CMIP6/PMIP4 simulations of the mid-Holocene and Last Interglacial using HadGEM3: comparison to the pre-industrial era, previous model versions, 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 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 PI and the two paleo 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). 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.

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 interglacial simulations: the mid-Holocene (MH) at ~6 ka and the Last Interglacial (LIG) at ~127 ka (although spanning ~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 has been investigated by a number of model simulations, 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).

The driving mechanism producing the climate and environmental changes indicated by the palaeodata for the MH and LIG 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, 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 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. In addition to a global assessment, a secondary focus of this paper is on the fidelity of the temperature anomalies and the degree of precipitation enhancement in the Sahara, the latter of which has proved problematic for several generations of models. Following this introduction, Section 2 describes the model, the experimental design, 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 simulation comparison and following with the model-data comparison, and finally section 4 summarises and concludes.

2. MODEL, EXPERIMENT DESIGN, 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 warm climate simulations conducted here, and the *piControl* simulation (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. 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, 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*. 2020a). 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*. [2020a]) as they are not directly affected by medium-frequency climate processes.

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), full-depth ocean temperature (OceTemp) and full-depth ocean salinity (OceSal) (Menary *et al*. 2018). See Section 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. Once the simulations were considered in an acceptable level of equilibrium (see Section 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 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, annual global mean 1.5 m air temperature and TOA radiation from both warm climate simulations, compared to the *piControl*, are summarised in Table 2; see Supplementary Material (SM2) for the timeseries of these fields. For the warm climate simulations, despite considerable interannual variability and arguably more so than in the *piControl* (see SM2), both are showing 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). 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 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.

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). OceTemp is steadily increasing throughout the *piControl*, and this continues in both warm climate simulations, whereas there is a dramatic fall in ocean salinity in these simulations (SM4). 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* it is not meeting the salinity criterion (-0.0004 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

3.1. Production runs results

The warm climate production runs were undertaken following the spin-up phase, with 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 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, the focus is on two separate measures: i) to describe and understand the differences between the two most recent 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 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'.

3.1.1. 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 annual and seasonal mean summer/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 simulation 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 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 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*. [2020b]).

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 calendar adjusted mean JJA (the primary monsoon region) 850mb wind anomalies (relative to the *piControl*). 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 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, drawing in more moisture from the tropical Atlantic, which are consistent with previous work documenting the intensified monsoon circulation (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 central rainbelt is more pronounced in the *lig127k* simulation (Figure 5c).

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 is 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).

3.1.2. 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, here we bring in a comparison with newly-available 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 vs simulated data and HadGEM3 vs previous versions. The aim of this is to firstly 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 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). 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 at many locations suggests a warmer MH relative to PI, is consistent with previous work using other models (e.g. 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, 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 indicate 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, 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 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 model, and indeed previous generations, also appears to be seasonally dependent, with the most recent *lig127k* 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.

3.1.3. 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 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 available proxy data. 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 being 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 the CMIP6/PMIP4 protocol. Both simulations were run for a 350-400 year spin-up phase, during which atmospheric and oceanic equilibrium were assessed, and once an acceptable level of equilibrium had been reached, the production runs were started.

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 one 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. 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. 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. 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 practical reasons some of the boundary conditions were left as PI, such as vegetation, 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), located at https://esgf-node.llnl.gov/projects/cmip6/, but are not yet publicly available. The simulations are, however, available to the public 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.: 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, 2011

Berger, A. L.: 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, 1978

Berger, A. L. & Loutre, M. F.: 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, 1991

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

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

Braconnot, P., Otto-Bliesner, B., Harrison, S., et al.: 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, 2007

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

Capron, E., Govin, A., Stone, E. J., et al.: 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, 2014

Capron, E., Govin, A., Feng R. et al.: 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, 2017

Carlson, A. E.: 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, 2008

Drake, N. A., Blench, R. M., Armitage, S. J., et al.: 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, 2011

Fischer, H., Meissner, K. J., Mix, A. C. et al.: 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, 2018

Gates, W. L.: 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, 1976

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

Guarino, M. V., Sime, L., Schroeder, D., et al.: 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, 2020

Guarino, M. V. et al.: 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.: New estimates of radiative forcing due to well mixed greenhouse gases, Geophys. Res. Lett, 25 (14): 2715-2718, https://doi.org/10.1029/98GL01908, 1998

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

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

Haxeltine, A. & Prentice, I. C.: 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, 1996

Haywood, A. M., Dowsett, H. J., Dolan, A. M. et al.: The Pliocene Model Intercomparison Project (PlioMIP) Phase 2: scientific objectives and experimental design, Clim. Past, 12: 663-675, https://doi.org/10.5194/cp-12-663-2016, 2016

Haug, G., Hughen, K. A., Sigman, D. M., et al.: Southward migration of the intertropical convergence zone through the Holocene, Science, 293: 1304-1308, DOI: 10.1126/science.1059725, 2001

Hoelzmann, P., Jolly, D., Harrison, S. P., et al.: Mid-Holocene land-surface conditions in northern Africa and the Arabian Peninsula: A data set for the analysis of biogeophysical feedbacks in the climate system, Global Biogeochemical Cycles, 12 (1): 35-51, https://doi.org/10.1029/97GB02733, 1998

Hoffman, J. S., Clark, P. U., Parnell, A. C., et al.: Regional and global sea-surface temperatures during the last interglaciation, Science, 355: 276-279, DOI: 10.1126/science.aai8464, 2017

Hollingsworth, A., Kållberg, P., Renner, V. & Burridge, D. M.: 'An internal symmetric computational instability, QJRMS, 109: 417–428, https://doi.org/10.1002/qj.49710946012, 1983

Holloway, M. D., Sime, L. C., Singarayer, J. S., et al.: Antarctic last interglacial isotope peak in response to sea ice retreat not ice-sheet collapse, Nature Comms, 7: 12293, https://doi.org/10.1038/ncomms12293, 2016

