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

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

[1]Department of Geography, University College London, London, WC1E 6BT, UK
[2]Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AB, UK
[3]Laboratoire des Sciences du Climat et de l'Environnement-IPSL, Unité Mixte CEA-CNRS-UVSQ, Université Paris-Saclay, Orme des Merisiers, Gif-sur-Yvette, France
[4]Department of Meteorology, University of Reading, Reading, RG6 6BB, UK
[5]School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK
[6]Alfred-Wegener-Institut Helmholtz-Zentrum f ur Polar- und Meeresforschung, Bremerhaven, Germany
[7]Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[8]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA.
[9]Blue Skies Research Ltd, The Old Chapel, Albert Hill, Settle, BD24 9HE, UK
[10]Department of Earth Sciences, University of Southern California, Los Angeles, California, USA
[11]Max Planck Institute for Meteorology, Hamburg, Germany
[12]Department of Physics, University of Toronto, 60 St George Street, Toronto, Ontario, M5S1A7, Canada
[13]LOCEAN Laboratory, Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Paris, France
[14]CIDIS-LID-Facultad de Ciencias y Filosofía-Universidad Peruana Cayetano Heredia, Lima, Peru
[15]Department of Geography, University of Oregon, Eugene, OR 97403, USA
[16]LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[17]NORCE Norwegian Research Centre, Bjerknes Center for Climate Research, Bergen, Norway
[18]Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, 10691, Stockholm, Sweden
[19]Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119333, Russia
[20]Climate and Global Dynamics Laboratory, National Center for Atmospheric Research (NCAR), Boulder, CO 80305, USA
[21]Institute of Geography, Russian Academy of Sciences, Staromonetny L. 29, Moscow, 110917, Russia
[22]NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA
[23]School of Atmospheric Sciences, Nanjing University of Information Science & Technology Nanjing, 210044, China
[24]Atmospheric and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan

Correspondence: c.brierley@ucl.ac.uk

Abstract. The mid-Holocene (6,000 years ago) is a standard experiment for the evaluation of the simulated response of global climate models using paleoclimate reconstructions. The latest mid-Holocene simulations are a contribution by the Palaeoclimate Model Intercomparison Project (PMIP4) to the current phase of the Coupled Model Intercomparison Project (CMIP6).

Here we provide an initial analysis and evaluation of the results of the experiment for the mid-Holocene. We show that state-of-
5   the-art models produce climate changes that are broadly consistent with theory and observations, including increased summer
warming of the northern hemisphere and associated shifts in tropical rainfall. Many features of the PMIP4-CMIP6 simulations
were present in the previous generation (PMIP3-CMIP5) of simulations. The PMIP4-CMIP6 ensemble for the mid-Holocene
has a global mean temperature change of -0.3 K, which is -0.2 K cooler that the PMIP3-CMIP5 simulations predominantly
as a result of the prescription of realistic greenhouse gas concentrations in PMIP4-CMIP6. Neither this difference nor the im-
10   provement in model complexity and resolution seems to improve the realism of the simulations. Biases in the magnitude and
the sign of regional responses identified in PMIP3-CMIP5, such as the amplification of the northern African monsoon, precip-
itation changes over Europe and simulated aridity in mid-Eurasia, are still present in the PMIP4-CMIP6 simulations. Despite
these issues, PMIP4-CMIP6 and the mid-Holocene provide an opportunity both for quantitative evaluation and derivation of
emergent constraints on climate sensitivity and feedback strength.

1   Introduction

Future climate changes pose a major challenge for Human civilisation, yet uncertainty remains about the nature of those
changes. This arises from societal decisions about future emissions, and also uncertainty stemming from differences between
the models used to make the projections (Hawkins and Sutton, 2011; Collins et al., 2013). Coupled general circulation models
(GCMs) can be used to simulate past changes in climate as well as those of the future. Palaeoclimate simulations allow us to
20   test the theoretical response of such models and provide an independent evaluation of them. The Coupled Model Intercompari-
son Project (CMIP; Eyring et al., 2016), which coordinates efforts to compare climate model simulations, includes simulations
designed to test model performance under past climate regimes. Evaluation of these palaeoclimate simulations against palaeo-
climate reconstructions, coordinated through the Palaeoclimate Modelling Intercomparion Project (PMIP; Kageyama et al.,
2018), provides an independent test of the ability of state-of-the-art models to simulate climate change.

25   The mid-Holocene (6000 years ago, 6ka) is one of the palaeoclimate simulations included in the current phase of CMIP
(PMIP4-CMIP6; Otto-Bliesner et al., 2017). This period is characterised by an altered seasonal and latitudinal distribution of
incoming solar radiation, because of larger obliquity and orbital precession, meaning that the Earth was closest to the Sun in
austral spring (rather than in austral summer as today) and that the northern latitudes received more solar radiation than today.
The mid-Holocene has been a baseline experiment for PMIP since its inception (Joussaume et al., 1999; Braconnot et al., 2007,
30   2012). As such, it has been a focus for synthesis of palaeoenvironmental data (see summary in Harrison et al., 2016) and for the
reconstruction of palaeoclimate variables from these data (e.g. Kohfeld and Harrison, 2000; Bartlein et al., 2011), facilitating
systematic model evaluation (e.g. Hargreaves et al., 2013; Jiang et al., 2013; Prado et al., 2013; Harrison et al., 2014; Mauri
et al., 2014; Perez-Sanz et al., 2014; Harrison et al., 2015; Bartlein et al., 2017).

The PMIP4-CMIP6 simulations differ from previous palaeoclimate simulations in two ways. Firstly, they represent a new
35   generation of climate models with greater complexity, improved parameterisations and often run at higher resolution. Changes
to the model configuration have, in some cases (e.g. CCSM4/CESM2, HadGEM2/HadGEM3, IPSL-CM5A/IPSL-CM6A),

resulted in substantially higher climate sensitivity than the previous PMIP3-CMIP5 version of the same model, although this is not a feature of all of the models (Tab. 1,2). Secondly, the protocol for the PMIP4-CMIP6 mid-Holocene experiment (called *midHolocene* on the Earth System Grid Federation, and henceforth herein) simulations was designed to better represent the

40 observed conditions than previous simulations (Otto-Bliesner et al., 2017). In addition to the change in orbital configuration, which was the only change imposed in the PMIP3-CMIP5 experiments, the current experiments include a realistic specification of changes in atmospheric greenhouse gas concentrations. Because of these small changes, the PMIP4-CMIP6 simulations are expected to be slightly colder that in the previous PMIP phase (Otto-Bliesner et al., 2017). The model configuration and all other forcings are the same as in the pre-industrial control simulation (*piControl*, 1850 CE). This means that models with

45 dynamic vegetation in the *piControl* are run with dynamic vegetation in the *midHolocene* experiment, so the PMIP4-CMIP6 ensemble includes a mixture of simulations with prescribed or interactive vegetation. Although some of the PMIP3-CMIP5 models were run with an interactive carbon cycle, none included fully-dynamic vegetation.

