

Line 400: remove 'often'

This has now been corrected

Line 437: small number of reconstructions?

As part of other changes, this has now been removed

Line 449: remove double comma

As part of other changes, this has now been removed

Line 499: refer to figure 10?

As part of other changes, this has now been removed

Line 505: smaller northward displacement?

As part of other changes, this has now been removed

Line 590: 'auspices', not sure if that is the right word for it.

This has now been corrected

Line 590: replace comma by a dot.

This has now been corrected

Line 591: remove 'time'?

This has now been corrected

Line 592: were assessed?

This has now been corrected

Line 603: 'time', are we talking seasons or different geological intervals?

This has now been corrected

Line 626: 'necessity' is perhaps a bit too strong in this context.

This has now been corrected

Line 1007: better not to use the & symbol.

This has now been corrected

Line 1008: for each

As part of other changes, this has now been removed

Line 1018: in this caption and some others the words 'simulated gridded anomalies' are used. This sounds a little double to me since nearly all climate models work on spatial grids so the output is per definition also gridded.

This has now been corrected

Lines 1024-1027: Is there no overlap between these two data sets? No single core was used in both of them?

As discussed in the Methodology section (section 2.3), the two datasets use different reference chronologies and methodologies to infer temporal surface temperature changes. Whilst they may use the same core, the methodologies are very different, and therefore they should not be combined

Line 1033: erroneous bracket?

As part of other changes, this has now been removed

REVIEWER 2

MAJOR COMMENTS:

1. It shall be elaborated what is new in this paper in terms of method, result and conclusion as compared to previous studies. Data model comparison in SST data and the question of seasonality could be more elaborated. It is not understandable that the SST comparison has not been performed on the MH experiment, although the data quality is higher and especially the dating uncertainty is much lower. Uncertainty is mentioned quite often, but not really elaborated. For the LIG, one could follow ideas outlined in Pfeiffer and Lohmann (CP) dealing with seasons and dating. For the MH, several data sets are available (e.g. Alkenone and Mg/Ca), again with uncertainties in the season or recorder depth.

Sentences have been added throughout the manuscript, including the abstract, introduction, methodology and conclusion, to elaborate and emphasise the novelty of this study. In short, we explain that although older versions of the UK model have been included in previous iterations of CMIP, and although present-day and future simulations from this model are included in CMIP6, this study is completely new because it is the first time this version of the model has been used to conduct any paleoclimate simulations. Give that these paleoclimate periods are out-of-sample in that they were not used in any way to tune or develop this model, these simulations provide a critical independent evaluation of the model's strengths and weaknesses.

Regarding the comment about including a model-data comparison with mid-Holocene SST data, we note that the manuscript already contains 5 separate datasets: i) land-surface temperature from the mid-Holocene, ii) land-surface precipitation from the mid-Holocene, iii) SST from the Last Interglacial (from 2 separate sources), and iv) precipitation from the Last Interglacial. The manuscript is already quite long, and we feel that the addition of more mid-Holocene SST data would not bring added information to the study. Moreover, another study (involving many of the co-authors here) is currently under review, looking specifically at the large-scale features during the mid-Holocene from CMIP6 models, including ours: Brierley et al. (2020). 'Large-scale features and evaluation of the PMIP4-CMIP6 midHolocene simulations'. *Clim. Past*. Under review. That study includes a significant model-data comparison section, including Holocene SSTs, and therefore we propose not to add any more model-data comparison in this paper, but rather to direct the reader on (lines 352-354).

2. The paper is too descriptive and focuses only on simulated temperature and precipitation. As a special contribution to CMIP6/PMIP4 is based on a single model, I would expect more comprehensive analysis, like the atmospheric and oceanic circulation, ocean states, and the potential relationship or mechanisms between different components. With such I believe the paper will meet the high standard of CP.

An example of atmospheric circulation changes (Figure 5) was already included in the original version of the manuscript, but this has now been elaborated in the text (lines 546-565). Moreover, in agreement with the reviewer, a measure of oceanic circulation has now been added (Figure 4), namely an example of the meridional overturning circulation. We find that there is almost no change in ocean circulation between the PI, mid-Holocene and LIG simulations.

3. The authors show precipitation only for Africa. As a paper contributing to the CMIP6/PMIP4, it shall show the model behavior on global rather than regional scale.

This has now been done, showing both precipitation and temperature at both global and Africa-wide scales

MORE SPECIFIC COMMENTS:

1. Lines 94-104: This paragraph describes the previous studies on the modeled and observed MH and LIG states, which I find is too brief. As there are so many modelling studies and proxy papers, and this is directly linked to the present manuscript, thus I suggest to make more complete references. It is suggested to split the texts into two paragraphs, one describing the previous simulation results, the other the proxy issues.

This has now been addressed, with this section being expanded to include a number of other studies, as well as being divided into firstly a paragraph on the proxy data, and then a paragraph on modelling studies (lines 213-155)

2. Lines 106-108: The authors mention that the past warming are indeed different from future warming, as they are driven by quite different thermal forcing mechanisms, orbital parameters and greenhouse gases. I suggest to also mention that, i) the orbital forcing is shortwave and greenhouse gases are related to mainly the longwave radiation flux, ii) difference in orbital parameters leads to uneven horizontal and seasonal changes, but greenhouse gases can cause more uniform anomalies. Furthermore: iii) It is helpful to know the changes of greenhouse gases between MH/LIG and PI are equal to how much radiation flux anomalies? How to calculate such anomalies based on CO2 changes can be found in some papers (e.g., Myhre, et al. 1998, GRL).

The first two elements have now been incorporated into the text (lines 163-165). Regarding the 3rd point, we feel that providing the precise calculation of the radiative forcing due to changes in MH and LIG greenhouse gases is beyond the scope of this study, and would not provide a great deal of added information. However, this has been acknowledged and clarified in the text, with a reference to the Gunnar et al. (1998) study (lines 165-169).

3. Lines 161-203 Too detailed information in terms of the changes in model version is give here. I would recommend to simplify the text and to show what aspect/process can be improved in the newest model version. Details could be provided as supplementary material.

This has now been done, with much of the text being transferred to the Supplementary Material

4. Lines 205-209: The sensitivity and control experiments are performed on different platforms. I worry about how different the simulated climate can be. If possible, one shall show in the supplement the anomalies of surface temperature based on the same experiment

This issue is discussed in section 2.1.2 (lines 291-300), where a previous study (Guarino *et al*. 2020) compares simulations across different platforms and finds that the various climate variables discussed in this paper are not significantly different across platforms. Please see Figure 6 in Guarino *et al*. (2020) for an example of this.

5. Table 2 and Fig. 2a, the 1.5 m air temperature of LIG still show significant trend in the final years. Could you please show a trend map to check where such trend mainly occurs? Does it happen in the region of interest?

This has now been addressed, with a 1.5 m air temperature trend map for both climate simulations being shown in the supplementary material (SM3) and discussed in section 2.4 (lines 405-408)

6. Lines 297-323: I think it is not so necessary to describe the spin-up in such a detail. Just show the tables, and I also recommend to put Fig. 2 and Fig. 3 into the supplement.

As part of other changes, this has already been done (see above comments to Reviewer 1)

7. Fig. 4 and Fig. 5: Perform a Student's t-test to identify in which regions the anomalies are significant and which regions related to internal variabilities. Given the relatively short length of the MH and LIG experiments, it is very important to do so.

This has now been done to the new version of Figures 4 and 5

8. Line 334 'and'=>', and'

As part of other changes, this has now been removed

9. Lines 336-337: 'in order to' => 'to'

This has been corrected

10. Line 342: Title is confusing. The CMIP6 HadGEM3 simulations include the PI, right?

This has been corrected

11. Line 351: 'central' => 'Central'

This has been corrected

12. Line 359 and a lot of other places in the paper: please make the experiment name consistent throughout the paper, for example, use either MH or midHolocene, the same for LIG and lig127k, piControl and PI.

A new section detailing the terminology has now been added (section 2.1.1), to clarify exactly what term refers to either the simulations or the geological intervals (lines 216-226)

13. Line 371: greater land-sea contrast... Is it also the same case in your model? I would recommend to check the moist static energy instead of surface temperature, to also include the aspect of moisture.

As part of other changes, this has now been removed

14. Lines 374-377: the small anomalies... Again please use Student's t-test. Results discussed in the texts should have a significance level above 95%.

As above, this has now been included in the new versions of Figures 2 and 3, showing the 99% significant levels

15. Lines 373-374: Comparing Fig. 5a and 5b, I observe no obvious shift in ITCZ, only stronger monsoon rainfall in LIG compared to MH.

As part of other changes, this sentence has now been modified and clarified

16. caption of Fig. 6, 9, and 12: Generally West Africa should be within 20W-15E. Why take 20W-30E?

This has now been corrected, such that all zonal mean plots go from 20°W-15°E

17. Lines 398-400: Please explain where the large uncertainty in proxy comes from.

This has now been corrected

18. Lines 422-424: Can this underestimation of the warming be used to explain the "Holocene temperature conundrum"? Or, might the "Holocene temperature conundrum" be caused by the fact that most of the proxy locate in regions with positive temperature anomalies? The proxy data represent seasonal or annual mean value? It might be helpful to discuss these issues. See, e.g. Lohmann et al. (2013, CP) for a comprehensive comparison for SST changes during the Holocene.

We agree with the reviewer that this term is ambiguous, and it has therefore been removed

19. Line 396: It would be better to clarify here the threshold of RMSE (is there any?) for a reasonable simulation result, in terms of surface temperature, precipitation and sst.

We do not use a threshold of RMSE, but we have clarified this in the text (lines 614)

20. Line 447: if => but

As part of other changes, this has now been removed

21. Fig. 10 and 11: Again, please show significance (t-test).

As part of other changes, these figures have now been removed

22. Line 557. The model used prescribed vegetation, and does not consider dust. Please discuss the influence of the lack of interactive vegetation and dust on the Africa monsoon rainfall.

This has now been addressed (lines 933-935)

23. Optional: I encourage the author to make a separate discussion section.

As part of other changes, the summary and conclusions have now been restructured and rewritten, and further discussion has been added throughout the results section

CMIP6/PMIP4 simulations of the mid-Holocene and Last Interglacial using HadGEM3: comparison to the pre-industrial era, previous model versions, and proxy data

~~The UK contribution to CMIP6/PMIP4: mid-Holocene and Last Interglacial experiments with HadGEM3, and comparison to the pre-industrial era and proxy data~~

Charles J. R. Williams[1,5], Maria-Vittoria Guarino[2], Emilie Capron[3], Irene Malmierca-Vallet[1,2], Joy S. Singarayer[4,1], Louise C. Sime[2], Daniel J. Lunt[1], Paul J. Valdes[1]

[1]School of Geographical Sciences, University of Bristol, UK (c.j.r.williams@bristol.ac.uk)
[2]British Antarctic Survey, Cambridge, UK
[3]Physics of Ice, Climate and Earth, Niels Bohr Institute, University of Copenhagen, Denmark
[4]Department of Meteorology & School of Archaeology, Geography and Environmental Science, University of Reading, UK
[5]NCAS-Climate / Department of Meteorology, University of Reading, UK

Corresponding author address:
Room 1.2n, School of Geographical Sciences,
University Road, Bristol, BS8 1SS
United Kingdom

Email: c.j.r.williams@bristol.ac.uk

Short title: mid-Holocene and Last Interglacial experiments with HadGEM3

Keywords: Palaeoclimate, Quaternary change, mid-Holocene, Last Interglacial

ABSTRACT

Palaeoclimate model simulations are an important tool to improve our understanding of the mechanisms of climate change. These simulations also provide tests of the ability of models to simulate climates very different to today. Here we present the results from two brand-new simulations using the latest version of the UK's physical climate model, HadGEM3-GC3.1; the mid-Holocene (~6 ka) and Last Interglacial (~127 ka) simulations, both conducted under the auspices of CMIP6/PMIP4. This is the first time this version of the UK model has been used to conduct paleoclimate simulations. These periods are of particular interest to PMIP4 because they represent the two most recent warm periods in Earth history, where atmospheric concentration of greenhouse gases and continental configuration ~~is~~ are similar to the pre-industrial period, but where there were significant changes to the Earth's orbital configuration, resulting in a very different seasonal cycle of radiative forcing.

Results for these simulations are assessed firstly against the same model's preindustrial control simulation (a simulation comparison, to describe and understand the differences between the preindustrial and the two paleoclimate simulations), and secondly against previous versions of the same model relative to newly-available proxy data (a model-data comparison, to compare all available simulations from the same model with proxy data to assess any improvements due to model advances)~~, previous versions of the UK model, and models from the previous CMIP5 exercise~~. The introduction of this newly available proxy data adds further novelty to this study. Globally, for metrics such as 1.5m temperature and surface rainfall, whilst both the recent paleoclimate simulations are mostly capturing the expected sign and, in some places, magnitude of change relative to the preindustrial, this is geographically and seasonally dependent. Compared to newly-available proxy data (including SST and rainfall), and also incorporating data from previous versions of the model, shows that the relative accuracy of the simulations appears to vary according to metric, proxy reconstruction used for comparison and geographical location. In some instances, such as mean rainfall in the mid-Holocene, there is a clear and linear improvement, relative to proxy data, from the oldest to the newest generation of the model. When zooming into northern Africa, a region known to be problematic for models in terms of rainfall enhancement, the behaviour of the West African monsoon in both recent paleoclimate simulations is consistent with current understanding, suggesting a wetter monsoon during the mid-Holocene and (more so) the Last Interglacial, relative to the preindustrial era. However, regarding the well-documented 'Saharan greening' during the mid-Holocene, results here suggest that the most recent version of the UK's physical model is still unable to reproduce the increases suggested by proxy data, consistent with all other previous models to date. ~~When the current version is compared to the previous generation of the UK model, the most recent version suggests limited improvement. In common with these previous model versions, the simulations reproduce global land and ocean temperatures (both surface and at 1.5 m) and a West~~

~~African monsoon that is consistent with the latitudinal and seasonal distribution of insolation. The Last Interglacial simulation appears to accurately capture Northern Hemisphere temperature changes, but without the addition of Last Interglacial meltwater forcing cannot capture the magnitude of Southern Hemisphere changes. Model-data comparisons indicate that some geographical regions, and some seasons, produce better matches to the palaeodata (relative to pre-industrial) than others. Model-model comparisons, relative to previous generations same model and other models, indicate similarity between generations in terms of both the intensity and northward enhancement of the mid-Holocene West African monsoon, both of which are underestimated. On the 'Saharan greening' which occurred the mid-Holocene African Humid Period, simulation results are likewise consistent with other models. The most recent version of the UK model appears to still be unable to reproduce the amount of rainfall necessary to support grassland across the Sahara.~~

1. INTRODUCTION

Simulating past climates has been instrumental in improving our understanding of the mechanisms of climate change (e.g. Gates 1976, Haywood *et al*. 2016, Jungclaus *et al*. 2017, Kageyama *et al*. 2017, Kageyama *et al*. 2018, Kohfeld *et al*. 2013, Lunt *et al*. 2008, Otto-Bliesner *et al*. 2017, Ramstein *et al*. 1997), as well as in identifying and assessing discrepancies in palaeoclimate reconstructions (e.g. Rind & Peteet 1985). Palaeoclimate scenarios can also provide tests of the ability of models to simulate climates that are very different to today, often termed 'out-of-sample' tests. This notion underpins the idea that robust simulations of past climates improve our confidence in future climate change projections (Braconnot *et al*. 2011, Harrison *et al*. 2014, Taylor *et al*. 2011). Palaeoclimate scenarios have also been used to provide additional tuning targets for models (e.g. Gregoire *et al*. 2011), in combination with historical or pre-industrial conditions.

The international Climate Model Intercomparison Project (CMIP) and the Palaeoclimate Model Intercomparison Project (PMIP) have spearheaded the coordination of the international palaeoclimate modelling community to run key scenarios with multiple models, perform data syntheses, and undertake model-data comparisons since their initiation twenty-five years ago (Joussaume & Taylor 1995). Now in its fourth incarnation, PMIP4 (part of the sixth phase of CMIP, CMIP6), it includes a larger set of models than previously, and more palaeoclimate scenarios and experiments covering the Quaternary (documented in Jungclaus *et al*. 2017, Kageyama *et al*. 2017, Kageyama *et al*. 2018 and Otto-Bliesner *et al*. 2017) and Pliocene (documented in Haywood *et al*. 2016).

PMIP4 specifies experiment set-ups for two ~~warm~~ interglacial simulations: the mid-Holocene (MH) at ~6 ka and the Last Interglacial (LIG) at ~127 ka (although spanning ~~covering~~ ~129-116 ka in its entirety). These are the two most recent warm periods (particularly in the Northern Hemisphere) in Earth history, and are of particular interest to PMIP4; indeed, the MH experiment is one of the two entry cards into PMIP (Otto-Bliesner *et al*. 2017). This is because whilst the atmospheric concentration of greenhouse gases, the extent of land ice, and the continental configuration is similar in these PMIP4 set-ups compared to the pre-industrial (PI) period, significant changes to the seasonal cycle of radiative forcing, relative to today, do occur during these periods due to long-term variations in the Earth's orbital configuration. The MH and LIG both have higher boreal summer insolation and lower boreal winter insolation compared to the PI, as shown by Figure 1, leading to an enhanced seasonal cycle in insolation as well as a change in its latitudinal distribution. The change is more significant in the LIG than the MH, due to the larger eccentricity of the Earth's orbit at that time. Note that, in this figure and indeed all subsequent figures using monthly or seasonal data, the data have been calendar adjusted (Joussaume & Braconnot 1997) according to the method of Pollard & Reusch (2002) and Marzocchi *et al*. (2015); see the Supplementary Material (SM1) for the same figure but using the modern calendar.

Palaeodata syntheses indicate globally warmer surface conditions of potentially ~0.7°C than PI in the MH (Marcott *et al*. 2013) and up to ~1.3°C in the LIG (Fischer *et al*. 2018). During both warm periods there is abundant palaeodata evidence indicating enhancement of Northern Hemisphere summer monsoons (e.g. Wang *et al*. 2008) and in the case of the Sahara, replacement of desert by shrubs and steppe vegetation (e.g. Drake *et al*. 2011, Hoelzmann *et al*. 1998), grassland and xerophytic woodland/scrubland (e.g. Jolly *et al*. 1998a, Jolly *et al*. 1998b, Joussaume *et al*. 1999) and inland water bodies (e.g. Drake *et al*. 2011, Lezine *et al*. 2011). Recent palaeodata compilations involving either air temperatures or SST (Capron *et al*. 2014, Hoffman *et al*. 2017) reveal that the maximum temperatures were reached asynchronously in the LIG between the Northern and Southern Hemispheres. Concerning precipitation, historically this has been lacking relative to temperature or SST reconstructions. One often-cited study for the MH is that of Bartlein *et al*. (2011), comprising a combination of existing quantitative reconstructions based on pollen and plant macrofossils; this provides evidence of the interaction between orbital variations and greenhouse gas forcing, and the atmospheric circulation response. More recently, one newly-published dataset of LIG precipitation proxy data (which the current study benefits from as part of the model-data comparison, see below) is that of Scussolini *et al*. (2019). Here, a number of climate models are assessed against this brand-new dataset, finding an agreement with proxy data over Northern Hemisphere landmasses, but less so in the Southern Hemisphere (Scussolini *et al*. 2019).