Jungclaus, J. H., Bard, E., Baroni, M. et al.: The PMIP4 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations, GMD, 10: 4005-4033, https://doi.org/10.5194/gmd-10-4005-2017, 2017

Jolly, D., Harrison, S. P., Damnati, B. & Bonnefille, R.: Simulated climate and biomes of Africa during the Late Quaternary: Comparison with pollen and lake status data, Quat. Sci. Rev, 17 (6-7): 629-657, https://doi.org/10.1016/S0277-3791(98)00015-8, 1998a

Jolly, D., Prentice, I. C., Bonnefille, R. et al.: Biome reconstruction from pollen and plant macrofossil data for Africa and the Arabian peninsula at 0 and 6000 years, J. Biogeography, 25 (6): 1007-1027, https://www.jstor.org/stable/2846197, 1998b

Joussaume, S. & Braconnot, P.: Sensitivity of paleoclimate simulation results to season definitions, J. Geophys. Res.-Atmos, 102: 1943-1956, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/96JD01989, 1997

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

Joussaume, S., Taylor, K. E., Braconnot, P. et al.: 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, 1999

Kageyama, M., Albani, S., Braconnot, P. et al.: 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, 2017

Kageyama, M., Braconnot, P., Harrison, S. P. et al.: 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, 2018

Kuhlbrodt, T., Jones, C. G., Sellar, A. et al.: 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, 2018

Kutzbach, J. E., Liu, X., Liu, Z. & Chen, G.: 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, 2008

Kohfeld, K. E., Graham, R. M., de Boer, A. M. et al.: 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, 2013

Lézine, A. M., Hély, C., Grenier, C. et al.: 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, 2011

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

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

Lunt, D. J., Foster, G. L., Haywood, A. M. & Stone, E. J.: Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO2 levels, Nature, 454 (7208): 1102, DOI: 10.1038/nature07223, 2008

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

Marzocchi, A., Lunt, D. J., Flecker, R. et al.: 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, 2015

McGee, D., Donohoe, A., Marshall, J. & Ferreira, D.: 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, 2014

Menary, M. B., Kuhlbrodt, T., Ridley, J. et al.: Pre-industrial control simulations with HadGEM3-GC3.1 for CMIP6, JAMES, 10: 3049–3075, https://doi.org/10.1029/2018MS001495, 2018

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

Otto-Bliesner, B. L., Braconnot, P., Harrison, S. P. et al.: 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, 2017

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

Rachmayani, R., Prange, M., Lunt, D. J., et al.: 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, 2017

Ramstein, G., Fluteau, F., Besse, J. & Joussaume, S.: 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, 1997

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

Rind, D. & Peteet, D.: 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, 1985

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

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

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

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

Smith, R. S. & Gregory, J. M.: 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, 2009

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

Storkey, D., Megann, A., Mathiot, P. et al.: 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, 2017

Taylor, K. E., Stouffer, R. J. & Meehl, G. A.: An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc, 93: 485-498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2011

Turney, C. S. M. & Jones, R. T.: Does the Agulhas Current amplify global temperatures during super-interglacials?, J. Quat. Sci, 25 (6): 839-843, https://doi.org/10.1002/jqs.1423, 2010

Walters, D. N., A., Baran, I., Boutle, M. E. et al.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, GMD, 12: 1909-1923, https://doi.org/10.5194/gmd-12-1909-2019, 2017

Wang, X., et al.: Interhemispheric anti-phasing of rainfall during the last glacial period, Quat. Sci. Rev, 25: 3391-3403, DOI: 10.1016/j.quascirev.2006.02.009, 2006

Wang, Y., Cheng, H., Edwards, R.L., et al.: Millennial-and orbital-scale changes in the East Asian monsoon over the past 224,000 years, Nature, 451 (7182): 1090, DOI: 10.1038/nature06692, 2008

Wang, P. X., et al.: The global monsoon across timescales: coherent variability of regional monsoons, Clim. Past, 10: 2007-2052, https://doi.org/10.5194/cp-10-2007-2014, 2014

Williams, K. D., Copsey, D., Blockley E. W., et al.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, JAMES, 10 (2): 357-380, https://doi.org/10.1002/2017MS001115, 2017

LIST OF TABLES

Table 1 - Astronomical parameters and atmospheric trace gas concentrations used in HadGEM3 *midHolocene* and *lig127k* simulations

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)

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

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

LIST OF FIGURES

Figure 1 - Calendar adjusted 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 & 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 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)