Here, we provide a preliminary analysis of the PMIP4-CMIP6 *midHolocene* simulations, focusing on surface temperature changes (sec. 3.1), hydrological changes (sec. 3.2 & 3.3) and the deep ocean circulation (sec. 3.4). We examine the impact

50 of changes in model configuration and experiment protocol on these simulations, specifically how far these changes improve known biases in the simulated changes. We draw on a more extended set of observation-derived benchmarks to evaluate these simulations. Finally we discuss the implications of this evaluation for future climate changes, for example by investigating whether the different climate sensitivities between CMIP6 and CMIP5 generation models has an impact on the new results.

2   Methods

55 ### 2.1   Experimental Setup and Models

The protocol and experimental design for the PMIP4-CMIP6 *midHolocene* simulations are described by Otto-Bliesner et al. (2017). The *midHolocene* simulations are run with known orbital parameters for 6000 yr BP and atmospheric trace greenhouse gas concentrations (GHGs) derived from ice-core records (as described by Otto-Bliesner et al., 2017). Eccentricity is increased by 0.001918 in the *midHolocene* simulations relative to the *piControl*, obliquity is increased by 0.646°, and perihe-

60 lion is changed from 100.33° in the *piControl* to 0.87° in the *midHolocene* (near the boreal autumn equinox). The result of these astronomical changes is a difference in the seasonal and latitudinal distribution of top-of-atmosphere (TOA) insolation. During boreal summer, anomalies between 40-50°N are 25 W/m^2 higher in the *midHolocene* simulations than in the *piControl* (Otto-Bliesner et al., 2017). The long-lived greenhouse gases are specified at their observed concentrations. Carbon dioxide is specified at 264.4 ppm (vs 284.3 ppm during the pre-industrial) and methane at 597 ppb (vs 808 ppb) and N_2O at 262 ppb

65 (versus 273 ppb). These changes in GHG concentrations lead to an effective radiative forcing of -0.3 W/m^2 (Otto-Bliesner et al., 2017).

Twelve models (Tab. 1) have performed the PMIP4-CMIP6 *midHolocene* simulations. A similar number of models have performed the equivalent PMIP3-CMIP5 *midHolocene* simulation (Tab. 2). The PMIP4-CMIP6 simulations are either available from the Earth System Grid Federation (from which they are freely downloadable) or will be lodged there in the near future. We

70   evaluate these simulations as part of an ensemble and do not always identify individual models. Most of the models included in the PMIP4-CMIP6 ensemble are state-of-the-art climate models, but we also include some results from models that are either lower resolution or less complex (and therefore faster). Even though all models have the same orbital parameters and trace gases in the *midholocene* experiment, there may be other differences compared to the *piControl* which can mean that the forcing is not exactly the same. For example, the different models have different solar constants (see Table 1), reflecting

75   choices made by the different groups for the *piControl* simulations. Similarly, the orbital parameters used by some groups for the *piControl* are the same as for the historical simulation and the trace gases are slightly different from the PMIP4-CMIP6 recommendations. Differences in the pre-industrial planetary albedo, resulting from surface albedo and clouds, may also mean the effective solar forcing is different between models (Braconnot et al., 2012). Experimental setup and spin-up procedure shall be documented for each *midHolocene* simulation individually elsewhere (Otto-Bliesner et al., 2017).

80   ## 2.2   Calendar adjustments and analysis techniques

Model outputs for the *midHolocene* simulation are currently mostly available as monthly values. The altered orbital configuration in the mid-Holocene resulted in a change in the Earth's transit speed through the seasonal cycle such that, considered as angular fractions of the Earth's orbit, the month lengths differed during the mid-Holocene (Joussaume and Braconnot, 1997; Bartlein and Shafer, 2019). Northern Hemisphere winter (December, January, February, DJF) was longer and summer (June,

85   July, August, JJA) correspondingly shorter than in the present day and the *piControl* simulation. To take account of these differences in calculating monthly or seasonal variables, we use the PaleoCalAdjust software (Bartlein and Shafer, 2019), which interpolates from non-adjusted monthly averages to daily values and calculates the average values for adjusted months defined as angular fractions of the orbit. However, since calendar adjustment is not necessary for annual measures of climate, we use the original outputs to calculate annual variables (as the automated scripts Phillips et al. (2014) weight each month evenly). The

90   PaleoCalAdjust software computes adjusted monthly variables from original monthly means which could impact the accuracy of variables that change abruptly throughout the year, rather than gradually, such as the sudden increase in precipitation in monsoon regions. To explore whether potential interpolation errors from PaleoCalAdjust are justified in such situations, we analysed the averaged rain rate during the monsoon season over the South American monsoon domain in the IPSL-CM6A-LR *midHolocene*, for which daily-resolution data is also provided on the Earth System Grid Federation. The areal extent of South

95   American monsoon domain varies when using different temporal data, so comparisons were only made over the common grid points. The average monsoon rain rate from the daily-resolution data is 7.0 mm/day: compared to 6.7 mm/day from calendar-adjusted monthly data and 7.1 mm/day using monthly data without this adjustment. The average monsoon rain rate in the *piControl* is 7.5 mm/day. We have therefore not applied the calendar adjustment when analysing monsoon variables.

Although fixed monsoon domains are often used when investigating variability and future changes in monsoon rainfall (e.g.