Many modelling studies have been undertaken in an attempt to reproduce the changes suggested by proxy data throughout the Quaternary, and especially during the interglacial periods discussed here, and there is not scope in this current study to give a full review here. An overview of multi-model assessments during the LIG can be found in Lunt *et al*. (2013). However, one example is the aforementioned monsoon enhancement (and expansion/contraction) during the Quaternary, and previous studies have focused on various aspects of this, such as whether any expansion was hemispherically consistent or asynchronous between hemispheres (e.g. Kutzbach *et al*. 2008, McGee *et al*. 2014, Singarayer & Burrough 2015, Singarayer *et al*. 2017, Wang *et al*. 2006, Wang *et al*. 2014). During the LIG, the aforementioned asynchronous temperature distribution between the hemispheres ~~Furthermore,~~ has been investigated by a number of model simulations~~, suggest~~ suggesting that this may have been caused by meltwater induced shutdown of the Atlantic Meridional Overturning Circulation (AMOC) in the early part of the LIG, due to the melting of the Northern Hemisphere ice sheets during the preceding deglaciation (e.g. Carlson 2008, Smith & Gregory 2009, Stone *et al*. 2016). ~~During both warm periods there is abundant palaeodata evidence indicating enhancement of Northern Hemisphere summer monsoons (e.g. Wang *et al*. 2008) and in the case of~~

~~the Sahara, replacement of desert by shrubs and steppe vegetation (e.g. Drake *et al*. 2011, Hoelzmann *et al*. 1998) and inland water bodies (e.g. Drake *et al*. 2011, Lezine *et al*. 2011).~~

The driving mechanism producing the climate and environmental changes indicated by the palaeodata for the MH and LIG ~~and MH~~ is different to current and future anthropogenic warming, as the former results from orbital forcing changes whilst the latter results from increases in greenhouse gases. Moreover, the orbital forcing primarily acts on shortwave radiation whereas greenhouse gas changes primarily act upon the longwave radiation flux, and the orbital forcing can lead to uneven horizontal and seasonal changes whereas greenhouse gas forcing can cause more uniform anomalies (it should be noted that whilst a precise calculation of the radiative forcing due to changes in MH and LIG greenhouse gases is beyond the scope of this study, such a calculation could follow the methodology of Gunnar *et al*. [1998]). Nevertheless, despite these differences in driving mechanism, ~~However,~~ these past high latitude (and mainly Northern Hemisphere) warm intervals are a unique opportunity to understand the magnitudes of forcings and feedbacks in the climate system that produce warm interglacial conditions, which can help us understand and constrain future climate projections (e.g. Holloway *et al*. 2016, Rachmayani *et al*. 2017, Schmidt *et al*. 2014). Running the same model scenarios with ever newer models enables the testing of whether model developments are producing improvements in palaeo model-data comparisons, assuming appropriate boundary conditions are used. Previous iterations of PMIP, with older versions of the PMIP4 models, have uncovered persistent shortcomings (Harrison *et al*. 2015) that have not been eliminated despite developments in resolution, model physics, and addition of further Earth system components. One key example of this is the continued underestimation of the increase in rainfall over the Sahara in the MH PMIP simulations (e.g. Braconnot *et al*. 2012).

In this study we run and assess the latest version of the UK's physical climate model, HadGEM3-GC3.1.~~–~~ Whilst older versions of the UK model have been included in previous iterations of CMIP, and whilst present-day and future simulations from this model are included in CMIP6, the novelty of this study is that this is the first time this version has been used to conduct any paleoclimate simulations. In Global Coupled (GC) version 3 (and therefore in the following GC3.1), there have been many updates and improvements, relative to its predecessors, which are discussed extensively in Williams *et al*. (2017) and a number of companion scientific model development papers (see Section 2.1). As a brief introduction, however, GC3 includes a new aerosol scheme, multilayer snow scheme, multilayer sea ice and several other parametrization changes, including a set relating to cloud and radiation, as well as a revision to the numerics of atmospheric convection (Williams *et al*. 2017). In addition, the ocean component of GC3 has other changes including ~~a new~~ an updated ocean and sea ice model, a new cloud scheme, and further revisions to all parametrization schemes (Williams *et al*. 2017). See Section 2.1 for further details.

Following the CMIP6/PMIP4 protocol, here the PMIP4 MH and LIG simulations have been conducted and assessed, with the assessment adopting a two-pronged approach. Firstly a simulation comparison is made between these simulations and the same model's PI simulation (to describe and understand the differences between them). Secondly a model-data comparison is made between the current and previous versions of the same model relative to newly-available proxy data, thereby assessing any improvements due to model advances~~comparing the results with available proxy data, previous versions of the UK's same physical climate model, and other models from CMIP5~~. In addition to a global assessment, a secondary ~~The~~ focus of this paper is on the fidelity of the temperature anomalies ~~globally~~ and the degree of precipitation enhancement in the Sahara, the latter of which has proved problematic for several generations of models. ~~The results discussed here are split into two sections: after an assessment of the level of equilibrium gained during the spin-up phase, the main focus is on the model-data and model-model comparisons using the production runs.~~ Following this introduction, Section 2 describes the model, the experimental design, ~~and~~ the proxy data used for the model-data comparisons, and a brief discussion of the simulation spin-up phases. Section 3 then presents the results, beginning with the model-model comparison and following with the model-data comparison, and ~~f~~~~divided into two subsections: i) equilibrium during the spin-up phase; and ii) model-data and model-model comparisons from the production runs. Finally~~finally, section 4 summarises and concludes.

2. MODEL, EXPERIMENT DESIGN, ~~AND~~ DATA AND SPIN-UP SIMULATIONS

2.1. Model

2.1.1. Model terminology

In this paper, and consistent with CMIP nomenclature, the 'spin-up phase' of the simulations refers to when they are spinning up to atmospheric and oceanic equilibrium, whereas the 'production run' refers to the end parts (usually the last 50 or 100 years) of the simulation used to calculate the climatologies, presented as the results. When discussed as geological intervals, the preindustrial, mid-Holocene and Last Interglacial are referred to as the PI, MH and LIG respectively. In contrast, when discussed as the three most recent simulations using HadGEM3 (see below), consistent with CMIP they are referred to as the *piControl*, *midHolocene* and *lig127k* simulations, respectively. When the *midHolocene* and *lig127k* are discussed collectively, they are referred to as the 'warm climate simulations'; whilst it is acknowledged that other factors differentiate these simulations such as orbital configuration or CO_2, 'warm climate simulations' was deemed an appropriate collective noun.

2.1.2. Model details

The ~~MH and LIG simulations conducted here (referred to as *midHolocene* and *lig127k*, respectively, and collectively as the~~ warm climate simulations conducted here~~)~~, and ~~indeed~~ the *piControl* ~~PI~~

simulation (~~*piControl*,~~ conducted elsewhere as part of the UK's CMIP6 runs and used here for comparative purposes) were all run using the same fully-coupled GCM: the Global Coupled 3 configuration of the UK's physical climate model, HadGEM3-GC3.1. ~~-~~Full details on HadGEM3-GC3.1, and a comparison to previous configurations, are given in Williams *et al*. (2017) and Kuhlbrodt *et al*. (2018). Here, the model was run using the Unified Model (UM), version 10.7, and including the following components: i) Global Atmosphere (GA) version 7.1, with an N96 atmospheric spatial resolution (approximately 1.875° longitude by 1.25° latitude) and 85 vertical levels; ii) the NEMO ocean component, version 3.6, including Global Ocean (GO) version 6.0 (ORCA1), with an isotropic Mercator grid which, despite varying in both meridional and zonal directions, has an approximate spatial resolution of 1° by 1° and 75 vertical levels; iii) the Global Sea Ice (GIS) component, version 8.0 (GSI8.0); iv) the Global Land (GL) configuration, version 7.0, of the Joint UK Land Environment Simulator (JULES); and v) the OASIS3 MCT coupler. The official title for this configuration of HadGEM3-GC3.1 is HadGEM3-GC31-LL N96ORCA1 UM10.7 NEMO3.6 (for brevity, hereafter HadGEM3).

All of the above individual components are summarised by Williams *et al*. (2017) and detailed individually by a suite of companion papers (see Walters *et al*. 2017 for GA7 and GL7, Storkey *et al*. 2017 for GO6 and Ridley *et al*. 2017 for GIS8). However, a brief description of the major changes relative to its predecessor are given in the Supplementary Material ~~here~~. ~~Beginning with GA7 and GL7, a once-in-a-decade replacement of the model's dynamical core, implementing ENDGame, was undertaken for the previous version (GA6) and therefore remains the same in GA7 (Walters *et al*. 2017). In addition, a number of bottom-up and top-down developments were included in GA7. For the former, these include improvements to the radiation scheme to allow better treatment of gases absorption, improvements to how warm rain and ice clouds are treated, and an improvement to the numerics of the convection scheme (Walters *et al*. 2017). For the latter, these include further improvements to the microphysics as well as an incremental development of ENDGame (Walters *et al*. 2017). Together these led to reductions in four model errors that were deemed critical in the previous configuration: i) South Asian monsoon rainfall biases over India; ii) biases in both temperature and humidity in the tropical tropopause; iii) shortcomings in the numerical conservation; and iv) biases in surface radiation fluxes over the Southern Ocean (Walters *et al*. 2017). In addition to these developments, two new parameterisation schemes were introduced in GA7: firstly the UK Chemistry and Aerosol (UKCA) GLOMAP-mode aerosol scheme, to improve the representation of tropospheric aerosols, and secondly a multi-layer snow scheme in JULES, to allow the first time inclusion of stochastic physics in UM climate simulations (Walters *et al*. 2017).~~

~~For the GO and GIS components, a number of improvements to GO6 have been made since the previous version, the first of which was an upgrade of the NEMO base code (to version 3.6) which~~

~~allowed a formulation for momentum advection (from Hollingsworth *et al*. 1983), a Lagrangian icebergs scheme, and a simulation of circulation below ice shelves (Storkey *et al*. 2018). Other developments included an improvement to the warm SST bias in the Southern Ocean (as detailed by Williams *et al*. 2017), as well as tuning to various parameters e.g. the isopycnal diffusion (Storkey *et al*. 2018). For GIS8, along with improvements to the albedo scheme and more realistic semi-implicit coupling, the biggest development since its predecessor is the inclusion of multilayer thermodynamics, giving a heat capacity to the sea ice and allowing vertical variation of conduction (Ridley *et al*. 2018). Testing of these two components produced a better simulation compared to its predecessor, with more realistic mixed layer depths in the Southern Ocean and the aforementioned reduced warm bias, the latter of which was deemed primarily due to the tuning of the different mixing (e.g. vertical and isopycnal) parameters (Storkey *et al*. 2018).~~

When all of these components are coupled together to give GC3, there have been several improvements relative to its predecessor (GC2), most noticeably to the large warm bias in the Southern Ocean (which was reduced by 75%), as well as an improved simulation of clouds, sea ice, the frequency of tropical cyclones in the Northern Hemisphere as well as the AMOC, and the Madden Julian Oscillation (MJO) (Williams *et al*. 2017). Relative to the previous fully-coupled version of the model (HadGEM2), which was submitted to the last CMIP5/PMIP3 exercise, many systematic errors have been improved including a reduction of the temperature bias in many regions~~including a reduction in many regions to the temperature bias~~, a better simulation of mid-latitude synoptic variability, and an improved simulation of tropical cyclones and the El Niño Southern Oscillation (ENSO) (Williams *et al*. 2017).

Here, the *midHolocene* and *lig127k* simulations were both run on the UK National Supercomputing Service, ARCHER, whereas the *piControl* was run on a different platform based within the UK Met Office's Hadley Centre. While this may mean that anomalies computed against the *piControl* are potentially influenced by different computing environments, and not purely the result of different climate forcings, the reproducibility of GC3.1 simulations across different platforms has been tested (Guarino *et al*. 2020). It was found that, although a simulation length of 200 years is recommended whenever possible to adequately capture climate variability across different platforms, the main climate variables considered here (e.g. surface temperature) are not expected to be significantly different on a 100- or 50-year timescale (see, for example, Fig. 6 in Guarino *et al*. [20202019]) as they are not directly affected by medium-frequency climate processes~~ such as ENSO~~.

~~Not including queueing time, both simulations were achieving 3-4 model years per day during the spin-up phase, and 1-2 model years per day during the production run; see below for the differences in output, and therefore speed, between the two phases.~~

2.2. Experiment design

Full details of the experimental design~~,~~ and results from the CMIP6 *piControl* simulation~~,~~ are documented in Menary *et al*. (2018). Both the warm climate simulations followed the experimental design given by Otto-Bliesner *et al*. (2017), and specified at https://pmip4.lsce.ipsl.fr/doku.php/exp_design:index. The primary differences from the *piControl* were to the astronomical parameters and the atmospheric trace greenhouse gas concentrations, summarised in Table 1. For the astronomical parameters, these were prescribed in Otto-Bliesner *et al*. (2017) according to orbital constants from Berger & Loutre (1991). However, in HadGEM3, the individual parameters (e.g. eccentricity, obliquity, etc) use orbital constants based on Berger (1978), according to the specified start date of the simulation. For the atmospheric trace greenhouse gas concentrations, these were based on recent reconstructions from a number of sources (see Table 1 for values, and section 2.2 in Otto-Bliesner *et al*. [2017] for a full list of references/sources).

All other boundary conditions, including solar activity, ice sheets, topography and coastlines, volcanic activity and aerosol emissions, are identical to the CMIP6 *piControl* simulation. Likewise, vegetation was prescribed to present-day values, to again match the CMIP6 *piControl* simulation. As such, the *piControl* and both the warm climate simulations actually include a prescribed fraction of urban land surface. As a result of this, our orbitally- and greenhouse gas-forced simulations should be considered as anomalies to the *piControl*, rather than absolute representations of the MH or LIG climate.

Both the warm climate simulations were started from the end of the *piControl* spin-up phase (which ran for approximately 600 years), after which time the *piControl* was considered to be in atmospheric and oceanic equilibrium (Menary *et al*. 2018). To assess this, four metrics were used, namely net radiative balance at the top of the atmosphere (TOA), surface air temperature (SAT), ~~and~~ full-depth ocean temperature (OceTemp) and full-depth ocean salinity (OceSal) (Menary *et al*. ~~(~~2018). See Section ~~3.1~~ 2.4 (and in particular Table 2) for an analysis of the equilibrium state of both the *piControl* and the warm climate simulations. Starting at the end of the *piControl*, these were then run for their own spin-up phases, 400 and 350 years for the *midHolocene* and *lig127k* respectively. ~~During this phase, ~700 diagnostics were output, containing mostly low temporal frequency (e.g. monthly, seasonal and annual) fields.~~ Once the simulations were considered in an acceptable level of equilibrium (see Section ~~3.1~~ 2.4), a production phase was run for 100 and 200 years for the *midHolocene* and *lig127k* respectively, during which the full CMIP6/PMIP4 diagnostic profile ~~(totalling ~1700 fields)~~ was implemented to output both high and low temporal frequency variables.

2.3. Data

Recent data syntheses compiling quantitative surface temperature and rainfall reconstructions were used in order to evaluate the warm climate simulations.

For the MH, the global-scale continental surface mean annual temperature (MAT) and rainfall (or mean annual precipitation, MAP) reconstructions from Bartlein *et al*. (2011), with quantitative uncertainties accounting for climate parameter reconstruction methods, were used (see Data Availability for access details). They rely on a combination of existing quantitative reconstructions based on pollen and plant macrofossils and are inferred using a variety of methods (see Bartlein *et al*. 2011 for further details). At each site, the 6 ka anomaly (corresponding to the 5.5-6.5 ka average value), is given relative to the present day, and in the case where modern values could not be directly inferred from the record, modern climatology values (1961-1990) were extracted from the Climate Research Unit historical climatology data set (New *et al*. 2002). Further proxy data for the MH, such as SST reconstructions, are not included here, as an extensive model-data comparison is presented in a companion paper (Brierley *et al*. 2020).

For the LIG, two recent different sets of surface temperature data are available. Firstly, the Capron *et al*. (2017) 127 ka timeslice of SAT and sea surface temperature (SST) anomalies (relative to pre-industrial, 1870-1899), is based on polar ice cores and marine sediment data that are (i) located poleward of 40° latitude and (ii) have been placed on a common temporal framework (see Data Availability for access details). Polar ice core water isotope data are interpreted as annual mean surface air temperatures, while most marine sediment-based reconstructions are interpreted as summer (defined here as July-September, JAS) SST signals. For each site, the 127 ka value was calculated as the average value between 126 and 128 ka using the surface temperature curve resampled every 0.1 ka. Here, we use the SST anomalies only. Secondly, a global-scale time slice of SST anomalies, relative to pre-industrial (1870-1889), at 127 ka was built, based on the recent compilation from Hoffman *et al*. (2017), which includes both annual and summer SST reconstructions (see Data Availability for access details). This adds further novelty to this study, by using a new combined dataset based on this existing data. The 127 ka values at each site were extracted, following the methodology they proposed for inferring their 129, 125 and 120 ka time slices i.e. the SST value at 127 ka was taken on the provided mean 0.1 ka interpolated SST curve for each core location. Data syntheses from both Capron *et al*. (2014, 2017) and Hoffman *et al*. (2017) are associated with quantitative uncertainties accounting for relative dating and surface temperature reconstruction methods. Here, the two datasets are treated as independent data benchmarks, as they use different reference chronologies and methodologies to infer temporal surface temperature changes, and therefore they should not be combined. See Capron *et al*. (2017) for a detailed comparison of the two syntheses. A model-data comparison exercise using existing LIG data compilations focusing on continental surface temperature (e.g. Turney and Jones 2010) was not attempted, as they do no benefit

yet from a coherent chronological framework, preventing the definition of a robust time slice representing the 127 ka terrestrial climate conditions (Capron *et al*. 2017).

A brand-new, recently-published dataset of proxy precipitation anomalies (again, relative to the preindustrial) is also used for model-data comparison purposes here, adding further novelty to this study. The proxy data are compiled from existing literature by Scusscolini *et al*. (2019), and the dataset includes 138 proxy locations from a number of paleoclimatic archives including pollen, fossils other than pollen, lacustrine or marine sediment composition, loess deposits, and other multi-proxy sources. Note that, as Scusscolini *et al*. (2019) observe, unlike temperature anomalies the majority of precipitation anomalies in the existing literature are not quantitative. To allow a quantitative comparison, Scusscolini *et al*. (2019) use a semi-quantitative scale, based on their expert judgement, to show a LIG that is 'much wetter', 'wetter', 'no discernible change', 'drier' and 'much drier', relative to the PI. The same scale is therefore used here. See Scusscolini *et al*. (2019) for further information, and see Data Availability for access details).