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

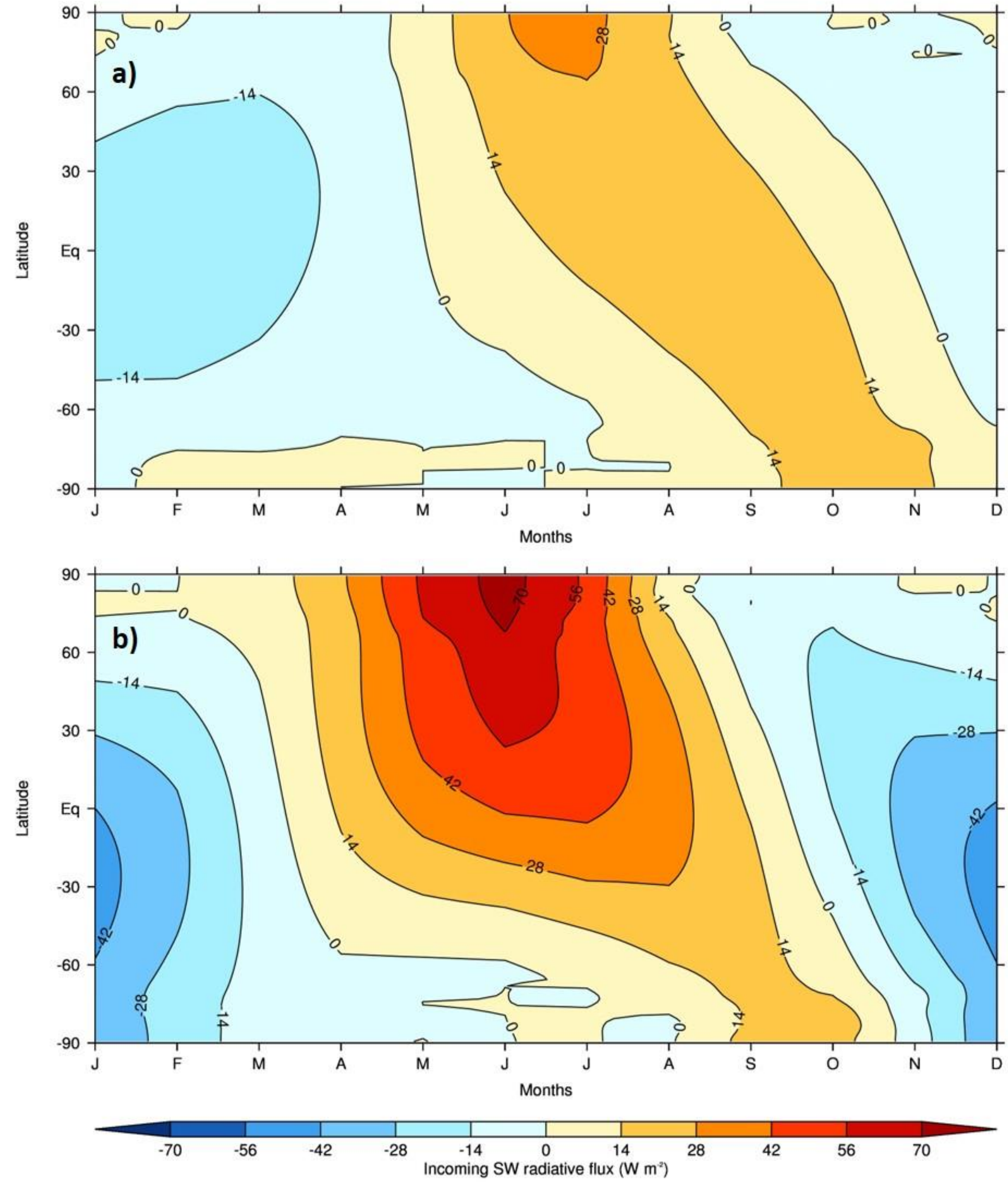

Figure 1 - Calendar adjusted latitude-month insolation (incoming SW radiative flux) anomalies: a) *midHolocene* - *piControl*; b) *lig127k* - *piControl*

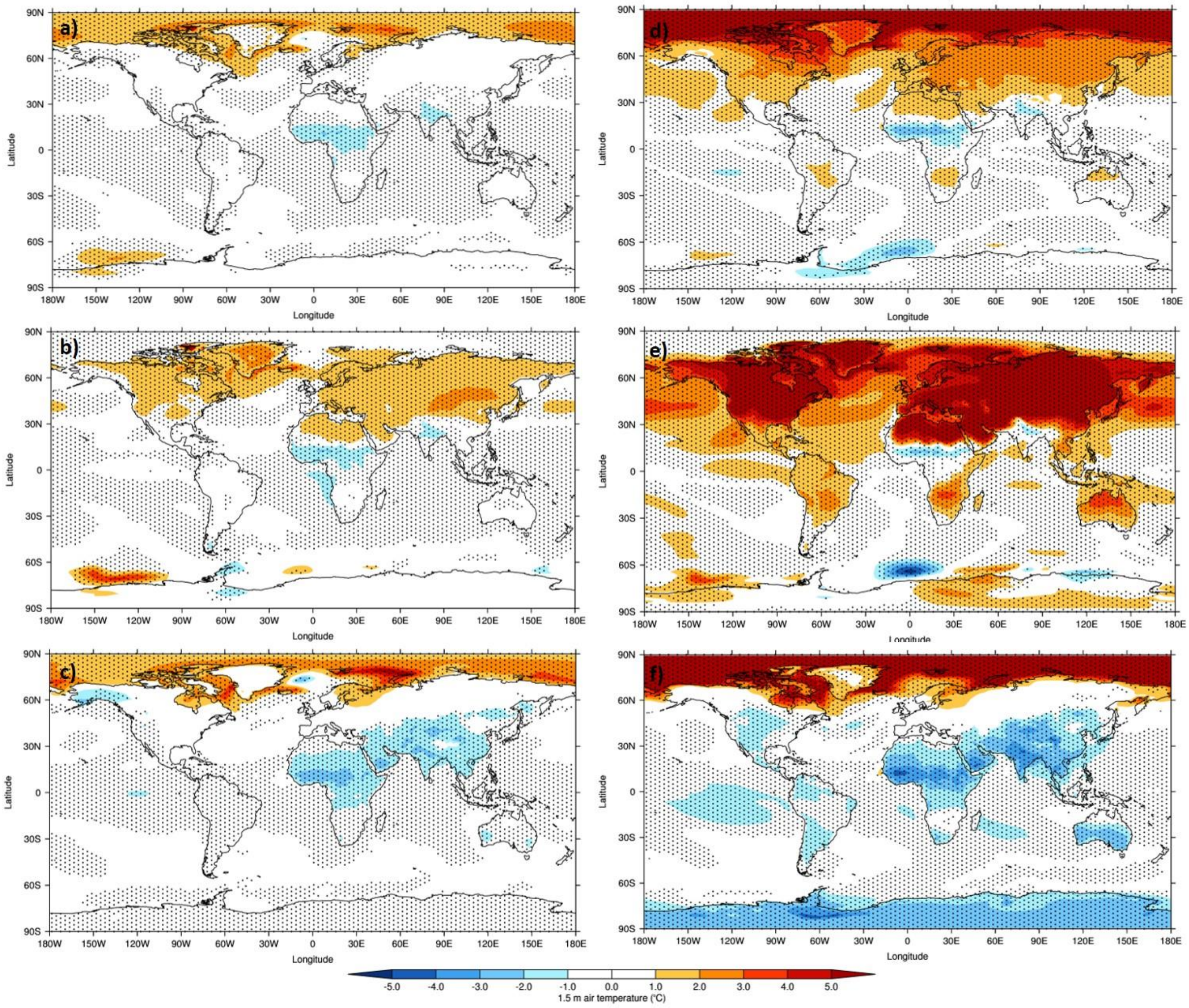

Figure 2 - Calendar adjusted 1.5 m air temperature climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a-c) *midHolocene – piControl*; d-f) *lig127k – piControl*. Top row: Annual; Middle row: Northern Hemisphere summer (JJA); Bottom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