100   Christensen et al., 2013), this is not appropriate in the mid-Holocene when the monsoons were greatly extended. Following Jiang et al. (2015), we adopt the definition of Wang et al. (2011) for analysis of monsoon regions: a grid point is considered to be affected by the monsoon if the rainfall predominantly falls in the summer (MJJAS in the Northern Hemisphere, NDJFM in the Southern Hemisphere; assessed using summer rainfall forming at least 55% of the annual total) and the average rain rate

difference between summer and winter (called monsoon intensity) is at least 2 mm/day or more. The ensemble mean global

105 domain is determined by applying both these criteria to the ensemble mean summer rainfall and monsoon intensity. We also calculate the areal extent of 7 land-based monsoon systems annually (November-October), as well as determining the average precipitation rate within each domain. Interannual variability is characterised by the standard deviation of these two quantities. The integral of these values is the total monsoon rainfall.

The analysis presented here mainly uses generalised evaluation software tools derived from the Climate Variability Diagnos-

110 tics Package (Phillips et al., 2014), which has been adapted for palaeoclimate purposes (Brierley and Wainer, 2018). It uses the surface air temperature and precipitation rate variables ('tas' and 'pr' respectively in the ESGF controlled vocabulary; Juckes et al., 2019), as well as several different ocean overturning mass streamfunction variables. The software and routines used to create the figures presented here are available to download (see *code and data availability* statement).

2.3 Palaeoclimate reconstructions and model evaluation

115 We provide only a preliminary quantitative evaluation of the realism of the PMIP4-CMIP6 simulations, drawing attention to obvious similarities and mismatches between the simulations and observational evidence of past climates. Some of this evidence is qualitative (e.g. changes in surface hydrology evidenced by lakes and vegetation records; Kohfeld and Harrison, 2000; Prentice et al., 2000), but we also use quantitative reconstructions from a number of sources. Bartlein et al. (2011) provide pollen-based reconstructions of land climate, including mean annual temperature, mean temperature of the coldest month,

120 growing season temperature (indexed by growing degree days above a baseline of 0), mean annual precipitation and an index of soil moisture (alpha, the ratio of actual to potential evaporation). They combined the reconstructions at individual pollen sites to produce an estimate for a 2°x2°grid, a resolution comparable with the climate models; reconstruction uncertainties are estimated as a pooled estimate of the standard errors of the original reconstructions for all sites in each grid cell (Bartlein et al., 2011). This data set was used to evaluate the PMIP3-CMIP5 simulations (Harrison et al., 2014) and has good coverage

125 of northern hemisphere terrestrial sites, although there are gaps in the coverage especially for the tropics and southern hemisphere. We also use temperature reconstructions from the 'Temperature 12k' database (Kaufman et al., in press). We extracted anomalies for the mid-Holocene compared to the last millennium interval (6.0 ± 0.5ka $- 0.6 \pm 0.5$ka) for site-level comparison with the PMIP4-CMIP6 simulations. This database has 1332 time series reconstructions of temperature (mean annual, summer and winter temperature) based on a variety of different ecological, geochemical and biophysical recrine (212) and terrestrial

130 (472) archives (Kaufman et al., in press). Additionally area-averaged temperature anomalies (w.r.t. 1800-1900) over 30° latitudinal bands have been generated using five different methods (Kaufman et al., submitted) to yield a single composite value with confidence intervals. Differences in methodology and coverage preclude direct comparison between the Bartlein et al. (2011) and Kaufman et al. (in press) data sets. We use both data sets to provide a measure of the uncertainties in reconstructed climates.

3 Simulated mid-Holocene Climates

3.1 Temperature Response

As expected from the insolation forcing, the PMIP4-CMIP6 ensemble shows an increase in mean annual temperature (MAT) in the high northern and southern latitudes and over Europe (Fig. 1). Yet there is a decrease in MAT elsewhere, which is especially large over northern Africa and India. The ensemble produces a global cooling of -0.3°C compared to the *piControl* simulation (Tab. S1). The relatively small change in MAT is consistent with the fact that the *midHolocene* changes are largely driven by seasonal changes in insolation, yet are of a different sign than the +0.5°C reconstruction (Kaufman et al., submitted) derived from the Temperature 12k compilation (Kaufman et al., in press). As might be expected, the higher insolation in northern hemisphere (NH) summer results in a pronounced summer (JJA) warming, particularly over land (Fig. 2). The increase in summer temperature over land in the NH high latitudes in the ensemble mean is 1.1°C (Tab. S1). Increased NH summer insolation leads to a northward shift and intensification of the monsoons (sec. 3.2), with an acompanying JJA cooling in the monsoon-affected regions of northern Africa and and South Asia. Reduced insolation in the NH winter (DJF) results in cooling over the northern continents and this cooling extends into the northern tropical regions, although the Arctic is warmer than in the *piControl* simulation (Fig. 2). Although the Southern Ocean shows warmer temperatures in the *midHolocene* than the *piControl* simulations in austral summer (DJF) as a result of increased obliquity, this warming does not persist into the winter to the same extent as seen in the Arctic. The damped insolation seasonality, together with the large effective heat capacity of the ocean heavily damps seasonal variations in surface air temperature. The enhanced NH seasonality and the preponderance of land in the NH therefore results in large seasonal variations of the interhemispheric temperature gradient, which translate into a small increase in favour of the northern hemisphere in the annual, ensemble mean.

The geographic and seasonal patterns of temperature changes in the PMIP4-CMIP6 ensemble are very similar to those seen in the PMIP3-CMIP5 ensemble. However, the change in MAT with respect to the *piControl* in the PMIP4-CMIP6 ensemble is less than in the PMIP3-CMIP5 (Fig. 1). The PMIP4-CMIP6 ensemble is cooler than the PMIP3-CMIP5 ensemble in both summer and winter (Fig. 2). The difference in the experimental protocol between the two sets of simulations would be expected to cause a slight cooling, since the difference in GHG concentrations would result in an effective radiative forcing of -0.3 W/m^{-2} (Otto-Bliesner et al., 2017). To evaluate this, we estimate the ensemble-mean forced response (Fig. 1) based on the climate sensitivity of each model (Tab. 1) and pattern scaling (Brierley et al., 2019). The estimated global mean pattern-scaled anomaly is -0.28°C, similar to the difference between the two model generations (Fig. 1).