2.4. Spin-up simulations

As briefly mentioned above, both the warm climate simulations had a spin-up phase before the main production run was started~~.~~, briefly discussed here. As an example of atmospheric equilibrium, ~~A~~annual global mean 1.5 m air temperature and TOA radiation from both warm climate simulations, compared to the *piControl*, are ~~shown in Fig_atmos_equilib and~~ summarised in Table 2; see Supplementary Material (SM2) for the timeseries of these fields. ~~There is a clear increase in temperature during the beginning of this period, as the *piControl* slowly spins up from its original starting point; this levels off towards the end of the period, however, with a final temperature trend of 0.03°C century^{-1} (Table 2 and Fig_atmos_equilib a).~~ For the warm climate simulations, despite considerable interannual variability and arguably more so than in the *piControl* (~~particularly halfway through the *lig127k* simulation~~see SM2), both are showing ~~small~~ long-term trends of -0.06°C century^{-1} and -0.16°C century^{-1} for the last 100 years of the *midHolocene* and *lig127k*, respectively (Table 2 and ~~Fig_atmos_equilib a~~). ~~–~~ The spatial patterns of these trends, also shown in the Supplementary Material (SM3), are similar in both warm climate simulations, with much of the statistically significant cooling occurring over high latitude regions in both Hemispheres, and particularly so over Antarctica in the *lig127k* simulation (SM3). ~~The same is true for~~ The TOA radiation balance is also showing long-term (and again slightly negative) trends by the end of the simulations, with -0.05 W m^2 and -0.06 W m^2 for the the *midHolocene* and *lig127k*, respectively~~, where the *piControl* has a slow downward trend towards zero until equilibrium was reached, whereas the *midHolocene* and *lig127k* are relatively stable (Fig_atmos_equilib b).~~

~~For the ocean~~As an example of oceanic equilibrium, annual global mean full-depth OceTemp and OceSal are shown in Table 2 (and again visualised in the Supplementary Material, SM4)~~and Fig_ocean_equilib. There is again a clear increase in OceTemp during the *piControl* spin-up phase, which again stabilises at 0.035°C century^{-1} by the end of the period (Table 2). OceTemp~~.~~ is steadily increasing throughout the *piControl* , and this continues ~~to increase in the *li* *g127k* until it stabilises within the last ~50 years~~ in both warm climate simulations~~(Fig_ocean_equilib a)~~, whereas there is a dramatic fall in ocean salinity in these simulations (SM4). ~~A similar pattern is shown in OceSal, with a steady decrease in the *piControl* spin-up phase which continues during the *midHolocene* and, conversely, starts to increase before stabilising during the *lig127k* (Fig_ocean_equilib b).~~ Concerning the long-term trends, Menary *et al*. (2018) considered values acceptable for equilibrium to be < +/- 0.035°C century^{-1} and < +/-0.0001 psu century^{-1} (for OceTemp and OceSal, respectively); as shown in Table 2, although both warm climate simulations meet the temperature criterion, the *midHolocene* ~~neither meet the~~ it is not meeting the salinity criterion (-0.00047 psu ~~and 0.006 psu for the *midHolocene* and *lig127k*, respectively, compared to a criterion of 0.0001 psu~~). However, running for several thousands of years (and > 5 years of computer time), which would be needed to reach true oceanic equilibrium, was simply unfeasible here given time and resource constraints.

3. RESULTS

~~As briefly mentioned above, both the warm climate simulations had a spin-up phase before the main production run was started. The results discussed here are therefore split into two sections: firstly, assessing the level of atmospheric and oceanic equilibrium during (and, in particular, at the end of) the spin-up phase, and secondly assessing the 100-year climatology from the production run.~~

~~3.1. Spin-up~~

~~Annual global mean 1.5 m air temperature and TOA radiation from both warm climate simulations, compared to the *piControl*, are shown in Fig_atmos_equilib and summarised in Table 2. Note that the *piControl* spin-up phase was run in three separate parts, to accommodate for minor changes/updates in the model as the simulation progressed. There is a clear increase in temperature during the beginning of this period, as the *piControl* slowly spins up from its original starting point; this levels off towards the end of the period, however, with a final temperature trend of 0.03°C century^{-1} (Table 2 and Fig_atmos_equilib a). For the warm climate simulations, despite considerable interannual variability (particularly halfway through the *lig127k* simulation) both are showing small long-term trends of -0.06°C century^{-1} and -0.16°C century^{-1} for the last 100 years of the *midHolocene* and *lig127k*, respectively (Table 2 and Fig_atmos_equilib a). The same is true for TOA, where the *piControl* has a slow downward trend towards zero until equilibrium was reached, whereas the *midHolocene* and *lig127k* are relatively stable (Fig_atmos_equilib b).~~

~~For the ocean, annual global mean OceTemp and OceSal are shown in Table 2 and Fig_ocean_equilib. There is again a clear increase in OceTemp during the *piControl* spin-up phase, which again stabilises at 0.035°C century^{-1} by the end of the period (Table 2). Whilst OceTemp stabilises in the *midHolocene* and indeed has a smaller trend than the *piControl* (Table 2), it continues to increase in the *lig127k* until it stabilises within the last ~50 years (Fig_ocean_equilib a). A similar pattern is shown in OceSal, with a steady decrease in the *piControl* spin-up phase which continues during the *midHolocene* and, conversely, starts to increase before stabilising during the *lig127k* (Fig_ocean_equilib b). Concerning the long-term trends, Menary *et al.* (2018) considered values acceptable for equilibrium to be < +/- 0.035°C century^{-1} and < +/- 0.0001 psu century^{-1} (for OceTemp and OceSal, respectively); as shown in Table 2, although both warm climate simulations meet the temperature criterion, neither meet the salinity criterion (-0.007 psu and 0.006 psu for the *midHolocene* and *lig127k*, respectively, compared to a criterion of 0.0001 psu). However, running for several thousands of years (and > 5 years of computer time), which would be needed to reach true oceanic equilibrium, was simply unfeasible here given time and resource constraints.~~

3.21. Production runs results

The warm climate production runs were undertaken following the spin-up phase, with ~~a 100 year~~ the climatology of each simulation being compared to that from the *piControl*, as well as available proxy data, using either annual means or summer/winter seasonal means. For the latter, depending on the availability of the proxy data, Northern Hemisphere summer is defined as either June-August (JJA) or ~~July-September (~~JAS~~)~~, and Northern Hemisphere winter is defined as either December-February (DJF) or January-March (JFM); and vice versa for Southern Hemisphere summer/winter. – As briefly introduced in Section 1, ~~Using atmospheric diagnostics, the~~ the focus is on ~~three~~ two separate measures: i) to describe and understand the differences between the ~~current~~ two most recent ~~two~~ warm climate simulations and the *piControl* in terms of temperature, rainfall and atmospheric/oceanic circulation changes; and ii) to compare both current simulations, as well as simulations from previous versions of the UK model (where available), with ~~existing and~~ the aforementioned newly-available proxy data, to assess any improvements due to model advances. A final aim, discussed only briefly here but shown in the Supplementary Material, is to include previous CMIP5 models to address the question of whether any of the simulations produce enough rainfall to allow vegetation growth across the Sahara: the mid-Holocene 'Saharan greening'~~and iii) to compare both current simulations with those from previous versions of the UK model (where available), such as HadGEM2-ES or HadCM3, in order to assess any improvements due to model advances. In this aim, previous CMIP3 and 5 versions of the UK model, alongside other CMIP5 models, will be assessed to address the question of whether simulations produce enough rainfall to allow vegetation growth across the Sahara: the mid-Holocene 'Saharan greening' problem~~.

3.~~2~~1.1. Do the CMIP6 HadGEM3 warm climate simulations show temperature, rainfall and atmospheric/oceanic circulation differences when compared to the pre-industrial era?

Here we focus on mean differences between the HadGEM3 warm climate simulations and the corresponding *piControl*. Calendar adjusted ~~Seasonal~~ annual and seasonal mean summer ~~and~~ /winter 1.5 m air temperature anomalies (relative to the *piControl*) from both warm climate simulations are shown in Figure 2.~~–~~ As an example and for comparative purposes, the same figure but where the data are based on the modern calendar is shown in the Supplementary Material (SM5); this suggests that the impact of the calendar adjustments on this field, and at this spatial and temporal scale, is negligible, with the only obvious impact occurring over the Northern Hemisphere polar regions during JJA in both simulations, but more so in the *lig127k* simulation (due to the larger changes in insolation resulting in a larger change to the calendar, relative to the MH). Consistent with the seasonality of the changes, the differences between either simulations are less at the annual timescale (Figure 2a and d) than during individual seasons, but are still nevertheless to statistically significant at the 99% level. During JJA, the *midHolocene* is showing a widespread statistically significant increase in temperatures of up to 2°C across the entire Northern Hemisphere north of 30°N, more in some places e.g. Greenland (Figure 2b), consistent with the increased latitudinal and seasonal distribution of insolation caused by known differences in the Earth's axial tilt (Berger & Loutre 1991, Otto-Bliesner *et al*. 2017). The only places showing a reduction in temperature are West and ~~central~~ Central Africa (around 10°N) and northern India; this, as discussed below, is likely related to increased rainfall in response to a stronger summer monsoon, but could also be due to the resulting increase in cloud cover (reflecting more insolation) or a combination of the two. During DJF, only the Northern Hemisphere high latitudes (north of 60°N) continue this warming trend, with the rest of continental Africa and Asia showing a reduction in temperature (Figure 2c). These patterns are virtually the same ~~during~~ in the *lig127k* simulation (Figure 2e and f), just much more pronounced (with statistically significant temperature increases during JJA of 5°C or more); again, this is consistent with the differences in the Earth's axial tilt, which were more extreme (and therefore Northern Hemisphere summer experienced larger insolation changes) in the LIG relative to the MH (Berger & Loutre 1991, Otto-Bliesner *et al*. 2017). Another clear feature of these figures, at either annual or seasonal timescales, is polar amplification, which is likely associated with changes in sea-ice; as shown in the Supplementary Material (SM6), statistically significant decreases in sea-ice are shown throughout the polar regions of both hemispheres in the *midHolocene*, relative to the *piControl*. The same is true for the *lig127k* simulation, just more pronounced (not shown).

Calendar adjusted seasonal mean summer and winter surface daily rainfall anomalies (again relative to the *piControl*) from both warm climate simulations are shown in Figure 3. In line with the aforementioned increased latitudinal and seasonal distribution of insolation, the largest differences in either simulation occur during Northern Hemisphere summer (Figure 3b and e). Both warm climate

simulations are showing statistically against increases in rainfall around the monsoon regions, especially over northern India and equatorial Africa, more so in the *lig127k* (Figure 3e). Both simulations are also showing oceanic drying relative to the *piControl*, especially in the equatorial Atlantic and Pacific, again more pronounced in the *lig127k* (Figure 3e). In contrast, during DJF, less of an impact is seen in either simulation relative to the *piControl*, with a small but statistically significant increase in rainfall in oceanic equatorial regions but drying over tropical land regions e.g. southern Africa, central Brazil and northern Australia (Figure 3c and f). Again, consistent with the increased insulation changes during the LIG compared to the MH, these differences are stronger in the *lig127k* simulation (Figure 3f). Consistent with the temperature differences, these signals are again weaker at the annual timescale but are nevertheless statistically significant (Figure 3a and b).

A measure of oceanic circulation is also considered here, shown by the three HadGEM3 simulations of meridional overturning circulation (MOC) in the Atlantic basin and globally (Figure 4a-c and d-f, respectively). Although not identical, the differences are nevertheless negligible, with both warm climate simulations almost exactly reproducing the structures of weakly and strongly overturning MOC seen in the *piControl*; for example, the strongly overturning MOC in the upper levels of the Atlantic is marginally stronger in the *midHolocene* at ~30-40°N relative to the other two simulations, but the structures are very similar. This suggests that the changes to atmospheric fields such as P-E, energy fluxes and wind stress (in response to the insolation changes) are having a minimal impact on the overturning circulation, and this is consistent with other work (e.g. Guarino *et al*. [2020])

A key region of interest, concerning mean precipitation changes and changes to the extent and latitudinal distribution of monsoon regions, is northern Africa, primarily because of the aforementioned inability of previous models to reproduce the increases shown by the proxy data here (e.g. Braconnot *et al*. 2007, Braconnot *et al*. 2012). Therefore, Figure 5 reproduces the above precipitation changes but zooms into Africa and additionally includes ~~Mean~~ calendar adjusted mean JJA (the primary monsoon region) ~~rainfall and~~ 850mb wind anomalies (relative to the *piControl*) ~~from both warm climate simulations are shown in Fig_wafricarain_prod, which zooms into Africa~~. In response to the increased Northern Hemisphere summer insolation, the West African monsoon is enhanced in both simulations, with positive (negative) rainfall anomalies across sub-Saharan Africa (eastern equatorial Atlantic) suggesting a northward displacement of the ~~ITCZ~~ rainfall maxima. This is consistent with previous work, with a northward movement of the rainbelt being associated with increased advection of moisture into the continent (Huag *et al*. 2001, Singarayer *et al*. 2017, Wang *et al*. 2014). This increased advection of moisture is shown by the enhanced low-level westerlies at all latitudes but especially over the regions of rainfall maxima in Figure 5a and b~~Fig_wafricarain_prod~~, drawing in more moisture from the tropical Atlantic, which are consistent with previous work documenting the intensified monsoon circulation ~~associated with a greater land-sea temperature~~

~~contrast~~ (Huag *et al*. 2001, Singarayer *et al*. 2017, Wang *et al*. 2006). This pattern is enhanced in the *lig127k* relative to the *midHolocene*, again in response to the stronger insolation changes relative to the MH, and the northward displacement of the ~~ITCZ~~ central rainbelt is more pronounced in the *lig127k* simulation (Figure 5~~Fig_wafricarain_prod~~ c). ~~Interestingly, however, regarding very small anomalies (i.e. < 1 mm day^{-1}), the *midHolocene* is showing wetter conditions further north, throughout the Sahara and up to the Mediterranean, whereas the *lig127k* simulation has small dry anomalies in this region (Fig_wafricarain_prod a and b for the *midHolocene* and *lig127k*, respectively).~~

The change to the intensity and the spatial pattern (e.g. latitudinal positioning and extent) of the West African monsoon is further shown in Figure 6, which shows calendar adjusted daily JJA rainfall by latitude over West Africa (averaged over 20°W-15°E, land points only) from both warm climate simulations. This figure also includes MH and LIG simulations from previous generations of the same model. It should be noted that although LIG experiments have been conducted previously with both model-model and model-data comparisons being made (Lunt *et al*. 2013), all of these experiments were carried out using early versions of the models and were thus not included in CMIP5. Moreover, as part of their assessment Lunt *et al*. (2013) considered a set of four simulations, at 130, 128, 125 and 115 ka, none of which are directly comparable to the current HadGEM3 *lig127k* simulation. Instead, a LIG simulation has recently been undertaken using one of the original versions of the UK's physical climate model, HadCM3, and so this is used here to compare with the *lig127k* simulation.

Beginning with the recent paleoclimate HadGEM3 simulations, in line with the changes in insolation both warm climate simulations are showing higher absolute values at their peak (between ~7.5-10°N) than the *piControl* (Figure 6a). Concerning anomalies, both simulations are showing a large increase in rainfall relative to the *piControl* (of ~2 and 6 mm day^{-1} for the *midHolocene* and *lig127k*, respectively) over the monsoon region between ~10-12°N (Figure 6b). Relative to previous versions of the same model, the previous generation (HadGEM2-ES) is slightly drier then HadGEM3 over this region for its PI simulation and slightly wetter for its MH simulation; conversely, the version before that (HadCM3) is consistently wetter than HadGEM3, for all of its simulations (Figure 6a). There also appears to be a northward displacement in the oldest version, with the largest difference between the simulations and their corresponding PI simulations occurring at ~11°N in the two most recent versions of the model, whereas in HadCM3 this appears to be shifted northwards to ~12.5°N (Figure 6b). This northward displacement in certain models consistent with previous work (e.g. Huag et al. 2001, Otto-Bliesner et al. 2017, Singarayer et al. 2017, Wang et al. 2014). In terms of the latitudinal extent, the results suggest that all warm climate simulations (regardless of generation) are producing a wider Northern Hemisphere monsoon region (i.e. a greater northerly extent) relative to each version's

PI, with rainfall falling to near zero at ~18°N in the PI simulations but extending to 20°N (and above, in terms of the LIG simulations) in both warm climate simulations (Figure 6a). This is again consistent with previous work, where various theories are compared as to the reasons behind the latitudinal changes in the rainbelt's position, one which is a symmetric expansion during boreal summer (Singarayer & Burrough 2015, Singarayer et al. 2017).

3.21.2. ~~Model~~ Simulation comparison and Model-Data comparison: Do the CMIP6 HadGEM3 simulations reproduce the 'reconstructed' climate based on available proxy data, and has there been any noticeable improvement relative to previous versions of the same model?

Although the above analysis is useful and confirms that the most recent warm climate simulations are responding consistently to the increased latitudinal and seasonal distribution of insolation, it does not give any information on which (if any) of the simulations is most accurate or which version of the model is better at reproducing proxy-observed conditions. Therefore, h~~H~~ere we ~~focus on~~ bring in a comparison with newly-available ~~recent~~ proxy data, comparing these to all versions of the model, focusing on surface air temperature, SST and rainfall (drawing direct comparisons, as well as using the root mean square error (RMSE~~,~~) but without a cut-off threshold), between both proxy ~~and~~ vs simulated data and HadGEM3 vs previous versions~~, summarised in Table 4a),~~). The aim of this is to firstly ~~to~~ see how well the current warm climate simulations are reproducing the 'observed' approximate magnitudes and patterns of change, and secondly to assess any possible improvement from previous versions of the same model. It is worth noting that both simulated and proxy anomalies contain a high level of uncertainty (as measured by the standard deviation), and in many locations the uncertainty is ~~often~~ larger than the anomalies themselves (not shown). The following results should therefore be considered with this caveat in mind.