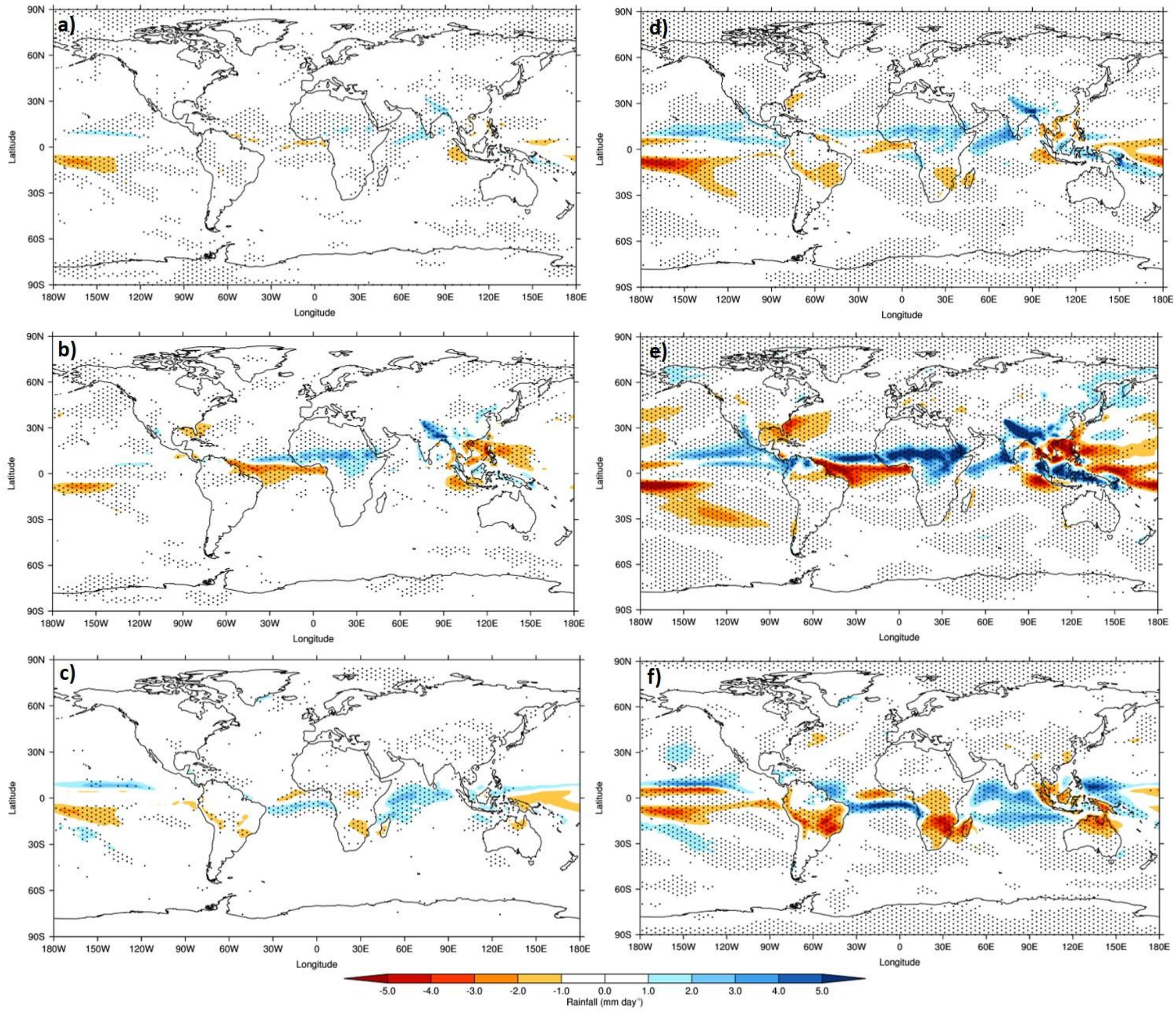

Figure 3 - Same as Figure 2, but for daily surface rainfall differences

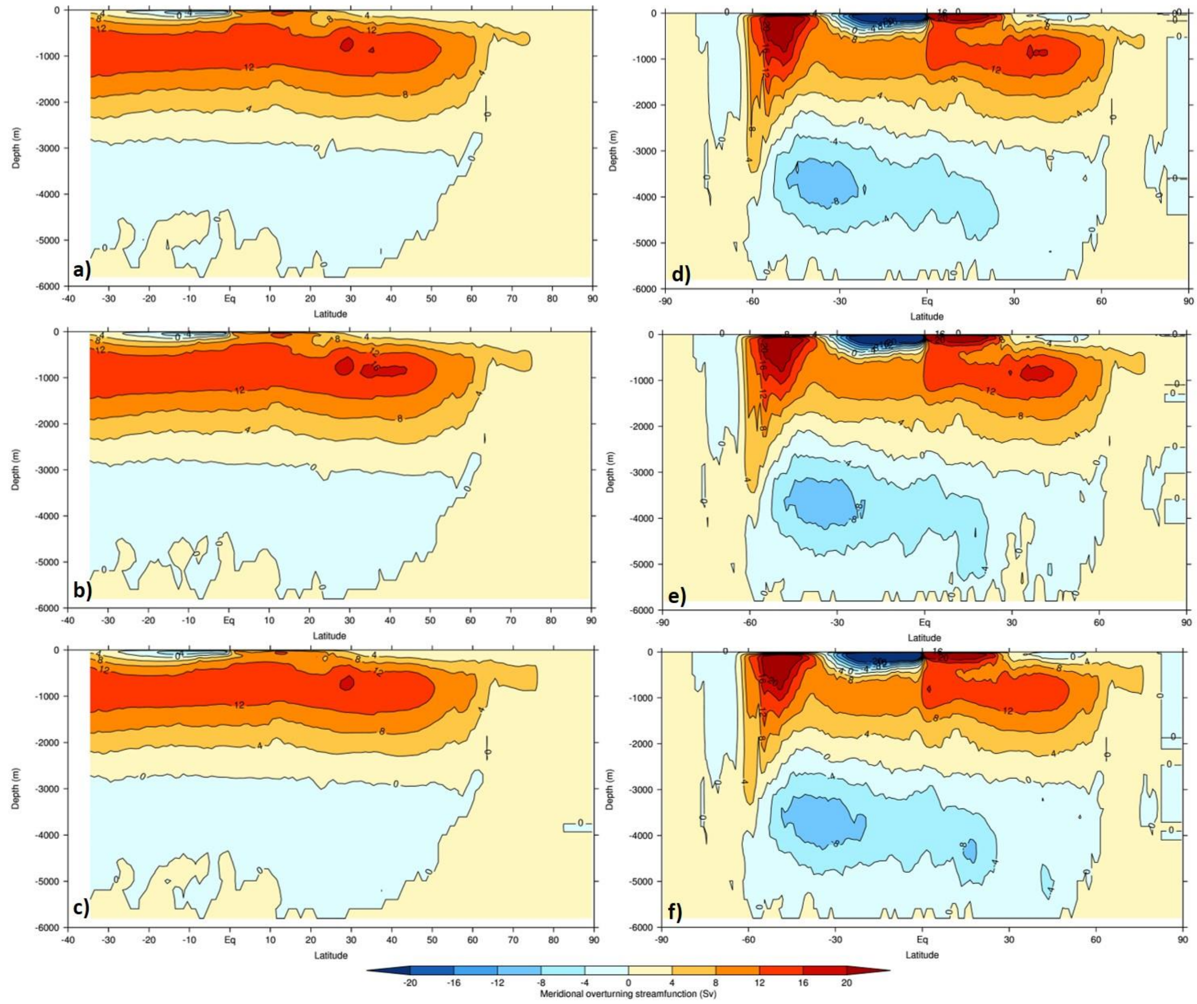

Figure 4 - Annual mean meridional overturning streamfunction climatologies from HadGEM3: a-c) Atlantic basin; d-f) Global. Top row: *piControl* simulation; Middle row: *midHolocene* simulation; Bottom row: *lig127k* simulation