Biases in the control simulation may influence the response to mid-Holocene forcing (Braconnot et al., 2012; Ohgaito and Abe-Ouchi, 2009; Harrison et al., 2014; Braconnot and Kageyama, 2015) and certainly affect the pattern and magnitude of simulated changes. There is some difficulty in diagnosing biases in the *piControl*, because there are few spatially-explicit observations for the pre-industrial, especially for precipitation. We therefore evaluate these simulations using reanalysed climatological temperatures (between 1871-1900 CE; Compo et al., 2011) for the spatial pattern (Fig. 3) and zonal averages of observed temperature (Fig. 4) for the period 1850-1900 CE from the HadCRUT4 dataset (Morice et al., 2012; Ilyas et al., 2017). We compare these with the mean difference between the pre-industrial climatology of each model (i.e. the ensemble

mean bias). The PMIP4-CMIP6 models are generally cooler than the observations, most noticeably in polar regions, over land

170 and over the NH oceans (Fig. 4). The models are still too warm over the eastern boundary upwelling currents, even though this bias has been reduced in some of the models compared to PMIP3-CMIP5. The colder conditions over the Labrador Current also indicate a difficulty with resolving the regional ocean circulation features sufficiently. The polar regions are noticeably too cold (Fig. 3 & 4), though the match between the models and the temperature observations/reanalysis appears satisfactory in the tropics. The magnitude of the simulated mid-Holocene temperature response in the Arctic is not significantly correlated

175 with the bias in the *piControl* simulation ($r = -0.28$, Fig. 4). Other factors such as ice albedo and ocean temperature advection affect the direct and indirect response to mid-Holocene forcing in these regions. PMIP4-CMIP6 also includes simulations with dynamic vegetation, for example. The associated vegetation-snow albedo feedback would tend to reduce the simulated cooling (e.g. O'ishi and Abe-Ouchi, 2011), but can introduce a larger cooling bias in the *piControl* simulation (Braconnot et al., 2019). However, changes in the treatment of aerosols in the PMIP4-CMIP6 ensemble could enhance the simulated cooling (Pausata

180 et al., 2016; Hopcroft and Valdes, 2019).

The reconstructed zonal temperature changes during the mid-Holocene suggest a warming at all latitudes, with maximum warming in the Arctic (Fig. 4). This feature is robust between the Bartlein et al. (2011) and Kaufman et al. (in press) reconstructions. The PMIP4-CMIP6 ensemble is equivocal about whether the polar regions were warmer or cooler on the annual mean. Furthermore, the PMIP4-CMIP6 models show a consistent cooling in the tropics. Tropical cooling was present but less

185 pronounced in the PMIP3-CMIP5 ensemble (Fig. 4). Tropical cooling is not consistent with the Temperature 12k area-averages (Kaufman et al., in press). Further work is required to determine whether the discrepancies between the temperature reconstructions and PMIP4-CMIP6 simulations indicate model deficiencies, or tell a more nuanced story (e.g. Liu et al., 2014b; Marsicek et al., 2018).

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

190 changes. The standard deviation of the temperature response across the PMIP4-CMIP6 ensemble is of the same magnitude as the ensemble mean for both annual (Fig. 1) and seasonal (Fig. 2) temperature changes. There is a very large spread in the high-latitude oceans and adjacent land areas in the winter hemisphere, where the spread originates from inter-model differences in the extent of the simulated sea ice. Ice-albedo feedback would enhance inter-model temperature differences (Berger et al., 2013). The second region characterised by large inter-model differences is where there are large changes in precipitation in the

195 tropics. This suggests that the spread originates in inter-model differences in simulated large scale water advection, evaporative cooling, cloud cover and precipitation changes. There is no systematic reduction in the spread of temperature responses within PMIP4-CMIP6 ensemble compared to the PMIP3-CMIP5 ensemble (Fig. 1, Fig. 2). Each of the ensembles include models of different complexity, and the lack of a systematic difference suggest that complexity and model tuning has a larger impact on the responses than differences in the protocol. Thus, even with the protocol-forced cooling of PMIP4-CMIP6 relative to

200 PMIP3-CMIP5, it may still be possible to consider them both as subsets of the same combined ensemble (Harrison et al., 2014). However, new approaches to classify models to highlight the impact of model complexity or of model biases on the response, following a fit-for-purpose approach, are clearly needed.

3.2 Monsoonal Response

The enhancement of the global monsoon is the most important consequence of the mid-Holocene changes in seasonal insolation

205 for the hydrological cycle (Jiang et al., 2015). The global monsoon domain is expanded in the PMIP4-CMIP6 *midHolocene* simulations: this occurs because of changes in both the summer rain rate and the monsoon intensity (Fig. 5). The weakening of the annual range of precipitation over the ocean and the strengthening over the continents indicates the changes reflect a redistribution of moisture (see e.g. Braconnot, 2004).

The most pronounced and robust changes in the monsoon occur over northern Africa and the Indian subcontinent (Fig. 6).

210 The areal extent of the northern African monsoon is 20-50% larger than in the *piControl* simulations, but the average rain rate only increases by 10% (Fig. 7). The intensification of precipitation on the southern flank of the Himalayas (Tab. S1) in the *midHolocene* simulations is offset by a reduction in the Philippines and Southeast Asia (Fig. 6), so the area-averaged reduction in rain rate is reduced over the South Asian monsoon domain (Fig. 7). There is an extension and intensification of the East Asian monsoon that is consistent across the PMIP4-CMIP6 ensemble, but the change is <10% (Fig. 7). This is a region

215 where previous analyses have shown that simulated changes in monsoon rainfall reflect the competition between enhanced contrast in moist static energy between land and ocean and increased local evaporation over the warmer oceans (Ohgaito et al., 2013). Ensemble mean changes in the North American Monsoon System, and the Southern Hemisphere monsoons are also small (Fig. 6), and less consistent across the ensemble although most of the models show a weakening and contraction of the Southern American Monsoon System and Southern African monsoon (Fig. 7). Changes in interannual variability within the

220 monsoon systems (characterised by standard deviations in both the areal extent and area-averaged rain rate; Fig. 7) are not consistent across the PMIP4-CMIP6 ensemble. Furthermore, those models that have the largest change in variability in one region are not necessarily the models that have large changes in other regions, which suggests that this variability is linked with regional feedbacks, rather than being an inherent characteristic of a model.