Before the spatial patterns are compared, it is useful to assess global means from the three HadGEM3 simulations (focusing on 1.5 m air temperature, calculated both annually and during Northern and Southern Hemisphere summer, JJA and DJF respectively) ~~for model-model comparisons~~. Table 3 shows these global means, where it is clear that when annual means are considered, the *midHolocene* simulation is actually cooler than the *piControl*. This discrepancy with the palaeodata, which ~~in general~~ at many locations suggests a warmer MH relative to PI, and this is consistent with previous work using other models (e.g. ~~also exists in previous models, and is termed the 'Holocene temperature conundrum' by~~ Lui *et al*. ~~(~~2014). The *lig127k* simulation is, however, warmer than the *piControl* simulation. Given the seasonal distribution of insolation in these two simulations, it is expected that the largest difference to the *piControl* occurs during boreal summer, and indeed it does; during JJA, there is a warmer *lig127k* and a slightly warmer *midHolocene* (1.69°C and 0.07°C, respectively). The opposite is true during DJF.

~~Concerning the spatial patterns during the MH, Fig_proxy_mh_loc shows simulated surface MAT and MAP anomalies from the *midHolocene* simulation versus MH proxy anomalies from Bartlein *et al*. (2011), both of which have over 600 proxy locations in total (Table 4), although mostly confined to the Northern Hemisphere. For MAT, globally the simulation looks reasonable (RMSE = 2.45°C), and appears to be able to reproduce the sign of temperature change for many locations, with both simulated and proxy anomalies suggesting increases in temperature North of 30°N (Fig_proxy_mh_loc a and b). This is not true everywhere, such as across the Mediterranean where the simulation suggests a small warming but the proxy data indicates cooling (Fig_proxy_mh_loc a and b). However, regarding the magnitude of change, the *midHolocene* simulation is underestimating the temperature increase across most of the Northern Hemisphere, with for example increases of up to 1°C across Europe from the simulation compared to 3-4°C increases from the proxy data (Fig_proxy_mh_loc a and b). In the simulation, temperature anomalies only reach these magnitudes in the Northern Hemisphere polar region (i.e. north of 70°N), not elsewhere. A similar conclusion can be drawn from MAP (RMSE = 280 mm yr^{-1}), where again the *midHolocene* simulation is correctly reproducing the sign of change across most of the Northern Hemisphere, but in some places not the magnitude. Over the eastern US, for example, rainfall decreases of up to 200 mm yr^{-1} are being shown by the simulation whereas the proxy data suggests a much stronger drying of up to 400 mm yr^{-1} (Fig_proxy_mh_loc c and d). Elsewhere, such as over Europe and Northern Hemisphere Africa, the simulation more accurately reproduces the magnitude of rainfall increases; both simulated and proxy anomalies show increases of 200-400 mm yr^{-1} (Fig_proxy_mh_loc c and d).~~

Concerning the spatial patterns during the MH, Figure 7 shows simulated surface MAT anomalies from the current *midHolocene* simulation and those from two previous versions of the same model, versus MH proxy anomalies from Bartlein *et al*. (2011). Note that, here, statistical significance of the simulated anomalies has not been shown, because firstly the aim here is to assess all differences regardless of significance and secondly because a measure of statistical significance (for HadGEM3) has already been presented in Figure 2; statistical significance from the other versions of the same model is virtually identical (not shown). Globally, all three models are showing a reasonable level of agreement to the proxy data, with RMSE = 2.45°C, 2.42°C and 2.37°C for HadGEM3, HadGEM2-ES and HadCM3, respectively (Table 4a). Using this metric, the oldest version of the model (HadCM3) is doing marginally better than the other models, relative to the proxy data. Spatially, however, there are differences to the proxy data and between model generations. Although all three generations appear to be able to reproduce the sign of temperature change for many locations, with both simulated and proxy anomalies suggesting increases in temperature North of 30°N and especially over northern Europe, the Arctic Circle increases are not as homogenous in HadCM3 (Figure 7d) and indeed this model shows cooling over the Greenland Sea. Although this cannot be corroborated by the proxy data, due to a lack of coverage, neither of the later generation models show this to the same extent

(Figure 7b and c). Discrepancies with the proxy data also occur in all three simulations across the Mediterranean region, where all three simulations suggest a small warming but the proxy data indicates cooling (Figure 7). Moreover, regarding the magnitude of change, all three simulations are underestimating the temperature increase across most of the Northern Hemisphere, with for example increases of up to 1°C across Europe from the simulations compared to 3-4°C increases from the proxy data. In the simulations, temperature anomalies only reach these magnitudes in the Northern Hemisphere polar region (i.e. north of 70°N), not elsewhere. Further equatorward, all three simulations are identifying a slight cooling over the West African monsoon region (as discussed above), but the accuracy of this relative to the proxy data is difficult to ascertain given the lack of coverage across Africa and, where there are data locations, a highly variable sign of change (Figure 7a).

A similar conclusion can be drawn from MAP, shown in Figure 8, where all three simulations are correctly reproducing the sign of change across most of the Northern Hemisphere, although more so in the two most recent generations of the model (HadGEM3 and HadGEM2-ES), but in some places not the magnitude. Over the eastern US, for example, rainfall decreases of up to 200 mm yr^{-1} are being shown by the simulations (Figure 8b-d) whereas the proxy data suggests a much stronger drying of up to 400 mm yr^{-1} (Figure 8a). Elsewhere, such as over Europe and Northern Hemisphere Africa, the simulations more accurately reproduce the magnitude of rainfall increases; both simulated and proxy anomalies show increases of 200-400 mm yr^{-1}. Globally, Table 4a suggests that the most recent generation model, HadGEM3, is doing better than the others, relative to the proxy data (RMSE = 285.9 mm yr^{-1}, 293.5 mm yr^{-1} and 304.7 mm yr^{-1} for HadGEM3, HadGEM2-ES and HadCM3, respectively). In terms of how the spatial patterns change according to model version, during the MH the two most recent simulations generally agree (RMSE = 90.8 mm $year^{-1}$, Table 4a) and show similar spatial patterns; focusing again on the African monsoon region (for the aforementioned reasons), both simulations show a drier equatorial Atlantic during the MH and then increased rainfall around 10°N (Figure 8b and c for HadGEM3 and HadGEM2-ES, respectively). Both simulations also suggest that the increases in rainfall extend longitudinally across the entire African continent, with the largest changes not only occurring across western and central regions but also further east. In contrast, globally HadCM3 agrees less with HadGEM3 (RMSE = 121.8 mm $year^{-1}$, Table 4a) and only suggests a wetter MH over West Africa, not further east. HadCM3, and indeed HadGEM2-ES, also differs from the most recent simulation over the equatorial Atlantic, showing a region of drying that is not only stronger in magnitude but also larger in terms of spatial extent; whilst still present in HadGEM3, this feature that is much weaker (Figure 8b-d).

~~Concerning the spatial patterns during the LIG, Fig8_proxy_lig_loc shows simulated mean SST anomalies (calculated both annually and during JAS/JFM) from the *lig127k* simulation and LIG proxy~~

~~anomalies from two sources, Capron *et al*. (2017) and Hoffman *et al*. (2017). When annual anomalies are considered, despite the lack of reconstructions in the Capron *et al*. (2017) data (Table 4), there is relatively good agreement (RMSE = 2.44°C and 2.94°C for the Capron *et al*. (2017) and Hoffman *et al*. (2017) data, respectively, and which is within the average uncertainty range), between simulated and observed SST anomalies in the Northern Hemisphere (and in particular in the North Atlantic), with both suggesting increased temperatures during the LIG of up to 3°C (Fig_proxy_lig_loc a). There are discrepancies, such as in the Norwegian Sea, where the Hoffman *et al*. (2017) reconstructions suggest a cooler LIG than preindustrial, whereas the *lig127k* simulation shows a consistent warming; this is, however, consistent with previous work, and earlier climate models have also failed to capture this cooling (Capron *et al*. 2014, Stone *et al*. 2016). Note that, over Greenland and Antarctica, the Capron *et al*. (2017) proxy data show SAT, not SST, and are therefore not compared in this figure; comparison with simulated SAT, however, suggests that the model is capturing the sign, if not the magnitude, of annual change over these regions (not shown). During Northern Hemisphere summer, JAS (during which period Capron *et al*. [2017] has the most proxy locations [Table 4]), the simulated anomalies are in agreement with many, but not all, of the proxy locations (RMSE = 3.11°C and 2.06°C for the Capron *et al*. (2017) and Hoffman *et al*. (2017) data, respectively); examples of where they differ, not just in magnitude but also sign, again include the Norwegian and Labrador Seas (Fig_proxy_lig_loc b). In Southern Hemisphere summer, JFM, the model suggests a general (but weak) cooling in the South Atlantic relative to preindustrial and a general (but weak) warming in the Southern Ocean (Fig_proxy_lig_loc c). However, certain proxy locations (such as off the coast of southern Africa) suggest a much warmer LIG than preindustrial (RMSE = 1.94°C and 4.24°C for the Capron *et al*. (2017) and Hoffman *et al*. (2017) data, respectively), which in stark contrast to the cooling in the same region from the *lig127k* simulation (Fig_proxy_lig_loc c) . In the Southern Ocean, the majority of simulated anomalies reproduce the observed sign of change, but not the magnitude; the *lig127k* simulation suggests temperature increases of up to 1°C, whereas both proxy datasets suggest SST increases of 2-3°C depending on location (Fig_proxy_lig_loc c).~~

Concerning the spatial patterns during the LIG, Figure 9 shows simulated mean SST anomalies (calculated both annually and during JAS/JFM) from the current *lig127k* simulation and that from the oldest version of the same model, versus LIG proxy anomalies from two sources, Capron *et al*. (2017) and Hoffman *et al*. (2017). No LIG simulation using HadGEM2-ES is currently available. When annual anomalies are considered, there is relatively good agreement globally between HadGEM3 and the proxy data where RMSE = 3.03°C and 2.42°C for the Capron *et al*. (2017) and Hoffman *et al*. (2017) data, respectively (Table 4b). HadCM3 performs marginally better when compared to the Capron *et al*. (2017) data, but worse when compared to the Hoffman *et al*. (2017) data (Table 4b). Similarly varying results also occur when JAS and JFM anomalies are considered, with HadGEM3

comparing slightly better or worse than HadCM3 according to season and proxy dataset used; all of the values, however, show relatively good agreement, with no simulation exceeding RMSE = 4.5°C in any season or with any dataset (Table 4b). Spatially, HadGEM3 is showing a general agreement between simulated and proxy annual and JAS anomalies in the Northern Hemisphere (and in particular in the North Atlantic), with both suggesting increased temperatures during the LIG of up to 5°C (Figure 9a and b). HadCM3 is not capturing these magnitudes at the annual timescale (Figure 9d) and, despite showing greater warming during JAS, is still lower than HadGEM3; this is more in agreement with the proxy data at higher latitudes (e.g. the western Norwegian Sea at ~70°N) but less so further south (Figure 9e). This might suggest that, in this region, HadGEM3 is actually overestimating the degree of warming. Nevertheless, in both versions of the model there are discrepancies concerning not just in the magnitude but also in the sign of change, such as in the eastern Norwegian Sea or the Labrador Sea, where reconstructions suggest a cooler LIG but both versions show a consistent warming (Figure 9b and e). This is, however, consistent with previous work, and earlier climate models have also failed to capture this cooling (Capron *et al*. 2014, Stone *et al*. 2016). In Southern Hemisphere summer, JFM, both versions agree on a general (but weak) cooling in the South Atlantic relative to preindustrial and a weak warming in the Southern Ocean (Figure 9c and f). In contrast certain proxy locations (such as off the coast of southern Africa) suggest a much warmer LIG than preindustrial, which is opposite to the simulated cooling in the same region (Figure 9c and f). Further south, the majority of simulated anomalies reproduce the observed sign of change, but not the magnitude; here, the simulations suggest temperature increases of up to 1°C, whereas both proxy datasets suggest SST increases of 2-3°C depending on location (Figure 9c and f).

For rainfall changes during the LIG, Figure 10 shows simulated annual mean surface rainfall anomalies from the current *lig127k* simulation and that from the oldest version of the same model, versus LIG proxy anomalies from Scusscolini et al. (2019). Note that the simulated anomalies shown here are annual anomalies, as opposed to daily anomalies in Figure 3, to be consistent with the proxy data. Note also that, for these proxy reconstructions, a semi-quantitative scale is used by Scusscolini et al. (2019) rather than actual anomalies and is therefore reproduced here; this ranges from a unitless -2 to 2, corresponding to 'Much wetter LIG anomaly', 'Wetter', 'No noticeable anomaly', 'Drier' and 'Much drier LIG anomaly'. It is for this reason that RMSE values have not been calculated here. As was suggested from the MH simulations (Figure 8), both versions of the model are showing similar patterns of rainfall changes, along the same lines as those seen during the MH but again enhanced (Figure 10). Both versions are showing enhanced rainfall across the Northern Hemisphere equatorial zone and in particular the monsoon regions during the LIG, often exceeding 500 mm year^{-1} in some places. In the Northern Hemisphere, both versions of the model are generally in agreement with the proxy data, with most proxy locations showing 'Wetter' or 'Much wetter' conditions. There are, however, some discrepancies elsewhere, such as the regions of tropical drying over e.g. Brazil and

southern Africa in the simulations being in stark contrast to the 'Wetter' conditions suggested by the proxy data (Figure 10). Concerning the differences in the spatial patterns between the model versions, although both generations qualitatively show similar patterns, there are subtle differences. Again focusing on the African monsoon region, HadGEM3 shows greatly increased rainfall across all of sub-Saharan Africa, centred on 10°N but extending from ~5°N to almost 20°N and longitudinally across the entire African continent (Figure 10a). In contrast, and similar to the MH results, in HadCM3 the largest rainfall increases are less apparent over East Africa (Figure 10b).

It would therefore be reasonable to say that, for both ~~warm climate~~ MH and LIG simulations, whilst the most recent version of the model is capturing the sign and magnitude of change relative to proxy reconstructions (for either temperature or rainfall) in some locations, this is highly geographically dependent and there are locations where the current simulation fails to capture even the sign of change. Compared to previous versions of the same model, any improvement also appears to be highly variable according to metric, proxy reconstruction used for comparison and geographical location, with for example HadGEM3 showing some improvement relative to previous versions for rainfall during the MH, but not surface air temperature. The accuracy of the most recent ~~The~~ model, and indeed previous generations, also appear~~s~~s to be seasonally dependent, with the most recent *lig127k* simulation ~~(but not the *midHolocene* simulation)~~ correctly reproducing both the sign and magnitude of change during Northern Hemisphere summer in some locations, but not during Southern Hemisphere summer or annually. It would also appear that, for both the MH and LIG simulations, whilst there is less difference between the most recent two configurations of the model, they are nevertheless quite different to the oldest version. For global mean annual rainfall during the MH, Table 4a shows a linear progression of improvement across the three versions of the model, as well as more agreement between the two most recent model generations. This is also true when just the region of rainfall maxima in northern Africa is considered, with both of the two most recent generations, and especially HadGEM2-ES, being marginally closer to the proxy data than HadCM3 (RMSE = 463.7 mm yr^{-1}, 424.5 mm yr^{-1} and 468.4 mm yr^{-1} for HadGEM3, HadGEM2-ES and HadCM3, respectively). In all simulations, although spatial patterns of rainfall are similar, there are discrepancies especially over the African monsoon region; the oldest version of the model, for example, only shows rainfall increases over West Africa, whereas the two most recent versions imply Africa-wide rainfall increases at this latitude. If a comparison is made with satellite-derived rainfall data for the modern West African monsoon (not shown), results suggest that rainfall maxima are not just limited to West Africa but also occur over the central region and East Africa, more consistent with the two most recent versions of the model. One reason for HadCM3 not identifying this longitudinal extent might be connected to the very coarse spatial resolution of this model, relative to the others, impacting any topographically-induced rainfall, especially over the East African Highlands.

3.2.3. Model-Model comparison: Do the CMIP6 HadGEM3 simulations show an improvement compared to older CMIP versions of the UK model?

Here we focus on model-model intercomparisons, comparing the HadGEM3 warm climate simulations with firstly those from previous versions of the UK model and secondly with those from other models included in CMIP5. It should be noted that although LIG experiments have been conducted previously with both model-model and model-data comparisons being made (Lunt *et al.* 2013), all of these experiments were carried out using early versions of the models and were thus not included in CMIP5. Moreover, as part of their assessment Lunt *et al.* (2013) considered a set of four simulations, at 130, 128, 125 and 115 ka, none of which are directly comparable to the current HadGEM3 *lig127k* simulation. Instead, a LIG simulation has recently been undertaken using one of the original versions of the UK's physical climate model, HadCM3, and so this is used here to compare with the *lig127k* simulation. As discussed above, this section is divided into two parts: firstly the mean climate state of the warm climate simulations will be compared to the model's predecessors, focusing again on hydroclimate of the West African monsoon (given the known problem of simulated rainfall underestimation in this region, see e.g. Braconnot *et al.* [2007]). Here, both direct comparisons and RMSE values will again be examined, this time calculating the RMSE between the simulated rainfall anomaly from two older versions of the UK model versus the current HadGEM3 *midHolocene* and *lig127k* simulations (summarised in Table 4b).

Secondly, previous generation simulations (from all available models included in CMIP5) will be compared to see whether the most recent HadGEM3 *midHolocene* simulation is now providing enough rainfall to allow vegetation growth across the Sahara; something which previous generations of models from CMIP5 did not (Braconnot *et al.* 2007).

3.2.3.1. Mean climate state from predecessors of HadGEM3

Regarding the magnitude and latitudinal extent of the West African monsoon, Fig_latrain_gen shows the JJA rainfall differences averaged over West Africa from the current *midHolocene* and *lig127k* simulation versus two of the model's predecessors. During the MH, the two most recent generations of the model (HadGEM3 and HadGEM2-ES) generally agree on drier conditions over the equatorial Atlantic and then wetter conditions over West Africa, however the oldest generation model (HadCM3) does not reproduce the Atlantic drying. Likewise the two most recent generations share a similar latitudinal distribution of rainfall above ~5°N, with a wetter MH over land, peaking at ~2-3 mm day^{-1} at ~11-12°N. Interestingly, the previous version of the model (HadGEM2-ES) shows the strongest and most northwardly displaced rainfall peak, as discussed in previous work (e.g. Huag *et al.* 2001, Otto-Bliesner *et al.* 2017, Singarayer *et al.* 2017, Wang *et al.* 2014); the most recent version, HadGEM3, has lower northward displacement compared to the two older versions of the model. Both

recent versions suggest that the monsoon region extends to ~17°N, above which the differences between the MH and PI reduce to near zero. In contrast, HadCM3 suggests a generally weaker, but latitudinally more extensive, monsoon region, suggesting a wetter MH (by ~1 mm day^{-1}) as far north as 20°N and beyond. For the LIG, HadGEM3 is showing a much stronger monsoon region relative to the *piControl*, compared to HadCM3. However, in terms of extent, similar results are shown to those for the MH, with HadCM3 showing a generally weaker, but more northwardly displaced, monsoon region. In this older generation model, positive rainfall anomalies of ~2-3 mm day^{-1} extend as far north as 17-18°N, whereas in HadGEM3 they fall to ~1 mm day^{-1} at these latitudes.