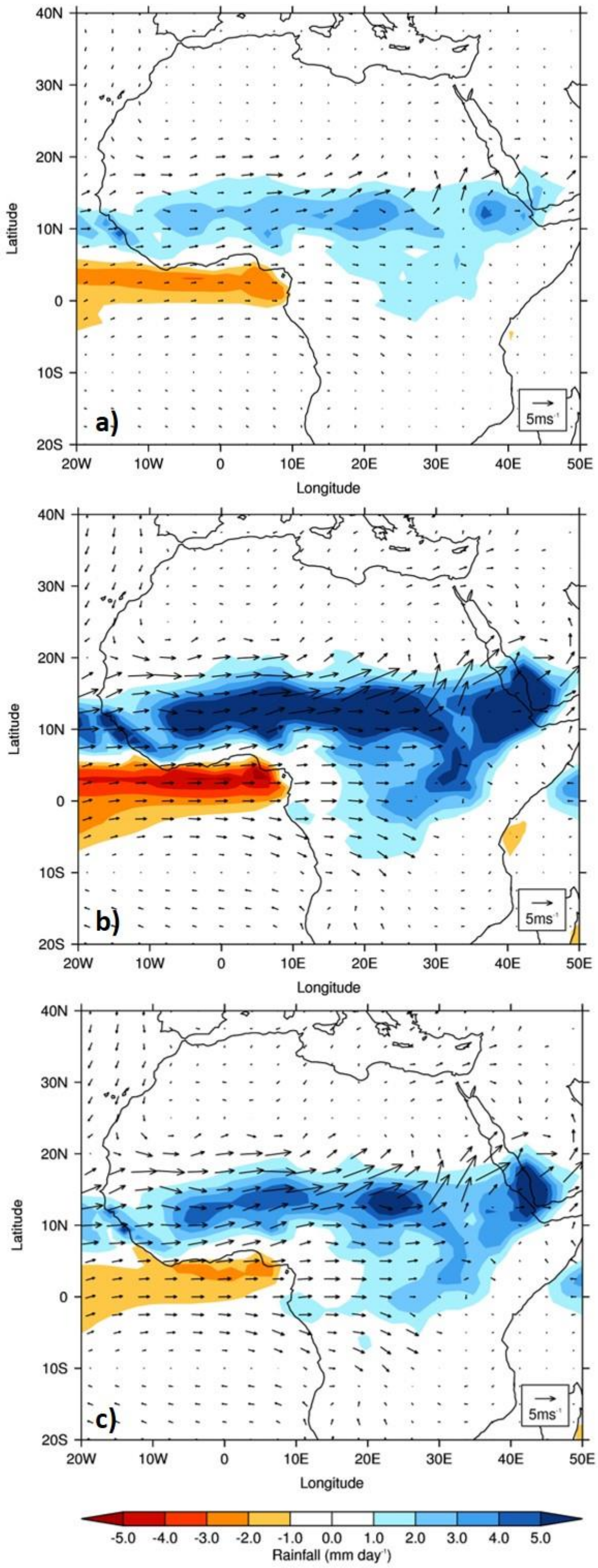

Figure 5 - Calendar adjusted JJA daily surface rainfall and 850mb wind climatology differences, HadGEM3 *midHolocene* and *lig127k* production runs versus HadGEM3 *piControl* production run: a) *midHolocene – piControl*; b) *lig127k – piControl*; c) *lig127k – midHolocene*