The broad scale changes in the PMIP4-CMIP6 simulations, with weaker southern and stronger and wider northern hemi-

225 sphere monsoons, were present in the PMIP3-CMIP5 simulations (Fig. 6). The response is robust across model results, indicating that all models produce the same large scale redistribution of moisture by the atmospheric circulation in response to the interhemispheric and land-sea gradients induced by the insolation and trace gas forcing. At a regional scale, however, there are differences between the two ensembles. The PMIP4-CMIP6 *midHolocene* ensemble shows wetter conditions over the Indian Ocean, a larger northward shift of the ITCZ in the Atlantic and a widening of the Pacific rain belt compared to the PMIP3-

230 CMIP5 models (Fig. 6). The expansion of the summer (JJA) monsoon in northern Africa is also greater in the PMIP4-CMIP6 than PMIP3-CMIP5 ensemble (Tab. S1) and the location of the northern boundary is more consistent between models. This is associated with a better representation of the northern edge of the rainbelt for the *piControl* simulation in the PMIP4-CMIP6 ensemble compared with previous generations (Fig. S1). However, there is no relationship between the amount of precipitation in the *piControl* simulations and the change in precipitation. The changes in precipitation appear to be more related to local

235 dynamics rather than orbitally-induced insolation changes (D'Agostino et al., 2019); some of the changes may be related to the inclusion of new land surface models, or dynamic vegetation in some PMIP4-CMIP6 models.

Although the PMIP4-CMIP6 models show the expected expansion of the monsoons, this expansion is weaker than indicated by palaeoclimate reconstructions (Fig. 8 & S1). This was a feature of the PMIP3-CMIP5 simulations (Braconnot et al., 2012; Perez-Sanz et al., 2014) and indeed previous generations of climate models (Joussaume et al., 1999; Braconnot et al.,

240 2007). It has been suggested that this persistent mismatch between simulations and reconstructions arises from biases in the *piControl* (Harrison et al., 2015). Indeed, the ensemble mean global monsoon domain in the PMIP4-CMIP6 ensemble is more equatorward in the *piControl* compared to the observations, particularly over the ocean (Fig. 5). In northern Africa, the expansion of the monsoon domain in the *midHolocene* simulations merely removes the underestimation of its poleward extent in the *piControl* simulations (Fig. 5). Furthermore, evaluation of the *piControl* simulations using climatological precipitation

245 data for the period between 1970 and the present day (Adler et al., 2003) shows the models fail to capture the magnitude of rainfall in the Intertropical Convergence Zone (ITCZ) and simulate a South Pacific Convergence Zone (SPCZ). The SPCZ is too zonal because of the poor representation of the SST gradient between the equator and 10°S in the west Pacific (Fig. 3; Brown et al., 2013). The PMIP4-CMIP6 models exhibit a dry bias over tropical and high northern latitude land areas, although the mid-latitude storm tracks are captured with varying levels of fidelity (Fig. 3).

250 There is large spread in the mid-Holocene precipitation response across both the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles (Fig. 6 & 8). Unsurprisingly, the ensembles exhibits the largest spread in its simulated mid-Holocene response where that response has the highest magnitude (Fig. 6).

3.3 Extratropical hydrological responses

Hydrological changes in the extratropics are comparatively muted in the PMIP4-CMIP6 ensemble, and closely resemble fea-

255 tures seen in the PMIP3-CMIP5 ensemble. There is a reduction in rainfall at the equatorward edge of the mid-latitude storm tracks, most noticeable over the ocean (Fig. 6). The NH extratropics are generally drier in the *midHolocene* simulations than in the *piControl*. There is a large inter-model spread in the summer rainfall changes over eastern North America and central Europe (Fig. 8). The spread in summer rainfall in both regions is clearly linked to the large inter-model spread in summer temperature (Fig. 6). Reconstructions from eastern North America suggest slightly drier conditions while reconstructions for

260 central Europe show somewhat wetter conditions, but in neither case are these incompatible with the simulations.

There are regions, however, where there is a substantial mismatch between the PMIP4-CMIP6 simulations and the pollen-based reconstructions. There is a simulated reduction in summer rainfall in mid-continental Eurasia (Fig. 6). This reduction is somewhat larger in the PMIP4-CMIP6 ensemble than in the PMIP3-CMIP5 ensemble, although this difference is likely not significant (Fig. 8). However, this reduction in precipitation and the consequent increase in mid-continental temperatures is

265 inconsistent with palaeoenvironmental evidence (and climate reconstructions), which show that this region was characterised by wetter and cooler conditions than today in the mid-Holocene (Fig. 8; Bartlein et al., 2017, Tab. S1). This indicates that model improvements have not resolved this persistent mismatch between simulated and observed mid-Holocene climate. Bartlein et al. (2017) pinpointed poor simulation of the extratropical atmospheric circulation as the underlying cause of this mismatch. The higher resolution of most PMIP4-CMIP6 models does not seem to improve the representation of the circulation. Poor

270 simulation of the extratropical circulation could also explain the failure to capture precipitation changes over Europe accurately

(Mauri et al., 2014). The PMIP4-CMIP6 ensemble shows little change in mean annual precipitation over Europe (Fig. 6) and fails to capture the north-south gradient of changes in mid-Holocene precipitation shown by reconstructions: with much wetter conditions in the Mediterranean, compared to modest increases in northern Europe (Fig. 8).

3.4 Ocean Circulation

275 The AMOC is an important factor affecting the Northern Hemisphere climate system and is a major source of decadal and multidecadal climate variability (e.g. Rahmstorf, 2002; Lynch-Stieglitz, 2017; Jackson et al., 2015). Recent studies have reported a ~15% decline in AMOC strength from the pre-industrial period to the present day (Rahmstorf et al., 2015; Dima and Lohmann, 2010; Caesar et al., 2018; Thornalley et al., 2018), at least partly in response to anthropogenic forcing. Reproducing the AMOC of the mid-Holocene is important for understanding the climate responses to external forcing at millennial

280 timescales. The members of both the PMIP4-CMIP6 and PMIP3-CMIP5 ensemble have different AMOC strengths in their *piControl* simulations (Fig. 9), although all models correctly predict that it is stronger at 30°N than at 50°N. There is a strong correlation (r=0.99 at 30°N) between the simulated strength of the AMOC in the *midHolocene* and the *piControl*. Furthermore, there is little change in the overall strength of the AMOC between the *midHolocene* and *piControl* experiments (Fig. 9) in either the PMIP4-CMIP6 or the PMIP3-CMIP5 simulations, and no consistency in whether this comparatively small (and probably

285 non-significant) change is positive or negative. The small difference between the *midHolocene* and *piControl* states is suprising given the magnitude of low frequency internal variability in AMOC. Shi and Lohmann (2016) detect large differences in simulated AMOC anomalies between models with coarse and higher resolutions. They suggest ocean and atmospheric processes affecting ocean salinity close to the sites of deep convection mean that higher resolution models tending to produce stronger *midHolocene* AMOC and lower resolution simulations a weaker AMOC than the *piControl*. The comparatively small changes

290 in the AMOC strength between the PMIP4-CMIP6 *piControl* and *midHolocene* simulations are consistent with these earlier results, where the simulated changes are generally of less than 2 Sv (Fig. 9).