In terms of the spatial patterns of the West African monsoon, Fig_wafricarain_gen_mh and Fig_wafricarain_gen_lig show the JJA daily rainfall climatology differences from the same three model generations for the MH and LIG, respectively. During the MH, consistent with Fig_latrain_gen, the two most recent simulations generally agree (RMSE = 0.46 mm day^{-1}) and show similar spatial patterns, with a drier equatorial Atlantic during the MH and then increased rainfall around 10°N (Fig_wafricarain_gen_mh a and b for HadGEM3 and HadGEM2-ES, respectively). Both simulations also suggest that the increases in rainfall extend longitudinally across the entire continent, with the largest changes not only occurring across western and central regions but also further east. In contrast, HadCM3 is less consistent than HadGEM3 (RMSE = 0.53 mm day^{-1}) and only suggests a wetter MH over West Africa; moreover, again consistent with Fig_latrain_gen, HadCM3 suggests that although the West African monsoon region is longitudinally narrower, it is latitudinally wider than the other two simulations (Fig_wafricarain_gen_mh c). HadCM3 also differs from the other simulations over the equatorial Atlantic, showing a region of drying that is not only stronger in magnitude (with the MH being over 5 mm day^{-1} drier than the PI in HadCM3, compared to ~2-3 mm day^{-1} in the two most recent simulations), but also larger in terms of latitude and longitude extent (Fig_wafricarain_gen_mh c).

During the LIG, only the most recent and oldest version of the model can be compared, as a LIG simulation using HadGEM2-ES is unavailable. In Fig_wafricarain_gen_lig there is a noticeable difference between generations and the level of agreement is the lowest across all simulation combinations (RMSE = 1.57 mm day^{-1}), with the most recent HadGEM3 showing greatly increased rainfall across all of northern Africa, centred on 10°N but extending from ~5°N to almost 20°N and beyond (Fig_wafricarain_gen_lig a), again consistent with Fig_latrain_gen. In contrast, and similar to the MH results, in HadCM3 the largest rainfall increases are confined to Western Africa only, rather than extending longitudinally across the continent (Fig_wafricarain_gen_lig b). However, in terms of latitudinal extent, HadCM3 is showing weak wet anomalies all the way to the Mediterranean, whereas the monsoon region diminishes further south (at ~30°N) in HadCM3 and dry anomalies are suggested North of this. Another noticeable difference is the region of drying, with the most recent generation

~~model placing this over the equatorial Atlantic (consistent with the MH) but HadCM3 shifting this further east, over most of central Africa (Fig_wafricarain_gen_lig b). The region of equatorial Atlantic drying shown by the more recent versions of the model is actually wetter during this HadCM3 LIG simulation.~~

~~It would therefore appear that, for the MH, whilst there is less difference between the most recent two configurations of the model (in terms of a more localised West African monsoon region), there nevertheless has been improvement since the oldest version of the UK's physical climate model. For the LIG, where unfortunately there is no intermediate generation, it would be reasonable to say that again considerable change has occurred since the oldest generation model, with the suggestion that, although HadCM3 is identifying an enhanced monsoon which extends to the Mediterranean (albeit with very weak anomalies), at lower latitudes it is not showing the level of northward displacement as the most recent version, apart from in the far western regions.~~

3.~~2~~1.3. ~~Rainfall across the~~ Saharan greening

Finally, a brief discussion is given on the 'Saharan greening' question. Given that the warm climate simulations, and indeed the *piControl*, did not use interactive~~, but rather prescribed,~~ vegetation, it is not possible to directly test if the model is reproducing the ~~'~~Saharan greening~~'~~ that proxy data suggest. For example, Jolly *et al*. (1998a, 1998b) analysed MH pollen assemblages across northern Africa and suggested that some areas south of 23°N (characterised by desert today) were grassland and xerophytic woodland/scrubland during the MH (Joussaume *et al*. 1999). To circumvent this caveat, Joussaume *et al*. (1999) developed a method for indirectly assessing Saharan greening, based on the annual mean rainfall anomaly relative to a given model's modern simulation. Using the water-balance module from the BIOME3 equilibrium vegetation model (Haxeltine & Prentice 1996), Joussaume *et al*. (1999) calculated the increase in mean annual rainfall, zonally averaged over 20°W-30°E, required to support grassland at each latitude from 0 to 30°N, compared to the modern rainfall at that latitude. This was then used to create maximum and minimum estimates, within which bounds the model's annual mean rainfall anomaly must lie to suggest enough of an increase to support grassland (Joussaume *et al*. 1999).

Therefore, an adapted version of Figure 3a in Joussaume *et al*. (1999) is shown in the Supplementary Material (SM7), which shows mean annual rainfall anomalies by latitude (to be consistent with the proxy data-based threshold) from not only the current *midHolocene* simulation, but also all previous MH simulations from CMIP5. Concerning the threshold required to support grassland, it is clear that although the current *midHolocene* simulation is just within the required bounds at lower latitudes (e.g. up to 17°N), north of this the current *midHolocene* simulation is not meeting the required threshold, neither are any of the other CMIP5 models after ~18°N (SM7). It would therefore appear that the

'Saharan greening' problem has yet to be resolved, and may well only be reproduced once interactive vegetation, and indeed interactive dust, is included in the simulation; given the current lack of an interactive vegetation/dust model, vegetation-related climate feedbacks (e.g. albedo) on the system are therefore currently missing.

4. SUMMARY AND CONCLUSIONS

This study has conducted and assessed the mid-Holocene and Last Interglacial simulations using the latest version of the UK's physical climate model, HadGEM3-GC3.1, comparing the results firstly with the model's preindustrial simulation and secondly with previous versions the same model, against ~~with~~ available proxy data~~, previous versions of the same model, and other models from CMIP's previous iteration, CMIP5~~. Therefore this study is novel, being the first time this version of the UK model has been used to conduct any paleoclimate simulations and therefore the first time we are in a position to include them as part of the UK's contribution to CMIP6/PMIP4. Both the *midHolocene* and *lig127k* simulations followed the experimental design defined in Otto-Bliesner *et al*. (2017) and ~~under the auspices of~~ and the CMIP6/PMIP4 protocol~~,~~. Both simulations were run for a 350-400 year spin-up phase, during which ~~time~~ atmospheric and oceanic equilibrium ~~was~~ were assessed, and once an acceptable level of equilibrium had been reached, the production runs were started.

~~Concerning the results from the spin-up phase, comparison to the metrics used to assess the CMIP6 *piControl* suggest that both warm climate simulations reached an acceptable state of equilibrium, in the atmosphere at least, to allow the production runs to be undertaken. From these, both simulations are showing global temperatures consistent with the latitudinal and seasonal distribution of insolation, and with previous work (e.g. Otto-Bliesner *et al*. 2017).~~ Globally, whilst both the recent simulations are mostly capturing the sign and, in some places, magnitude of change relative to the PI, similar to previous model simulations this is geographically and seasonally dependent. It should be noted that the proxy data (against which the simulations are evaluated) also contain a high level of uncertainty in both space and time (in terms of both seasons and geological era), and so it is encouraging that the simulations are generally reproducing the large-scale sign of change, if not at an individual location. Compared to previous versions of the same model, this appears to vary according to metric, proxy reconstruction used for comparison and geographical location. In some instances, such as annual mean rainfall in the MH, there is a clear and linear improvement (relative to proxy data) through the model generations when rainfall is considered globally; likewise there is more accuracy in the two recent versions (again relative to proxy data) than the oldest version when only the West African monsoon region is considered (see Table 4a and the RMSE values discussed in the concluding paragraph of Section 3.1.2).

Likewise, when zooming into Africa, the behaviour of the West African monsoon in both HadGEM3 warm climate simulations is consistent with current understanding (e.g. Huag et al. 2001, Singarayer et al. 2017, Wang et al. 2014), which suggests a wetter (and possibly latitudinally wider, and/or northwardly displaced) monsoon during the MH and LIG, relative to the PI. Regarding model development in simulating the West African monsoon, there are differences between model generations; the oldest version of the model, for example, limits the rainfall increases to over sub-Saharan West Africa only, whereas the two most recent versions imply Africa-wide (i.e. across all longitudes) rainfall increases at this same latitude~~a~~. A~~lthough there has been an improvement since the oldest version of the UK's physical climate model (HadCM3), the two most recent version of the model yield similar results in terms of both intensity and position.~~ Lastly, regarding the well-documented 'Saharan greening' during the MH, results here suggest that the most recent version of the UK's physical climate model is consistent with all other previous models to date.

In conclusion, the results suggest that the most recent version of the UK's physical climate model is reproducing climate conditions consistent with the known changes to insolation during these two warm periods~~, and is consistent with previous versions of the same model, and other models~~. Even though the *lig127k* simulation did not contain any influx of Northern Hemisphere meltwater, shown by previous work to be a critical forcing in LIG simulations (causing regions of both warming and cooling, according to location), it is still nevertheless showing increased temperatures in certain regions. ~~A potential caveat of this conclusion, however, is the matter of spin-up and the fact that neither of the current warm climate simulations were in oceanic equilibrium when the production runs were undertaken. The production runs were undertaken nevertheless because the resources required to run for several thousands of years (needed to reach true oceanic equilibrium) would have been impossible to obtain, but future simulations using this model should endeavour to obtain a better level of oceanic equilibrium.~~ Another limitation of using this particular version of the model is that certain processes, such as vegetation and atmospheric chemistry, were prescribed, rather than allowed to be dynamically evolving. Moreover, for ~~reasons of necessity~~ practical reasons some of the boundary conditions were left as PI, such as vegetation, ~~surface like,~~ anthropogenic deforestation and aerosols; a better simulation might be achieved if these were prescribed for the MH and LIG. Processes and boundary conditions such as these may be of critical importance regarding climate sensitivity during the MH and the LIG, and therefore ongoing work is underway to repeat both of these experiments using the most recent version of the UK's Earth Systems model, UKESM1. Here, although the atmospheric core is HadGEM3, UKESM1 contains many other earth system components (e.g. dynamic vegetation), and therefore in theory should be able to better reproduce these paleoclimate states.

DATA AVAILABILITY

The model simulations will be uploaded in the near future to the Earth System Grid Federation (ESGF) WCRP Coupled Model Intercomparison Project (Phase 6), but are not yet publicly available. The simulations are, however, available by directly contacting the lead author. For the MH reconstructions, the data can be found within the Supplementary Online Material of Bartlein *et al*. (2011), at https://link.springer.com/article/10.1007/s00382-010-0904-1. For the LIG temperature reconstructions, the data can be found within the Supplementary Online Material of Capron *et al*. (2017), at https://www.sciencedirect.com/science/article/pii/S0277379117303487?via%3Dihub, and the Supplementary Online Material of Hoffman *et al*. (2017), at https://science.sciencemag.org/content/suppl/2017/01/23/355.6322.276.DC1. The LIG temperature reconstructions created here, based on the above Hoffman *et al*. (2017) data, are currently available by directly contacting the lead author. For the LIG precipitation reconstructions, the data can be found within the Supplementary Online Material of Scusscolini *et al*. (2019), at https://advances.sciencemag.org/content/suppl/2019/11/18/5.11.eaax7047.DC1.

COMPETING INTERESTS

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTION

CJRW conducted the *midHolocene* simulation, carried out the analysis, produced the figures, wrote the majority of the manuscript, and led the paper. MVG conducted and provided the *lig127k* simulation, and contributed to some of the analysis and writing. EC provided the proxy data, and contributed to some of the writing. IMV provided the HadCM3 LIG simulation. PJV provided the HadCM3 MH simulation. JS contributed to some of the writing. All authors proofread the manuscript and provided comments.

ACKNOWLEDGEMENTS

CJRW acknowledges the financial support of the UK Natural Environment Research Council-funded SWEET project (Super-Warm Early Eocene Temperatures), research grant NE/P01903X/1. CJRW also acknowledges the financial support of the Belmont-funded PACMEDY (PAlaeo-Constraints on Monsoon Evolution and Dynamics) project, as does JS. MVG and LCS acknowledge the financial support of the NERC research grants NE/P013279/1 and NE/P009271/1. EC acknowledges financial support from the ChronoClimate project, funded by the Carlsberg Foundation.

REFERENCES

Bartlein, P. J., Harrison, S. P., Brewer, S., et al. (2011). 'Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis'. Clim. Dyn. 37: 775–802. DOI:10.1007/s00382-010-0904-1

Berger, A. L. (1978). 'Long-term variations of daily insolation and Quaternary climatic changes'. J. Atmos. Sci. 35: 2362-2367. https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2

Berger, A. L. & Loutre, M. F. (1991). Insolation values for the climate of the last 10,000,000 years. Quaternary Sci. Rev. 10: 297–317. https://doi.org/10.1016/0277-3791(91)90033-Q

Braconnot, P., Harrison, S. P., Kageyama, M., et al. (2012). 'Evaluation of climate models using palaeoclimatic data'. Nature Climate Change. 2 (6): 417. DOI: 10.1038/NCLIMATE1456

Braconnot, P., Harrison, S. P., Otto-Bliesner, B, et al. (2011). 'The palaeoclimate modelling intercomparison project contribution to CMIP5'. CLIVAR Exch. Newsl. 56: 15–19

Braconnot, P., Otto-Bliesner, B., Harrison, S., et al. (2007). 'Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features'. Clim. Past. 3: 261-277. https://doi.org/10.5194/cp-3-261-2007

Brierley, C. M., Zhao, A., Harrison, S., et al. (2020). 'Large-scale features and evaluation of the PMIP4-CMIP6 midHolocene simulations'. Clim. Past. Under review

Capron, E., Govin, A., Stone, E. J., et al. (2014). 'Temporal and spatial structure of multi-millennial temperature changes at high latitudes during the Last Interglacial'. Quat. Sci. Rev. 103: 116-133. https://doi.org/10.1016/j.quascirev.2014.08.018

Capron, E., Govin, A., Feng R. et al. (2017). 'Critical evaluation of climate syntheses to benchmark CMIP6/PMIP4 127 ka Last Interglacial simulations in the high-latitude regions'. Quat. Sci. Rev. 168: 137-150. DOI: 10.1016/j.quascirev.2017.04.019

Carlson, A. E. (2008). 'Why there was not a Younger Dryas-like event during the Penultimate Deglaciation'. Quaternary Sci. Rev. 27: 882-887. DOI: 10.1016/j.quascirev.2008.02.004

Drake, N. A., Blench, R. M., Armitage, S. J., et al. (2011). 'Ancient watercourses and biogeography of the Sahara explain the peopling of the desert'. Proceedings of the National Academy of Sciences. 108 (2): 458-462. DOI: 10.1073/pnas.1012231108

Fischer, H., Meissner, K. J., Mix, A. C. et al. (2018). 'Palaeoclimate constraints on the impact of 2°C anthropogenic warming and beyond'. Nature Geoscience. 11: 474-485. https://doi.org/10.1038/s41561-018-0146-0

Gates, W. L. (1976). 'The numerical simulation of ice-age climate with a global general circulation model'. J. Atmos. Sci. 33: 1844–1873. DOI: 10.1175/1520-0469(1976)033<1844:TNSOIA>2.0.CO;2

Gregoire, L. J., Valdes, P. J., Payne, A. J. & Kahana, R. (2011). 'Optimal tuning of a GCM using modern and glacial constraints'. Clim Dyn. 37: 705-719. DOI:10.1007/s00382-010-0934-8

Guarino, M. V., Sime, L., Schroeder, D., et al. (2020). 'Machine dependence and reproducibility for coupled climate simulations: The HadGEM3-GC3.1 CMIP Preindustrial simulation'. GMD. 13 (1): 139-154. https://doi.org/10.5194/gmd-2019-83

Guarino, M. V. et al. (2020). 'A sea ice-free Arctic during the Last Interglacial supports fast future loss'. Nature Climate Change. Under review

Gunnar, M., Highwood, E. J., Shin, K. P. & Stordal, F. (1998) "New estimates of radiative forcing due to well mixed greenhouse gases.". Geophys. Res. Lett. 25 (14): 2715-2718. https://doi.org/10.1029/98GL01908

Harrison, S. P., Bartlein, P. J., Brewer, S., et al. (2014). 'Climate model benchmarking with glacial and mid-Holocene climates'. Clim. Dyn. 43: 671–688. https://doi.org/10.1007/s00382-013-1922-6

Harrison, S. P., Bartlein, P. J., Izumi, K., et al. (2015). 'Evaluation of CMIP5 palaeo-simulations to improve climate projections'. Nature Climate Change. 5: 735. DOI: 10.1038/nclimate2649

Haxeltine, A. & Prentice, I. C. (1996). 'BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types'. Global Biogeochemical Cycles. 10 (4): 693-709. DOI: 10.1029/96GB02344

1786 Haywood, A. M., Dowsett, H. J., Dolan, A. M. et al. (2016). 'The Pliocene Model Intercomparison
1787 Project (PlioMIP) Phase 2: scientific objectives and experimental design'. Clim. Past. 12: 663-675.
1788 https://doi.org/10.5194/cp-12-663-2016
1789
1790 Haug, G., Hughen, K. A., Sigman, D. M., et al. (2001). 'Southward migration of the intertropical
1791 convergence zone through the Holocene'. Science. 293: 1304-1308. DOI: 10.1126/science.1059725
1792
1793 Hoelzmann, P., Jolly, D., Harrison, S. P., et al. (1998). 'Mid-Holocene land-surface conditions in
1794 northern Africa and the Arabian Peninsula: A data set for the analysis of biogeophysical feedbacks in
1795 the climate system'. Global Biogeochemical Cycles. 12 (1): 35-51.
1796 https://doi.org/10.1029/97GB02733
1797
1798 Hoffman, J. S., Clark, P. U., Parnell, A. C., et al. (2017). 'Regional and global sea-surface
1799 temperatures during the last interglaciation'. Science. 355: 276-279. DOI: 10.1126/science.aai8464
1800
1801 Hollingsworth, A., Kållberg, P., Renner, V. & Burridge, D. M. (1983). 'An internal symmetric
1802 computational instability'. QJRMS. 109: 417–428. https://doi.org/10.1002/qj.49710946012
1803
1804 Holloway, M. D., Sime, L. C., Singarayer, J. S., et al. (2016). 'Antarctic last interglacial isotope peak
1805 in response to sea ice retreat not ice-sheet collapse'. Nature Comms. 7: 12293.
1806 https://doi.org/10.1038/ncomms12293
1807
1808 Jungclaus, J. H., Bard, E., Baroni, M. et al. (2017). 'The PMIP4 contribution to CMIP6 – Part 3: The
1809 last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations'.
1810 GMD. 10: 4005-4033. https://doi.org/10.5194/gmd-10-4005-2017
1811
1812 Jolly, D., Harrison, S. P., Damnati, B. & Bonnefille, R. (1998a). 'Simulated climate and biomes of
1813 Africa during the Late Quaternary: Comparison with pollen and lake status data". Quat. Sci. Rev. 17
1814 (6-7): 629-657. https://doi.org/10.1016/S0277-3791(98)00015-8
1815
1816 Jolly, D., Prentice, I. C., Bonnefille, R. et al. (1998b). 'Biome reconstruction from pollen and plant
1817 macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years'. J. Biogeography. 25 (6):
1818 1007-1027
1819
1820 Joussaume, S. & Braconnot, P. (1997). 'Sensitivity of paleoclimate simulation results to season
1821 definitions'. J. Geophys. Res.-Atmos. 102: 1943-1956
1822