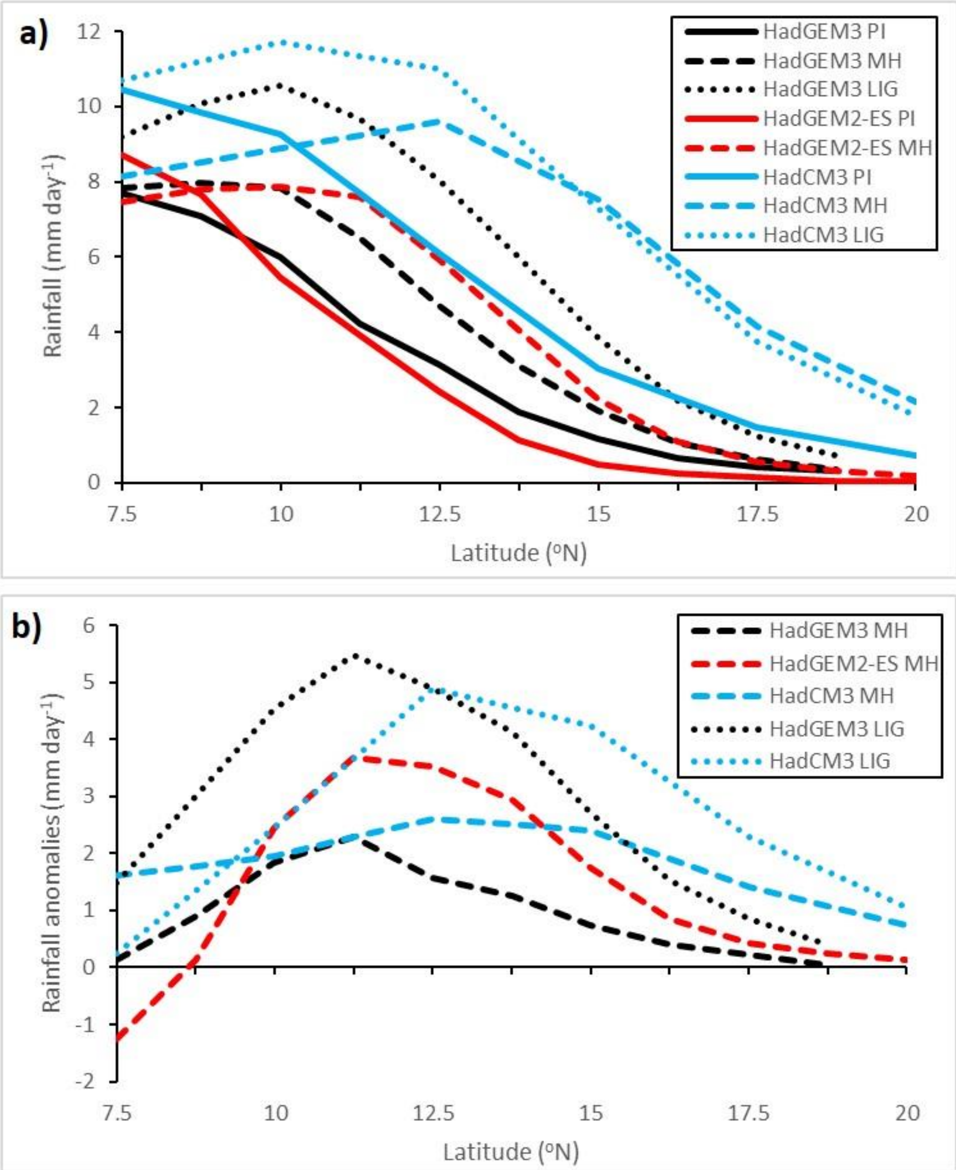

Figure 6 - Calendar adjusted JJA daily rainfall climatology by latitude, averaged over West Africa (20°W-15°E, land points only), for the various generations of the UK's physical climate model: a) Absolute values; b) Anomalies (MH or LIG – PI). Solid lines show PI simulations, dashed lines show MH simulations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