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

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

300 consistent with the simulated results (Fig. 9).

3.5 Evaluation of mid-Holocene climate features

Comparisons of the PMIP4-CMIP6 simulations with either palaeoenvironmental observations or palaeoclimate reconstructions have highlighted a number of regions where there are mismatches either in magnitude or sign of the simulated response.

Whilst it is possible to attempt assessment of overall performance of each model (e.g. Taylor, 2001, Fig. S2) or for individual
305   regional features (e.g. Fig. S3), their utility is unclear. Substantial further research is required before the quality of *midHolocene*
simulations can be used operationally to enhance future projections for climate services - although Schmidt et al. (2014a)
provide most of the necessary groundwork.

Analyses of key features of the *midHolocene* simulations, such as the monsoon amplification or the strength of the AMOC,
suggest that the PMIP4-CMIP6 simulations are from the same population as the PMIP3-CMIP5 simulations. We formally test
310   this by calculating Hotelling's T^2 statistic (Wilks, 2011), a multivariate generalization of the ordinary t-statistic that is often
used to examine differences in climate-model simulations (Chervin and Schenider, 1976), at each grid point of a common
1°grid for different combinations of climate variables. The patterns of "significant" (i.e. $p < 0.05$) tests (where one would
reject the null hypothesis that the PMIP4-CMIP6 and PMIP3-CMIP6 ensemble means are equal between groups) are quite
chaotic (Fig. 10) and show little relation to the largest climate anomalies (Fig. 1 & 6). There are few locations that do not fall
315   below the false discovery rate (Wilks, 2006). Consequently there is little support for the idea that the PMIP4-CMIP6 generation
of simulations differ from the PMIP3-CMIP5 simulations, which were themselves not significantly different from the PMIP3-
CMIP5 simulations (Harrison et al., 2015). This suggests, that all of these simulations could be considered as a single ensemble
for process-based analysis (e.g. D'Agostino et al., 2019) or for the investigation of emergent constraints (Yoshimori and Suzuki,
2019), which would considerably enhance the statistical power of such analyses.

320   Many of the PMIP4-CMIP6 models have a higher climate sensitivity, defined as the response of global temperature to a
doubling of CO_2 (Gregory et al., 2004), than earlier versions of the same model (Tab. 1, Tab. 2). Although increased sensitivity
could contribute to the PMIP4-CMIP6 ensemble being somewhat cooler than the PMIP3-CMIP5 ensemble, the change in the
experimental protol appears to be the dominant explanation for this change (Fig. 1). There is no inherent relationship between
climate sensitivity and seasonality, because of differences in the rôle of the ocean on seasonal compared to multi-annual
325   timescales. Nevertheless, since the change in climate sensitivity arises from differences in basic climate feedbacks, such as
water vapour or clouds, it is feasible that the change in climate sensitivity could affect the simulated changes in seasonality.
However, we find no inherent relationship between climate sensitivity and temperature seasonality, here shown for seasonality
changes in central Asia (Fig. 11). Although four of the individual models that have higher sensitivity in PMIP4-CMIP6 than
the corresponding version of that model in PMIP3-CMIP5 show an increase in the seasonality (Fig. 11), others show a decrease
330   in seasonality with increased sensitivity. Even if there is no ubiquitous relationship, the fact that changes in climate sensitivity
can be detected in the thermodynamic response to orbital forcing raises the possibility that the *midHolocene* simulations could
provide a constraint on climate sensitivity.

Circum-Pacific paleoclimate records document marked fluctuations in ENSO activity throughout the Holocene (Tudhope
et al., 2001; McGregor and Gagan, 2004; Koutavas and Joanides, 2012; McGregor et al., 2013; Cobb et al., 2013; Carré et al.,
335   2014; Chen et al., 2016; Grothe et al., 2019). In the central and eastern Pacific, the deepest reduction (around 2/3 in terms
of 2-7yr variance) are observed in the 3-5 ka BP interval, rather than around the canonical 6 ka midpoint (Emile-Geay et al.,
2016). This reduction has been simulated by models of various complexity (e.g. Clement et al., 2000; Liu et al., 2000; Zheng
et al., 2008; Chiang et al., 2009; An and Choi, 2014; Liu et al., 2014a) and is a feature of the PMIP4-CMIP6 *midHolocene*

simulations (Tab. S1 Brown et al., submitted). Analyses of simulated and reconstructed changes in tropical Pacific climate
340   variability (Emile-Geay et al., 2016) showed that the PMIP3-CMIP5 models rarely produced a reduction in ENSO as large as
shown by the paleoclimate observations, though mid-Holocene boundary conditions did increase those odds. This is also true
for most of the PMIP4-CMIP6 models (Table S1). With the exception of MIROC-ES2L, the models produce a reduction in
ENSO variability but this is much smaller than the reduction implied by the palaeoclimate records. A key result of Emile-Geay
et al. (2016) was that while models showed an inverse relationship between ENSO variance (inferred from 2-7yr bandpass
345   filtered metrics of ENSO) and seasonality (defined as the range of the monthly-mean annual cycle), the observations showed
either no or a weakly positive relationship. Proxy evidence also points to an increased zonal SST gradient in the equatorial
Pacific during the mid-Holocene (Koutavas et al., 2002; Linsley et al., 2010; Carré et al., 2014), whilst the PMIP4-CMIP6
ensemble yields a slight decrease in the gradient (Table S1). Analysis of equatorial Pacific climate change and variability finds
little evidence for simulated relationship between either the seasonality or SST gradient and ENSO variance in the PMIP4-
350   CMIP6 ensemble (Brown et al., submitted).