Joussaume, S. & Taylor K. E. (1995). 'Status of the Paleoclimate Modeling Intercomparison Project'. In: Proceedings of the First International AMIP Scientific Conference, WCRP-92 425. 430 Monterey, USA

Joussaume, S., Taylor, K. E., Braconnot, P. et al. (1999). 'Monsoon changes for 6000 years ago: Results of 18 simulations from the Paleoclimate Modeling Intercomparison Project (PMIP)'. GRL. 26 (7): 859-862. https://doi.org/10.1029/1999GL900126

Kageyama, M., Albani, S., Braconnot, P. et al. (2017). 'The PMIP4 contribution to CMIP6 – Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments'. GMD. 10: 4035-4055. https://doi.org/10.5194/gmd-10-4035-2017

Kageyama, M., Braconnot, P., Harrison, S. P. et al. (2018). 'The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan'. GMD. 11: 1033-1057. https://doi.org/10.5194/gmd-11-1033-2018

Kuhlbrodt, T., Jones, C. G., Sellar, A. et al. (2018). 'The low resolution version of HadGEM3 GC3.1: Development and evaluation for global climate'. J. Adv. Model. Earth Sy. 10: 2865-2888. https://doi.org/10.1029/2018MS001370

Kutzbach, J. E., Liu, X., Liu, Z. & Chen, G. (2008). 'Simulation of the evolutionary response of global summer monsoons to orbital forcing over the past 280,000 years'. Clim. Dyn. 30: 567-579. DOI: 10.1007/s00382-007-0308-z

Kohfeld, K. E., Graham, R. M., de Boer, A. M. et al. (2013). 'Southern Hemisphere westerly wind changes during the Last Glacial Maximum: paleo-data synthesis'. Quat. Sci. Rev. 68: 76-95. https://doi.org/10.1016/j.quascirev.2013.01.017

Lézine, A. M., Hély, C., Grenier, C. et al. (2011). 'Sahara and Sahel vulnerability to climate changes, lessons from Holocene hydrological data'. Quat. Sci. Rev. 30 (21-22): 3001-3012. DOI:10.1016/j.quascirev.2011.07.006

Liu, Z., Zhu, J., Rosenthal, Y. et al. (2014). 'The Holocene temperature conundrum'. PNAS. 111 (34): 3501-3505. DOI: 10.1073/pnas.1407229111

Lunt, D. J., Abe-Ouchi, A., Bakker, P. et al. (2013). 'A multi-model assessment of last interglacial temperatures'. Clim. Past. 9: 699–717. https://doi.org/10.5194/cp-9-699-2013

Lunt, D. J., Foster, G. L., Haywood, A. M. & Stone, E. J. (2008). 'Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels'. Nature. 454 (7208): 1102. DOI: 10.1038/nature07223.

Marcott, S. A., Shakun, J. D., Clark, P. U. & Mix, A. C. (2013). 'A reconstruction of regional and global temperature for the past 11,300 years'. Science. 399 (6124): 1198-1201. DOI: 10.1126/science.1228026.

Marzocchi, A., Lunt, D. J., Flecker, R. et al. (2015). 'Orbital control on late Miocene climate and the North African monsoon: insight from an ensemble of sub-precessional simulations'. Clim. Past. 11 (10): 1271-1295. https://doi.org/10.5194/cp-11-1271-2015

McGee, D., Donohoe, A., Marshall, J. & Ferreira, D. (2014). 'Changes in ITCZ location and cross-equatorial heat transport at the Last Glacial Maximum, Heinrich Stadial 1, and the Mid-Holocene'. Earth and Planetary Science Letters. 390: 69-79. https://doi.org/10.1016/j.epsl.2013.12.043

Menary, M. B., Kuhlbrodt, T., Ridley, J. et al. (2018). 'Pre-industrial control simulations with HadGEM3-GC3.1 for CMIP6'. JAMES. 10: 3049–3075. https://doi.org/10.1029/2018MS001495

New, M., Lister, D., Hulme, M. & Makin, I. (2002). 'A high-resolution data set of surface climate over global land areas'. Clim Res. 21: 2217–2238. DOI:10.3354/cr021001

Otto-Bliesner, B. L., Braconnot, P., Harrison, S. P. et al. (2017). 'The PMIP4 contribution to CMIP6 – Part 2: Two interglacials, scientific objective and experimental design for Holocene and Last Interglacial simulations'. GMD. 10: 3979-4003. https://doi.org/10.5194/gmd-10-3979-2017

Pollard, D. & Reusch, D. B. (2002). 'A calendar conversion method for monthly mean paleoclimate model output with orbital forcing'. J. Geophys. Res. 107 (D22). DOI:10.1029/2002JD002126

Rachmayani, R., Prange, M., Lunt, D. J., et al. (2017). 'Sensitivity of the Greenland Ice Sheet to interglacial climate forcing: MIS 5e versus MIS11'. Paleoceanography. 32 (11): 1089-1101. https://doi.org/10.1002/2017PA003149

Ramstein, G., Fluteau, F., Besse, J. & Joussaume, S. (1997). 'Effect of orogeny, plate motion and land–sea distribution on Eurasian climate change over the past 30 million years'. Nature. 386 (6627): 788. https://doi.org/10.1038/386788a0

Ridley, J., Blockley, E., Keen, A. B. et al. (2017). 'The sea ice model component of HadGEM3-GC3.1'. GMD. 11: 713-723. https://doi.org/10.5194/gmd-11-713-2018

Rind, D. & Peteet, D. (1985). 'Terrestrial conditions at the last glacial maximum and CLIMAP sea-surface temperature estimates: Are they consistent?' Quat. Res. 2: 1-22. DOI:10.1016/0033-5894(85)90080-8

Schmidt, G. A., Annan, J. D., Bartlein, P. J. et al. (2014). 'Using paleo-climate comparisons to constrain future projections in CMIP5'. Clim. Past. 10: 221-250. https://doi.org/10.5194/cp-10-221-2014

Scussolini, P., Bakker, P., Guo, C. et al. (2019). 'Agreement between reconstructed and modeled boreal precipitation of the Last Interglacial'. Sci. Adv. 5 (11): 1-11.
DOI: 10.1126/sciadv.aax7047

Singarayer, J. S. & Burrough, S. L. (2015). 'Interhemispheric dynamics of the African rainbelt during the late Quaternary'. Quaternary Science Reviews. 124: 48-67.
DOI: 10.1016/j.quascirev.2015.06.021

Singarayer, J. S., Valdes, P. J. & Roberts, W. H. G. (2017). 'Ocean dominated expansion and contraction of the late Quaternary tropical rainbelt'. Nature Scientific Reports. 7: 9382.
DOI:10.1038/s41598-017-09816-8

Smith, R. S. & Gregory, J. M. (2009). 'A study of the sensitivity of ocean overturning circulation and climate to freshwater input in different regions of the North Atlantic'. Geophys. Res. Lett. 36.
DOI:10.1029/2009GL038607

Stone, E. J., Capron, E., Lunt, D. J., et al. (2016). 'Impact of meltwater on high-latitude early Last Interglacial climate'. Clim. Past. 12: 1919–1932. https://doi.org/10.5194/cp-12-1919-2016

Storkey, D., Megann, A., Mathiot, P. et al. (2017). 'UK Global Ocean GO6 and GO7: A traceable hierarchy of model resolutions'. GMD. 11: 3187-3213. https://doi.org/10.5194/gmd-11-3187-2018

1932 Taylor, K. E., Stouffer, R. J. & Meehl, G. A. (2011). 'An overview of CMIP5 and the experiment
1933 design'. Bull. Am. Meteorol. Soc. 93: 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1
1934
1935 Turney, C. S. M. & Jones, R. T. (2010). 'Does the Agulhas Current amplify global temperatures
1936 during super-interglacials?' J. Quat. Sci. 25 (6): 839-843. https://doi.org/10.1002/jqs.1423
1937
1938 Walters, D. N., A., Baran, I., Boutle, M. E. et al. (2017). 'The Met Office Unified Model Global
1939 Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations'. GMD. 12: 1909-1923.
1940 https://doi.org/10.5194/gmd-12-1909-2019
1941
1942 Wang, X., et al. (2006). 'Interhemispheric anti-phasing of rainfall during the last glacial period'.
1943 Quat. Sci. Rev. 25: 3391-3403. DOI: 10.1016/j.quascirev.2006.02.009
1944
1945 Wang, Y., Cheng, H., Edwards, R.L., et al. (2008). 'Millennial-and orbital-scale changes in the East
1946 Asian monsoon over the past 224,000 years'. Nature. 451 (7182): 1090. DOI: 10.1038/nature06692
1947
1948 Wang, P. X., et al. (2014). 'The global monsoon across timescales: coherent variability of regional
1949 monsoons'. Clim. Past. 10: 2007-2052. https://doi.org/10.5194/cp-10-2007-2014
1950
1951 Williams, K. D., Copsey, D., Blockley E. W., et al. (2017). 'The Met Office Global Coupled Model
1952 3.0 and 3.1 (GC3.0 and GC3.1) Configurations'. JAMES. 10 (2): 357-380.
1953 https://doi.org/10.1002/2017MS001115
1954

LIST OF TABLES

LIST OF FIGURES

Figure 4 – Annual mean meridional overturning streamfunction climatologies from HadGEM3: a-c) Atlantic basin; d-f) Global. Top row: *piControl* simulation; Middle row: *midHolocene* simulation; Bottom row: *lig127k* simulation

Figure 5 – Calendar adjusted JJA daily surface rainfall & 850mb wind climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a) *midHolocene* – *piControl*; b) *lig127k* – *piControl*; c) *lig127k* – *midHolocene*

Figure 6 – Calendar adjusted JJA daily rainfall climatology by latitude, averaged over West Africa (20°W-15°E, land points only), for the various generations of the UK's physical climate model: a) Absolute values; b) Anomalies (MH or LIG – PI). Solid lines show PI simulations, dashed lines show MH simulations and dotted lines show LIG simulations

~~Fig_latrain_prod – JJA rainfall differences by latitude, averaged over West Africa (20°W-30°E, including both land and ocean points), HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run, 100-year climatology from each year~~

~~Fig_proxy_mh_loc – Simulated versus proxy MAT and MAP anomalies. Left-hand side panels show simulated gridded anomalies from HadGEM3 (*midHolocene* production run – *piControl* production run, 100-year climatology from each), right-hand side panels show proxy data from Bartlein *et al.* (2011) (MH – preindustrial). Proxy data locations are projected onto model grid: a) Simulated MAT; b) Proxy MAT; c) Simulated MAP; d) Proxy MAP~~

Figure 7 – Calendar adjusted mean annual surface air temperature anomalies from simulated model data versus proxy data. Background data show simulated anomalies (MH – PI) from different generations of the same model: a) Proxy data anomalies (MH – PI) from Bartlein et al. (2011), with locations projected onto model grid; b) HadGEM3; c) HadGEM2-ES; d) HadCM3

Figure 8 – Same as Figure 7, but for rainfall anomalies

Figure 9 - Calendar adjusted SST anomalies from model simulated data versus proxy data. Background data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Capron et al. (2017) and triangles show anomalies from Hoffman et al. (2017). Proxy data locations are projected onto model grid: a-c) HadGEM3; d-f) HadCM3. Top row: Annual; Middle row: Northern Hemisphere summer (JAS); Bottom row: Southern Hemisphere summer (JFM)

Figure 10 - Calendar adjusted annual surface rainfall anomalies from model simulated data versus proxy data. Background data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Scussolini et al. (2019). Proxy data locations are projected onto model grid: a) HadGEM3; b) HadCM3. Inset shows semi-quantitative scale of proxy data, adapted from Scussolini et al. (2019)

~~Fig_latrain_gen – JJA daily rainfall climatology differences (MH and LIG-PI) by latitude, averaged over West Africa (20°W-30°E, including both land and ocean points), for the various generations of the UK's physical climate model, 100-year climatology from each (50-year climatology for HadCM3 LIG). Solid lines show MH simulations, dotted lines show LIG simulations. Note that due to the low spatial resolution in HadCM3, values in between latitude points have been interpolated~~

~~Fig_wafricarain_gen_mh – JJA daily rainfall climatology differences (MH-PI) for the various generations of the UK's physical climate model, 100-year climatology from each: a) HadGEM3; b) HadGEM2-ES; c) HadCM3~~

~~Fig_wafricarain_gen_lig – JJA daily rainfall climatology differences (LIG-PI) for the various generations of the UK's physical climate model, 100-year climatology from HadGEM3, 50-year climatology from HadCM3: a) HadGEM3; b) HadCM3~~

~~Fig_greening_prod – Annual mean rainfall over West Africa, zonally averaged from 20°W-30°E, HadGEM3 and CMIP5 *midHolocene* production run minus corresponding *piControl* production runs, 100-year climatology. Solid line shows HadGEM3, dotted lines show CMIP5 simulations. Grey dashes show maximum and minimum bounds of the increase in rainfall required to support grassland at each latitude, within which simulations must lie if producing enough rainfall to support grassland~~

LIST OF SUPPLEMENTARY MATERIAL FIGURES

SM1 – Latitude-month insolation (incoming SW radiative flux) anomalies, using modern calendar: a) *midHolocene - piControl*; b) *lig127k – piControl*

SM2 - Annual global mean atmospheric fields from HadGEM3 *piControl*, *midHolocene* and *lig127k* spin-up phases: a) 1.5 m air temperature; b) TOA radiation balance. Thin lines in b) show annual TOA radiation balance, thick lines show 11-year running mean. Note that the *piControl* spin-up phase was run in three separate parts, to accommodate for minor changes/updates in the model as the simulation progressed. Note also that the first ~50 years of the *lig127k* simulation have been deliberately removed from this figure, because a number of model crashes caused the model to be initially unstable and give highly varied global mean temperatures.

SM3 – Centennial trends in 1.5m temperature for HadGEM3 warm climate simulations' spin-up phases, last 100 years only: a) *midHolocene* ; b) *lig127k*. Stippling shows statistical significance (as calculated by a Mann-Kendall test) at the 99% level

SM4 - Annual global mean (full depth) oceanic fields from HadGEM3 *piControl*, *midHolocene* and *lig127k* spin-up phases: a) OceTemp; b) OceSal

SM5 – Modern calendar 1.5 m air temperature climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a) *midHolocene – piControl*, JJA; b) *midHolocene – piControl*, DJF; c) *lig127k – piControl*, JJA; d) *lig127k – piControl*, DJF. Stippling shows statistical significance (as calculated by a Student's T-test) at the 99% level

SM6 – Annual mean sea-ice climatology differences, HadGEM3 *midHolocene* production run versus HadGEM3 *piControl* production run. Stippling shows statistical significance (as calculated by a Student's T-test) at the 99% level

SM7 – Annual mean rainfall over West Africa (averaged over 20°W-30°E, consistent with Joussaume *et al*. [1999]), HadGEM3 *midHolocene* simulation minus corresponding *piControl*, and likewise for previous models from CMIP5. Solid line shows HadGEM3, dotted lines show CMIP5 simulations. Grey dashes show maximum and minimum bounds of the increase in rainfall required to support grassland at each latitude, within which simulations must lie if producing enough rainfall to support grassland (adapted from Figure 3a in Joussaume *et al*. [1999])

TABLES

	piControl	*midHolocene*	*lig127k*
Astronomical parameters			
Eccentricity	0.016764	0.018682	0.039378
Obliquity	23.459	24.105°	24.04°
Perihelion-180°	100.33	0.87°	275.41°
Date of vernal equinox	March 21 at noon	March 21 at noon	March 21 at noon
Trace gases			
CO_2	284.3 ppm	264.4 ppm	275 ppm
CH_4	808.2 ppb	597 ppb	685 ppb
N_2O	273 ppb	262 ppb	255 ppb
Other GHG gases	CMIP DECK *piControl*	CMIP DECK *piControl*	CMIP DECK *piControl*

Table 1 - Astronomical parameters and atmospheric trace gas concentrations used in HadGEM3 simulations

Variable	***piControl***	***midHolocene***	***lig127k***
TOA (W m^2)	-0.002	-0.05	-0.06
1.5 m air temp (°C)	0.03	-0.06	-0.16
OceTemp (°C)	0.035	0.03	0.03
OceSal (psu)	0.0001	-0.0004	0.00007

Table 2 - Trends (per century) in global mean measures of climate equilibrium for the last hundred years of the simulations, adapted from and including *piControl* results from Menary *et al*. (2018)

Time period	**Means (°C)**			**Anomalies (°C)**	
	piControl	***midHolocene***	***lig127k***	***midHolocene – piControl***	***lig127k – piControl***
Annual	13.8	13.67	14.29	-0.12	0.49
JJA	15.68	15.75	17.37	0.07	1.69
DJF	11.86	11.55	11.39	-0.31	-0.47

Table 3 - Global 1.5 m air temperature means and anomalies from HadGEM3 *piControl*, *midHolocene* and *lig127k* production runs

a)

Metric	Simulations vs proxy data			Simulations vs simulations	
	HadGEM3	HadGEM2-ES	HadCM3	HadGEM2-ES v HadGEM3	HadCM3 v HadGEM3
MAT (°C)	2.45	2.42	2.37	0.65	0.57
No. of locations	*638*			*Global coverage*	
MAP (mm year^{-1})	285.9	293.5	304.7	90.8	121.8
No. of locations	*651*			*Global coverage*	

b)

Metric	Simulations vs proxy data					
	Yearly		JAS		JFM	
	HadGEM3	HadCM3	HadGEM3	HadCM3	HadGEM3	HadCM3
SST from Capron *et al.* (2017)	3.03	3.04	3.03	2.98	2.81	2.62
No. of locations	*3*		*24*		*15*	
SST from Hoffman *et al.* (2017)	2.42	3.02	1.99	2.78	4.28	3.97
No. of locations	*86*		*12*		*6*	

Table 4 - RMSE values (for various metrics) between simulations from different generations of the same model versus proxy data, and versus each other: a) MAT and MAP from the MH simulations versus proxy data from Bartlein et al. (2011); b) SST from the LIG simulations versus proxy data from Capron et al. (2017) and Hoffman et al. (2017). Regarding the proxy data comparisons in b), for JAS the simulated SST anomalies are compared to Northern Hemisphere summer reconstructions and for JFM the simulated SST anomalies are compared to Southern Hemisphere summer reconstructions