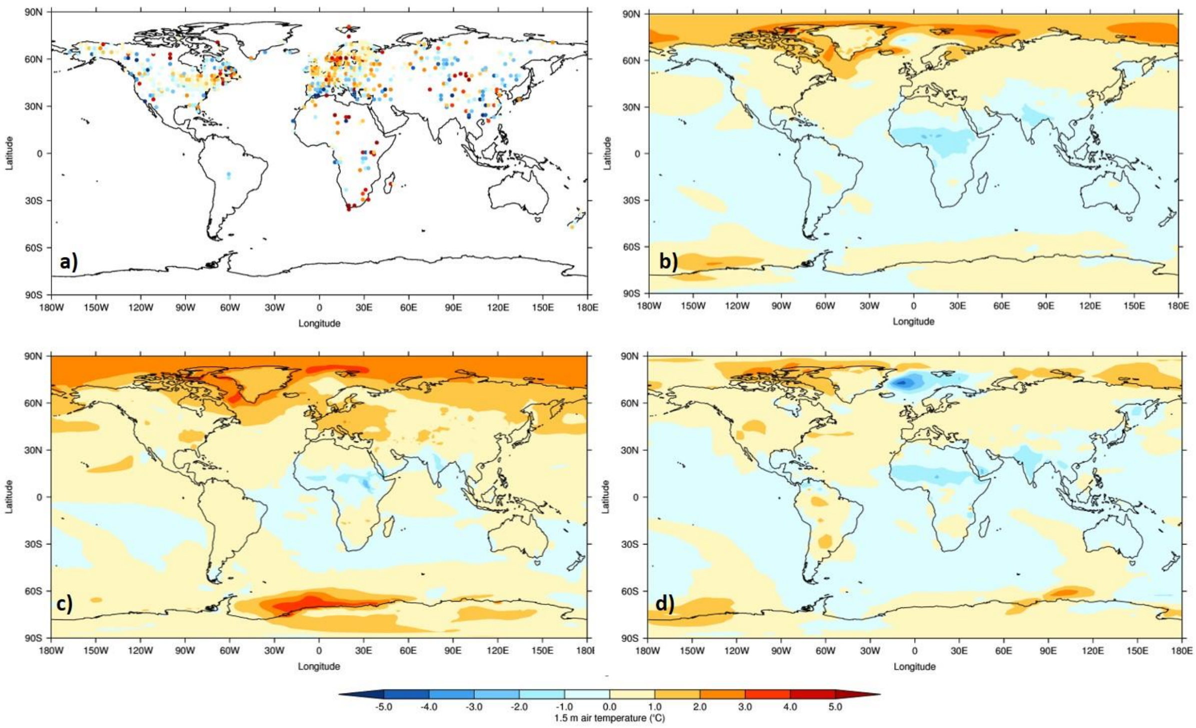

Figure 7 - Calendar adjusted mean annual surface air temperature anomalies from simulated model data versus proxy data. Background data show simulated anomalies (MH – PI) from different generations of the same model: a) Proxy data anomalies (MH – PI) from Bartlein et al. (2011), with locations projected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