4   Conclusions

The PMIP4-CMIP6 *midHolocene* simulations show changes in seasonal temperatures and precipitation that are consistent
with the expected response to changes in insolation forcing. The broadscale patterns of change are similar to those seen in
previous generations of models, most particularly the PMIP3-CMIP5 ensemble. Both ensembles show increased temperature
355   seasonality, but with enhanced warming year-round at high northern and southern latitudes resulting from higher obliquity and
feedbacks from sea ice and snow cover. Both show an enhancement of the Northern Hemisphere monsoons and a weakening of
the southern hemisphere monsoons. Neither the PMIP4-CMIP6 nor the PMIP3-CMIP5 models show a significant change in the
AMOC during the mid-Holocene. This suggests that the changes in wind forcing, temperature gradients, seasonality of sea-ice
and precipitation are not sufficient to alter the overall AMOC strength, although investigations into its various components may
360   deliver greater insight.

Although the geographic and seasonal patterns of temperature changes in the PMIP4-CMIP6 ensemble are very similar to
those seen in the PMIP3-CMIP5 ensemble, the PMIP4-CMIP6 ensemble is cooler than the PMIP3-CMIP5 ensemble in both
summer and winter. This difference is consistent with the change in radiative forcing induced by using realistic GHG concen-
trations in the PMIP4-CMIP6. Improvements in the models themselves could also contribute to this difference, in particular
365   changes in the implementation of aerosols. There is a considerable spread in simulated regional *midHolocene* responses be-
tween the PMIP4-CMIP6 models. In some cases, for example in the strength of the AMOC, this spread is clearly related to
the spread in the *piControl* simulations. Biases in the *piControl* may also help to explain the underestimation of the northward
expansion of the NH monsoons, since the global monsoon domain is underestimated by both CMIP/PMIP ensembles in the
piControl compared to observations.

370   This preliminary analysis of the PMIP4-CMIP6 *midHolocene* simulations already demonstrates the utility of running palaeo-
climate simulations to evaluate the ability of state-of-the-art models to simulate climate change and thus to simulate the likely

trajectory of future climate changes realistically. Although it is disappointing that the PMIP4-CMIP6 simulations are not significantly better than the PMIP3-CMIP5 models in capturing important features of the mid-Holocene climate, analyses of the mechanisms giving rise to these failures should shed light on the need for improved process representation in future versions

375   of the CMIP climate models. The examination of the how biases in the *piControl* simulations impact on the simulation of past climates is directly relevant to understanding how modern biases are propagated into future projections. Furthermore, the similarities between the PMIP4-CMIP6 and PMIP3-CMIP5 simulations provides an argument for combining these to create a single ensemble, which will considerably enhance the statistical skill of future analyses. Sensitivity tests, already planned within the framework of PMIP4-CMIP6 (Otto-Bliesner et al., 2017), should help to disentangle the impacts of specific feed-

380   backs on simulated climate changes. Finally, the PMIP4-CMIP6 *midHolocene* simulations provide an opportunity both for quantitative evaluation and derivation of emergent constraints on sensitivity and feedback strength.

Code and data availability.   The necessary output variables from both the *midHolocene* and *piControl* simulations shall be freely available from the Earth System Grid Federation at https://esgf-node.llnl.gov/search/cmip6/. (HadGEM3-GC31-LL, AWI-ESM-1-1-LR and UofT-CCSM-4 have committed to lodge their data as soon as practical). A GitHub repository is available at https://github.com/chrisbrierley/

385   PMIP4-midHolocene with the code used for this analysis. The Temperature 12k database, along with latitude-zone and global temperature reconstructions using multiple statistical methods, is available through the World Data Service (NOAA) Paleoclimatology (www.ncdc.noaa.gov/paleo/study/27330) [note to reviewers: the URL will be activated upon publication of the database]. The Bartlein et al. (2011) reconstructions are downloadable as an Electronic Supplementary Material of that article. The Compo et al. (2011) Reanalysis can be foudn at www.esrl.noaa.gov/psd/data/gridded/data.20thC_ReanV2c.html. The precipitation observations of Adler et al. (2003) and Xie and

390   Arkin (1997) are archived at https://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html and https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html respectively. The preindustrial latitudinal average temperatures were created using anomalies of Ilyas et al. (2017) from https://oasishub.co/dataset/global-monthly-temperature-ensemble-1850-to-2016 combined with the HadCRUT4 (Morice et al., 2012) absolute climatological temperatures from https://crudata.uea.ac.uk/cru/data/temperature/.

Author contributions.   There are three tiers of authorship for this research, with the latter two in reverse alphabetical order. C.M.B., A.Z.,

395   S.P.H. and P.Br. performed the bulk of the writing and analysis. C.J.R.W., D.J.R.T., X.S., J-Y.P., R.O., D.S.K., M.K., J.C.H., M.P.E., J.E-G., R.D'A., D.C., M.C. and P.Ba. contributed text and analysis to the research. The third tier of authors contributed data for the manuscript.

Competing interests.   The authors declare no competing interests

Acknowledgements.   We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the groups developing the climate models (listed in Tab. 1 & 2 of this paper) for producing and making

400 available their model output. C.M.B., S.P.H., P.Br., C.J.R.W., X.S., R.D'A., M.C. and G.L. received funded by JPI-Belmont Forum project entitled Palaeoclimate Constraints on Monsoon Evolution and Dynamics (PaCMEDy). C.M.B., C.J.R.W. and S.P.H. were funded in part by NERC (NE/P006752/1). D.J.R.T. and C.M.B. were funded in part by NERC (NE/S009736/1). S.P.H. acknowledges the ERC-funded project GC2.0 (Global Change 2.0: Unlocking the past for a clearer future, grant number 694481). R.O. acknowledges support from the Integrated Research Program for Advancing Climate Models (TOUGOU programme) from the Ministry of Education, Culture, Sports, Science and

405 Technology (MEXT), Japan. The simulations using MIROC models were conducted on the Earth Simulator of JAMSTEC. The NorESM simulations were performed on resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway. D.S.K.,C.R. and N.M. were funded by US-NSF-AGS-1602105. We thank R. Eyles (UCL) for some invaluable database management and preprocessing.