FIGURES

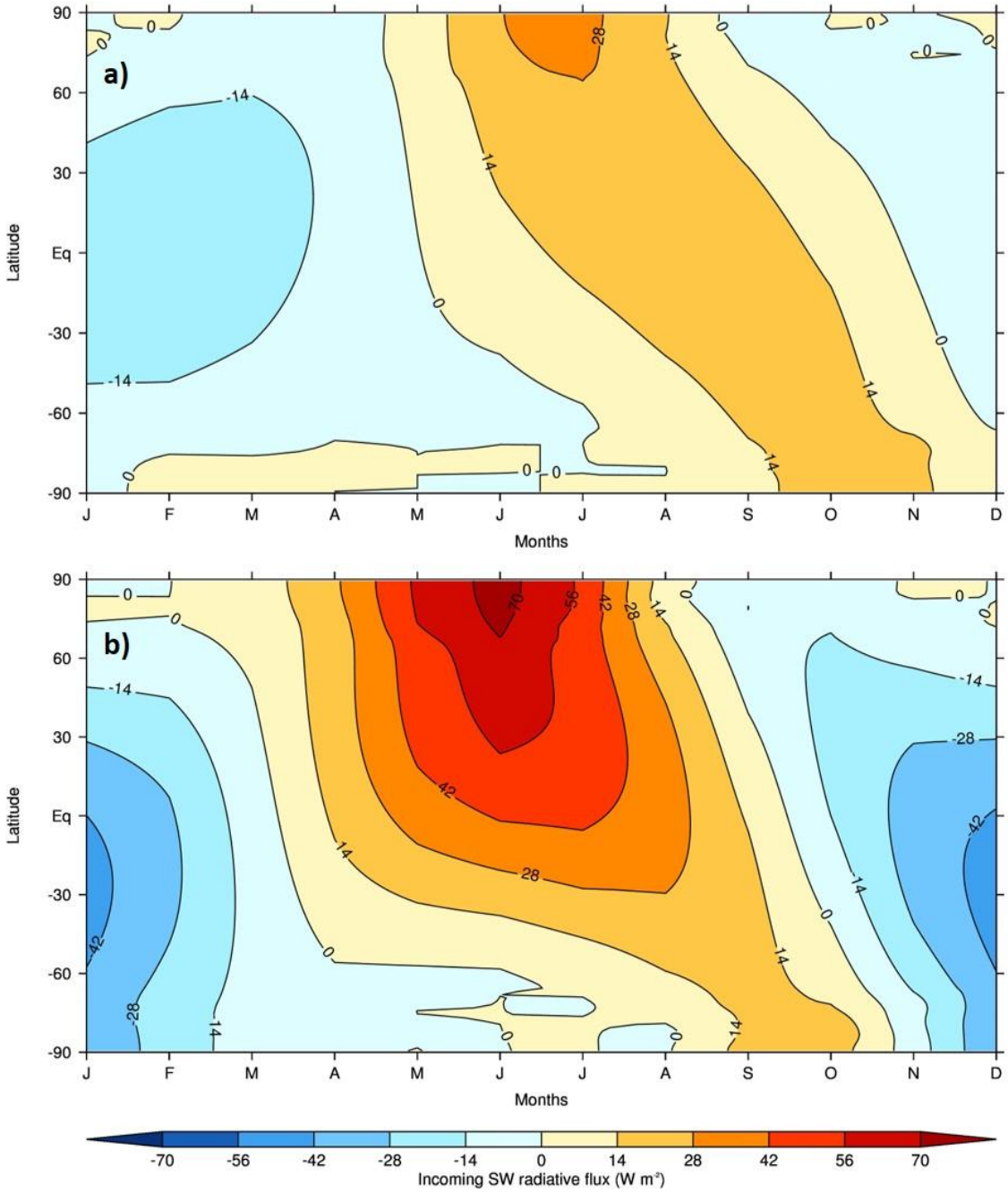

Figure 1 - Calendar adjusted ~~Latitude~~latitude-month insolation (incoming SW radiative flux) anomalies: a) *midHolocene - piControl*; b) *lig127k - piControl*

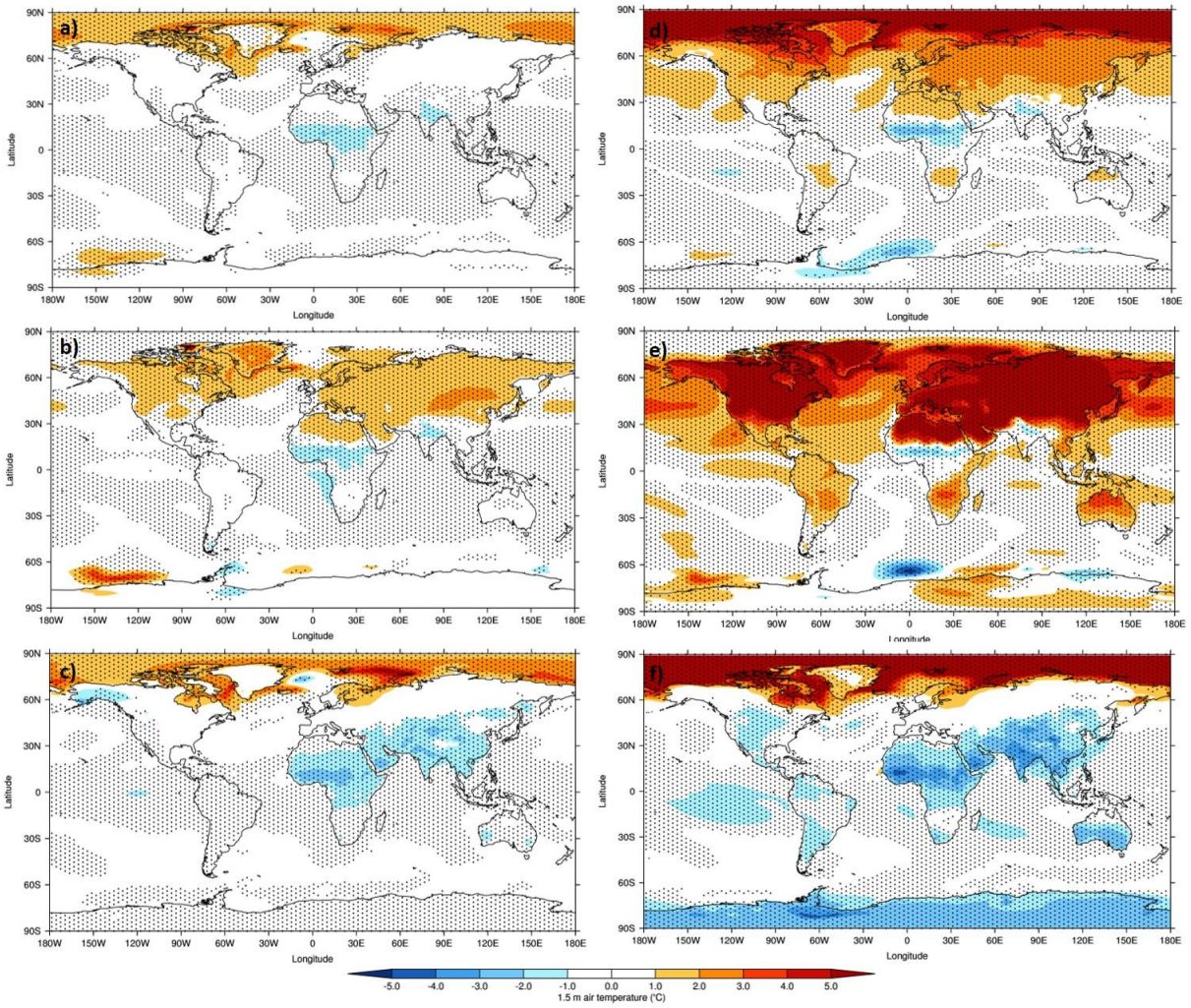

Figure 2 – Calendar adjusted 1.5 m air temperature climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a-c) *midHolocene – piControl*; d-f) *lig127k – piControl*. Top row: Annual; Middle row: Northern Hemisphere summer (JJA); Bottom row: Northern Hemisphere winter (DJF). Stippling shows statistical significance (as calculated by a Student's T-test) at the 99% level

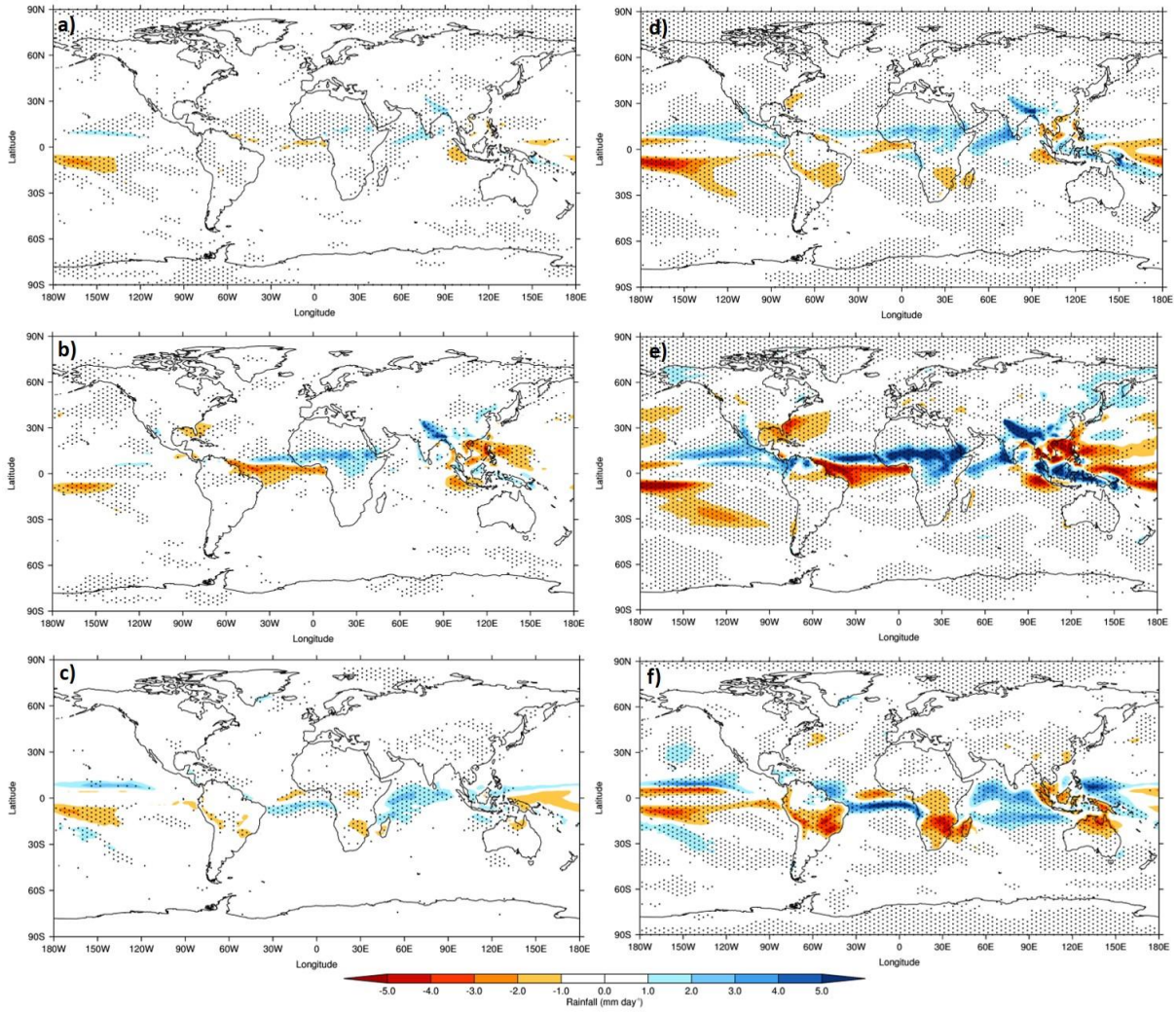

Figure 3 – Same as Figure 2, but for daily surface rainfall differences

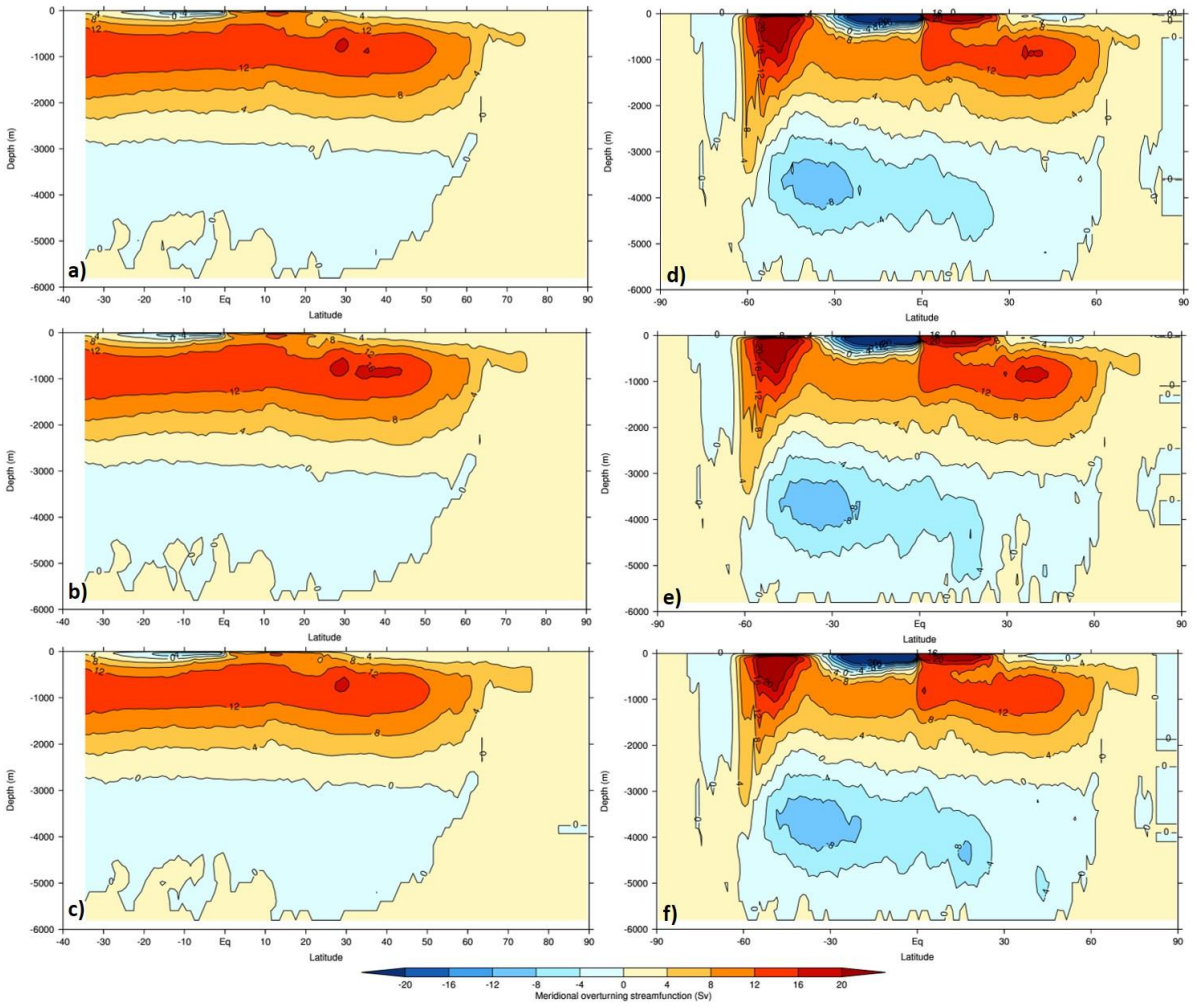

Figure 4 – Annual mean meridional overturning streamfunction climatologies from HadGEM3: a-c) Atlantic basin; d-f) Global. Top row: *piControl* simulation; Middle row: *midHolocene* simulation; Bottom row: *lig127k* simulation

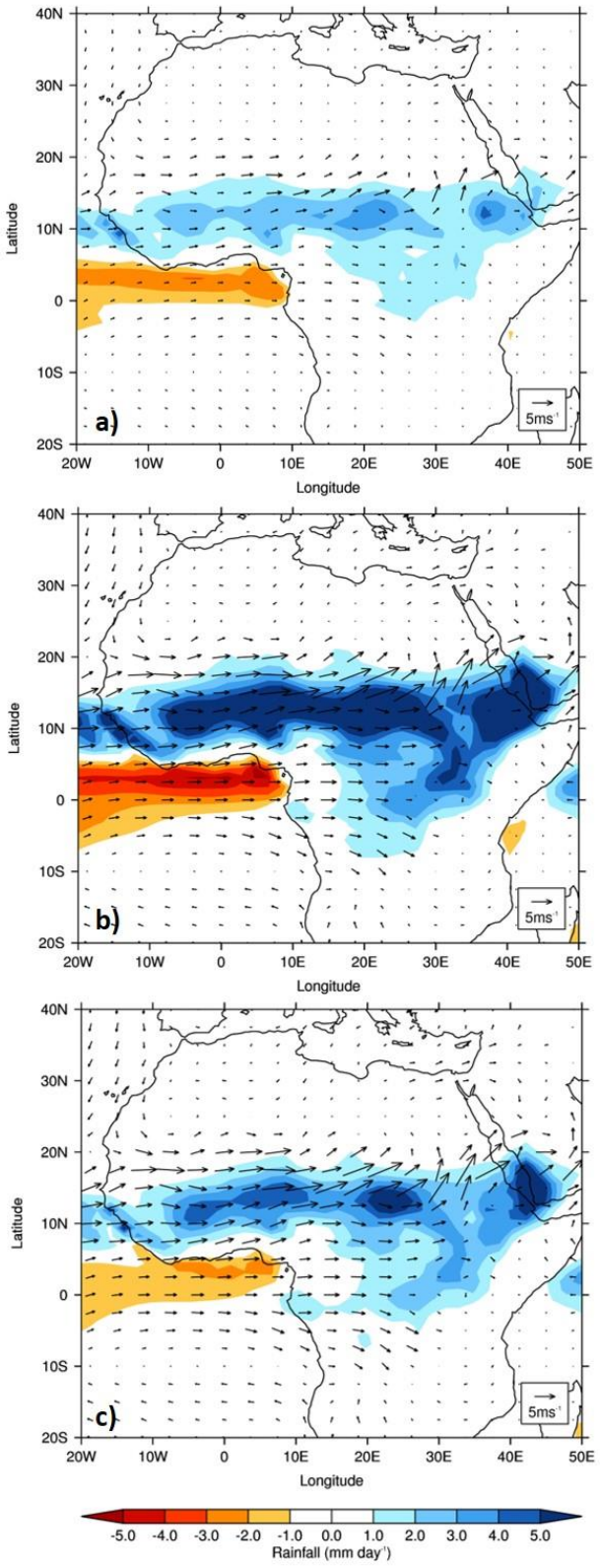

Figure 5 – Calendar adjusted JJA daily surface rainfall ~~&~~ and 850mb wind climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a) *midHolocene* – *piControl*; b) *lig127k* – *piControl*; c) *lig127k* – *midHolocene*