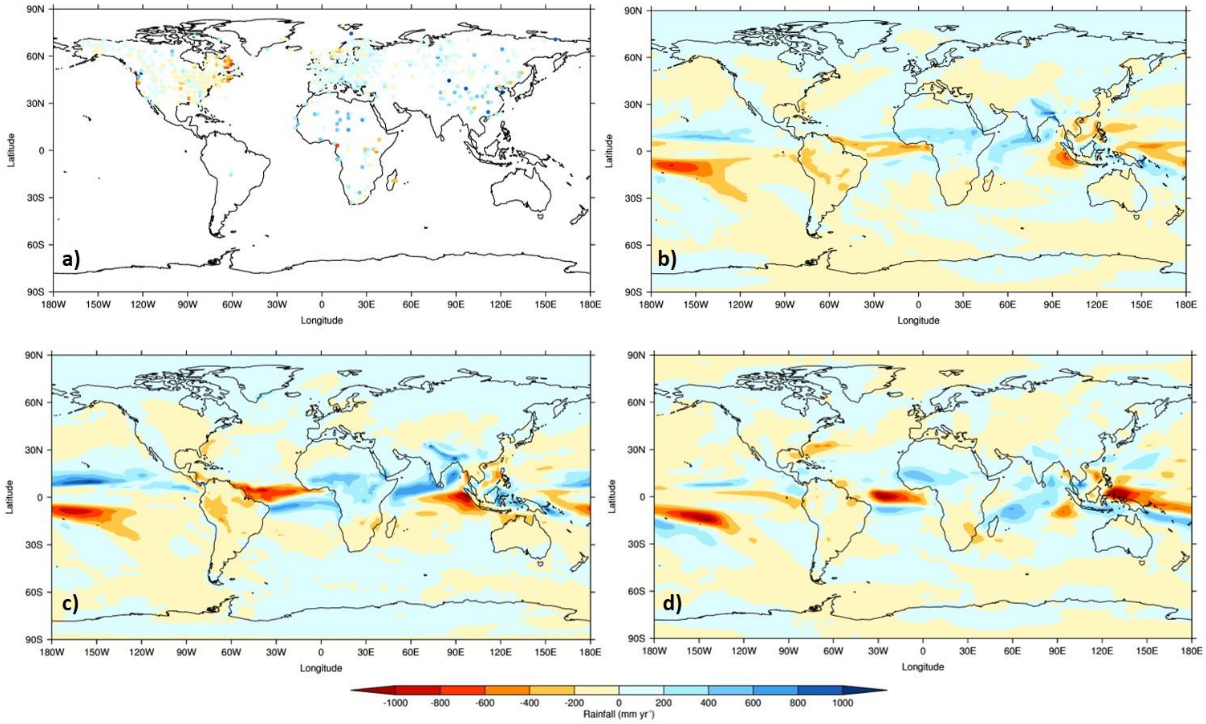

Figure 8 - Same as Figure 7, but for rainfall anomalies

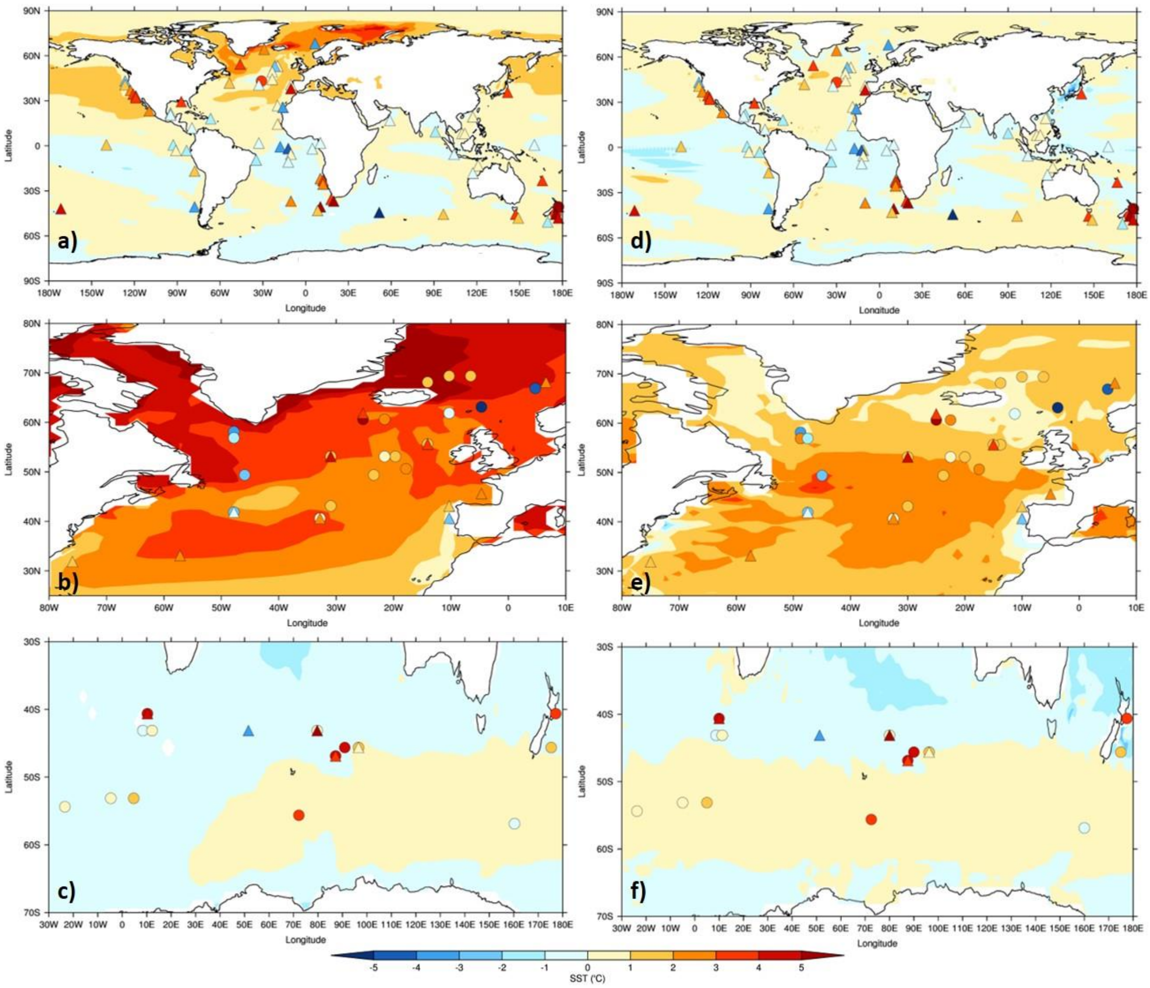

Figure 9 - Calendar adjusted SST anomalies from model simulated data versus proxy data. Background data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Capron et al. (2017) and triangles show anomalies 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)

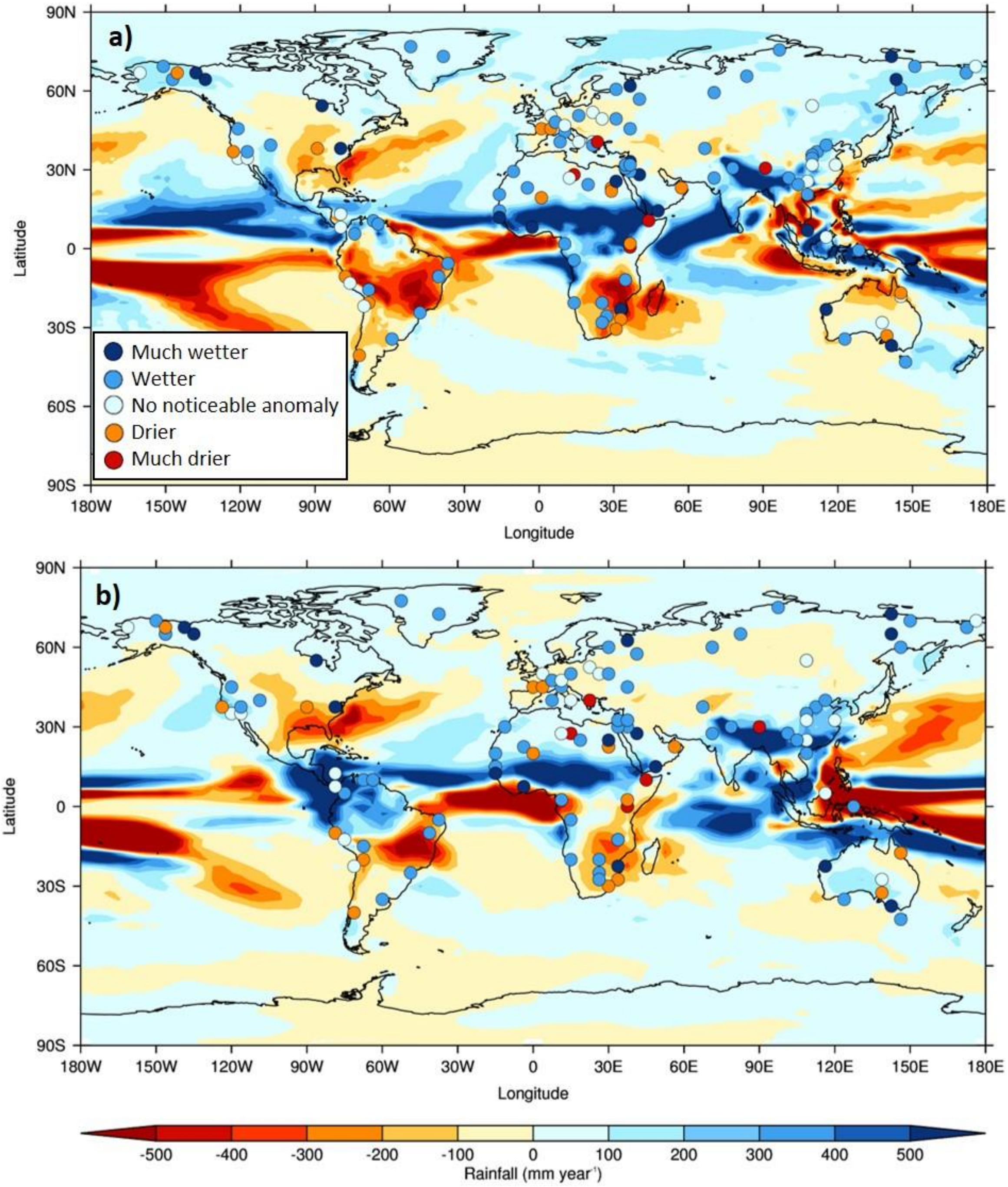

Figure 10 - Calendar adjusted annual surface rainfall anomalies from model simulated data versus proxy data. Background data show simulated anomalies (LIG - PI climatology) from different generations of the same model, circles show proxy data anomalies (LIG – preindustrial) from Scussolini et al. (2019). 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)