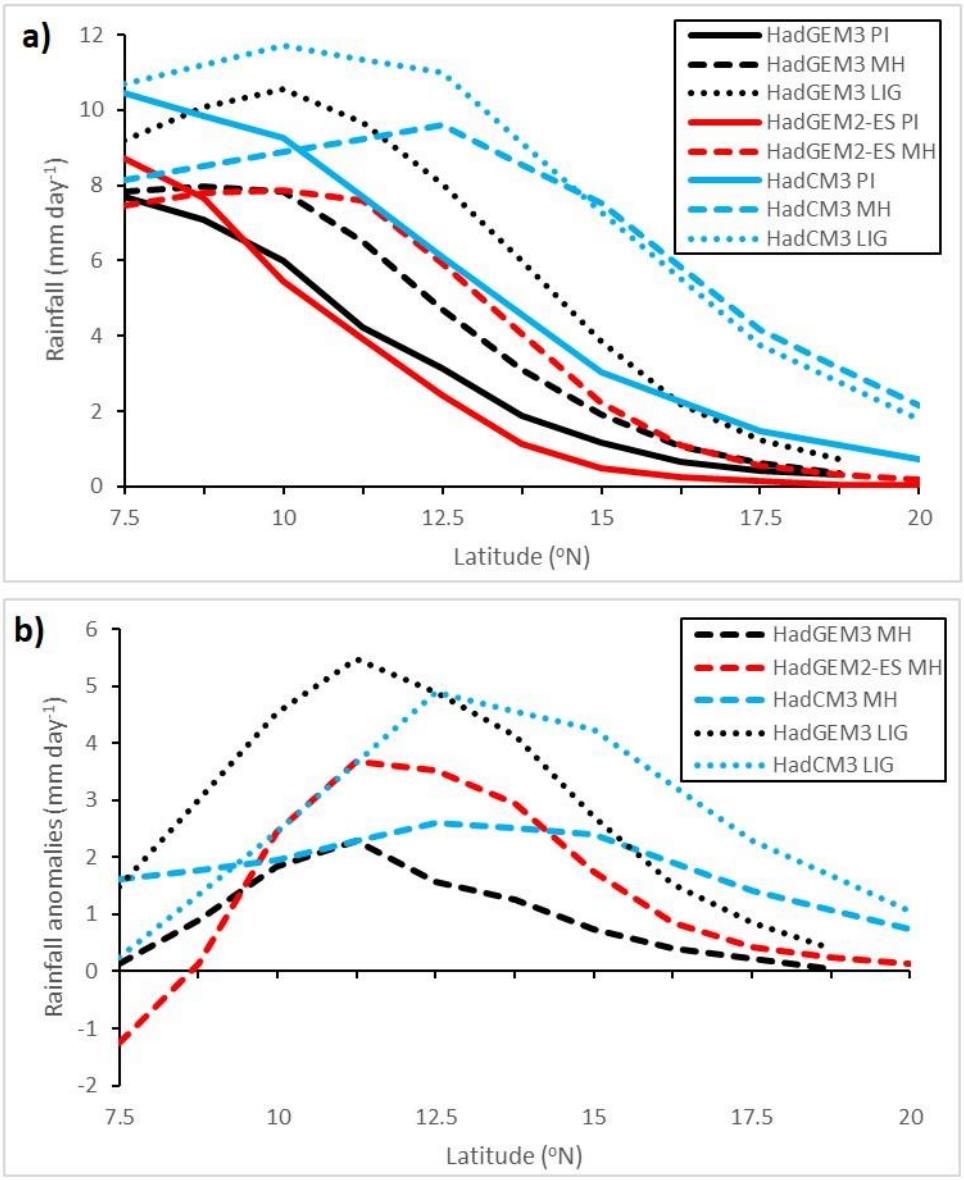

Figure 6 – Calendar adjusted JJA daily rainfall climatology by latitude, averaged over West Africa (20°W-15°E, land points only), for the various generations of the UK's physical climate model: a) Absolute values; b) Anomalies (MH or LIG – PI). Solid lines show PI simulations, dashed lines show MH simulations and dotted lines show LIG simulations

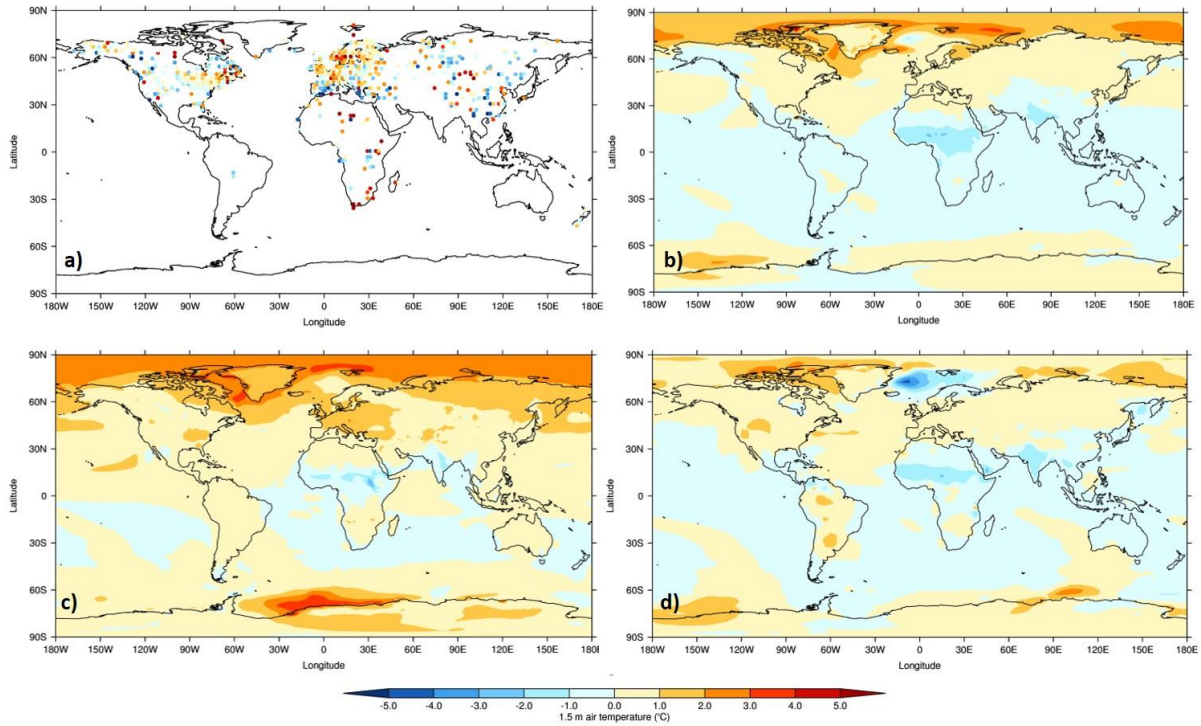

Figure 7 – Calendar adjusted ~~Mean~~ mean annual surface air temperature anomalies from simulated model data versus proxy data ~~versus calendar adjusted simulated anomalies~~. Background ~~gridded~~ data show simulated anomalies (MH – PI) from different generations of the same model: a) Proxy data anomalies (MH – PI) from Bartlein et al. (2011), with locations projected onto ~~HadGEM3~~ model grid; b) HadGEM3; c) HadGEM2-ES; d) HadCM3

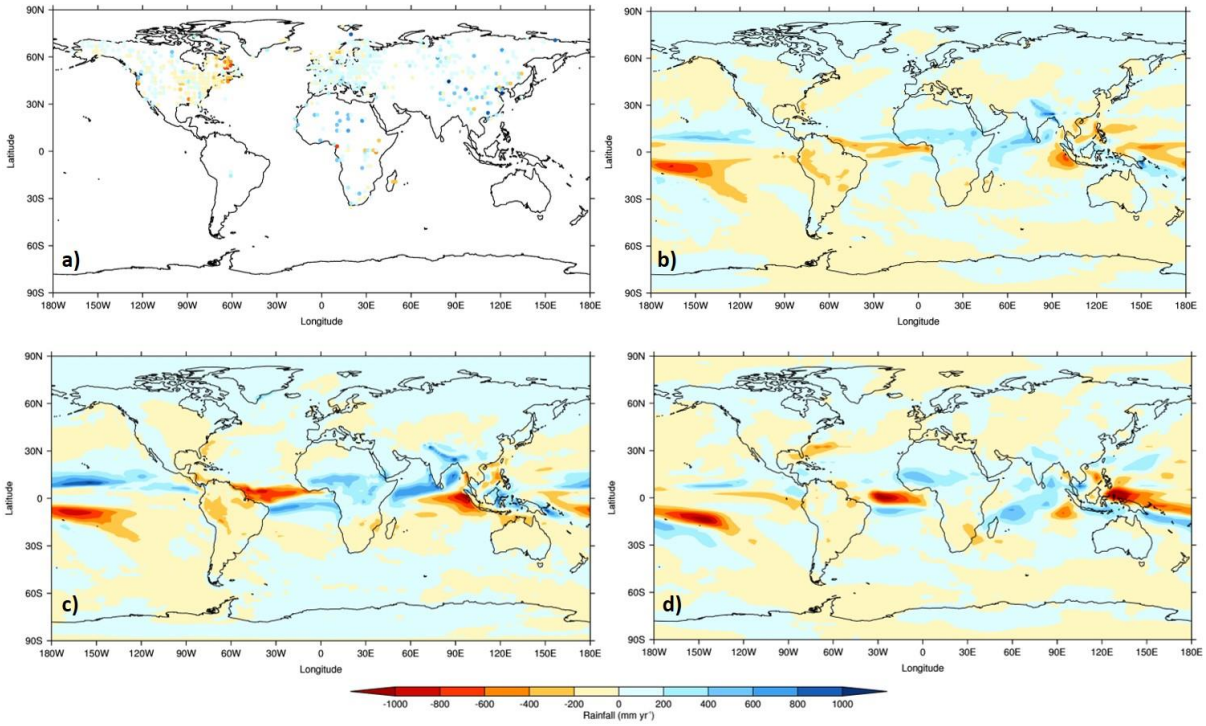

Figure 8 – Same as Figure 7, but for rainfall anomalies

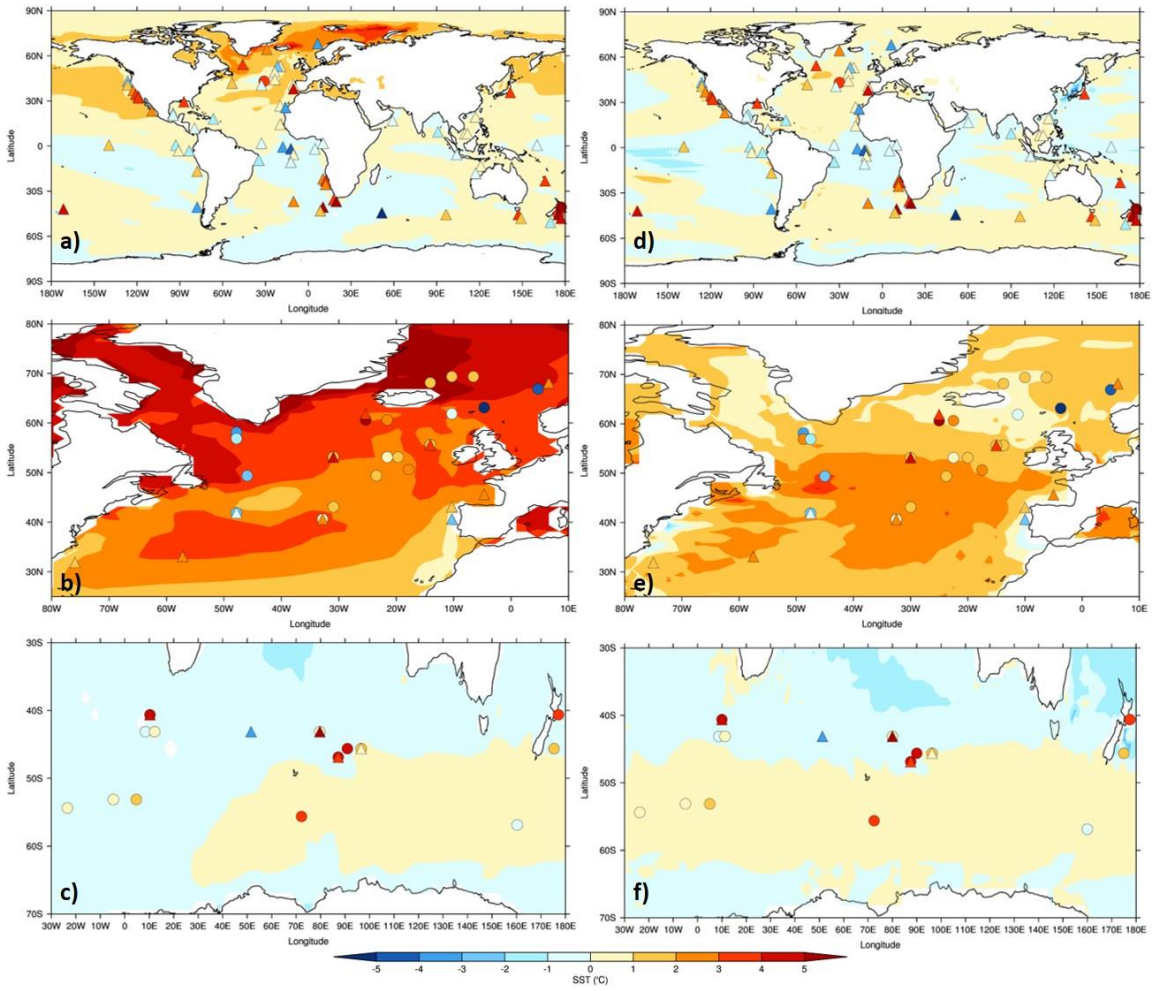

Figure 9 - Calendar adjusted SST anomalies from model simulated data versus proxy data. Background ~~gridded~~ data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Capron et al. (2017) and triangles show anomalies from Hoffman et al. (2017). Proxy data locations are projected onto model grid: a-c) HadGEM3; d-f) HadCM3. Top row: Annual; Middle row: Northern Hemisphere summer (JAS); Bottom row: Southern Hemisphere summer (JFM)

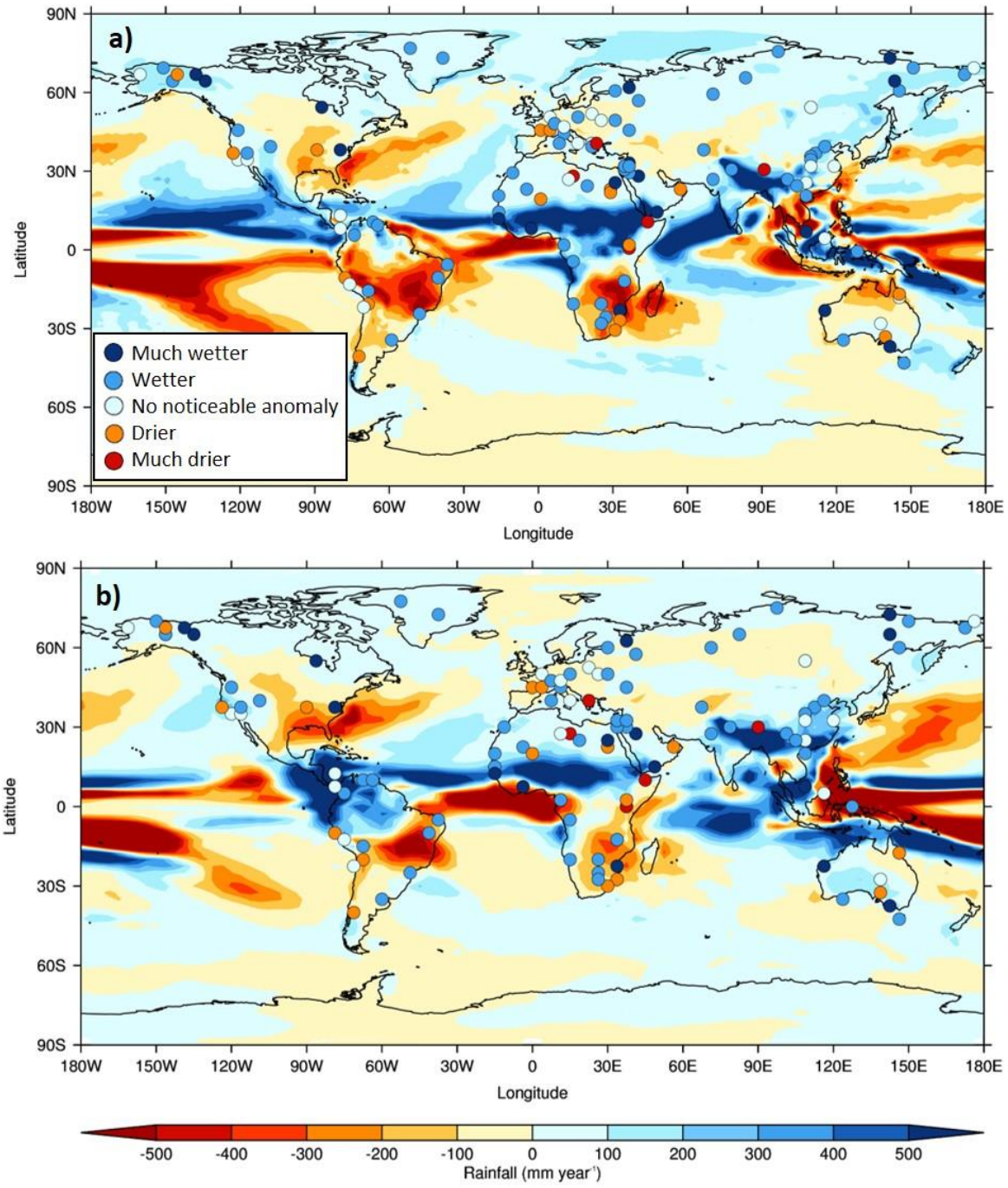

Figure 10 - Calendar adjusted annual surface rainfall anomalies from model simulated data versus proxy data. Background ~~gridded~~ data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Scussolini et al. (2019). Proxy data locations are projected onto model grid: a) HadGEM3; b) HadCM3. Inset shows semi-quantitative scale of proxy data, adapted from Scussolini et al. (2019)