References

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

An, S.-I. and Choi, J.: Mid-Holocene tropical Pacific climate state, annual cycle, and ENSO in PMIP2 and PMIP3, Climate Dynamics, 43, 957–970, https://doi.org/10.1007/s00382-013-1880-z, 2014.

415   Bao, Q., Lin, P., Zhou, T., Liu, Y., Yu, Y., Wu, G., He, B., He, J., Li, L., Li, J., et al.: The flexible global ocean-atmosphere-land system model, spectral version 2: FGOALS-s2, Advances in Atmospheric Sciences, 30, 561–576, https://doi.org/10.1007/s00376-012-2113-9, 2013.

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

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

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

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

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

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

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

Braconnot, P. and Kageyama, M.: Shortwave forcing and feedbacks in Last Glacial Maximum and Mid-Holocene PMIP3 simulations, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 20140424, https://doi.org/10.1098/rsta.2014.0424, 2015.

435   Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, T., Hewitt, C., et al.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum–Part 1: experiments and large-scale features, Climate of the Past, 3, 261–277, https://doi.org/10.5194/cp-3-261-2007, 2007.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nature Climate Change, 2, 417, https://doi.org/10.1038/nclimate1456, 2012.

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

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

Brierley, C., Koch, A., Ilyas, M., Wennyk, N., and Kikstra, J.: Half the world's population already experiences years 1.5 C warmer than
445   preindustrial, https://doi.org/10.31223/osf.io/sbc3f, 2019.

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

Brown, J. R., Moise, A. F., and Colman, R. A.: The South Pacific Convergence Zone in CMIP5 simulations of historical and future climate, Climate dynamics, 41, 2179–2197, https://doi.org/10.1007/s00382-012-1591-x, 2013.

450 Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed fingerprint of a weakening Atlantic Ocean overturning circulation, Nature, 556, 191, https://doi.org/10.1038/s41586-018-0006-5, 2018.

Cao, J., Wang, B., Young-Min, Y., Ma, L., Li, J., Sun, B., Bao, Y., He, J., Zhou, X., and Wu, L.: The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation, Geoscientific Model Development, 11, 2975–2993, https://doi.org/10.5194/gmd-11-2975-2018, 2018.

455 Carré, M., Sachs, J. P., Purca, S., Schauer, A. J., Braconnot, P., Falcón, R. A., Julien, M., and Lavallée, D.: Holocene history of ENSO variance and asymmetry in the eastern tropical Pacific, Science, 345, 1045–1048, https://doi.org/10.1126/science.1252220, 2014.

Chandan, D. and Peltier, W. R.: Regional and global climate for the mid-Pliocene using the University of Toronto version of CCSM4 and PlioMIP2 boundary conditions, Climate of the Past, 13, 919, https://doi.org/10.5194/cp-13-919-2017, 2017.

Chen, S., Hoffmann, S. S., Lund, D. C., Cobb, K. M., Emile-Geay, J., and Adkins, J. F.: A high-resolution speleothem record of
460 western equatorial Pacific rainfall: Implications for Holocene ENSO evolution, Earth and Planetary Science Letters, 442, 61–71, https://doi.org/10.1016/j.epsl.2016.02.050, http://www.sciencedirect.com/science/article/pii/S0012821X16300759, 2016.

Chervin, R. M. and Schenider, S. H.: On determining the statistical significance of climate experiments with general circulation models, Journal of the Atmospheric Sciences, 33, 405–412, https://doi.org/10.1175/1520-0469(1976)033<0405:ODTSSO>2.0.CO;2, 1976.

Chiang, J. C. H., Fang, Y., and Chang, P.: Pacific Climate Change and ENSO Activity in the Mid-Holocene, Journal of Climate, 22, 923–939,
465 https://doi.org/10.1175/2008JCLI2644.1, 2009.

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

Clement, A. C., Seager, R., and Cane, M. A.: Suppression of El Niño during the mid-Holocene by changes in the Earth's orbit, Paleoceanography, 15, 731–737, https://doi.org/10.1029/1999PA000466, 2000.

Cobb, K. M., Westphal, N., Sayani, H. R., Watson, J. T., Di Lorenzo, E., Cheng, H., Edwards, R. L., and Charles, C. D.: Highly Variable El
470 Niño–Southern Oscillation Throughout the Holocene, Science, 339, 67, https://doi.org/10.1126/science.1228246, 2013.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., et al.: Long-term climate change: projections, commitments and irreversibility, in: Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1029–1136, Cambridge University Press, 2013.

475 Collins, W., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C., Joshi, M., Liddicoat, S., et al.: Development and evaluation of an Earth-System model–HadGEM2, Geoscientific Model Development, 4, 1051–1075, https://doi.org/10.5194/gmd-4-1051-2011, 2011.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al.: The twentieth century reanalysis project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28,
480 https://doi.org/10.1002/qj.776, 2011.

D'Agostino, R., Bader, J., Bordoni, S., Ferreira, D., and Jungclaus, J.: Northern Hemisphere Monsoon Response to Mid-Holocene Orbital Forcing and Greenhouse Gas-Induced Global Warming, Geophysical Research Letters, 46, 1591–1601, https://doi.org/10.1029/2018GL081589, 2019.

Dima, M. and Lohmann, G.: Evidence for two distinct modes of large-scale ocean circulation changes over the last century, Journal of Climate, 23, 5–16, https://doi.org/10.1175/2009JCLI2867.1, 2010.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., et al.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Emile-Geay, J., Cobb, K. M., Carre, M., Braconnot, P., Leloup, J., Zhou, Y., Harrison, S. P., Correge, T., McGregor, H. V., Collins, M., Driscoll, R., Elliot, M., Schneider, B., and Tudhope, A.: Links between tropical Pacific seasonal, interannual and orbital variability during the Holocene, Nature Geosci, 9, 168–173, https://doi.org/10.1038/ngeo2608, 2016.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization., Geoscientific Model Development, 9, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models, in: Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 741–866, Cambridge University Press, 2013.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., et al.: The community climate system model version 4, Journal of Climate, 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

Gettelman, A., Hannay, C., Bacmeister, J., Neale, R., Pendergrass, A., Danabasoglu, G., Lamarque, J.-F., Fasullo, J., Bailey, D., Lawrence, D., et al.: High climate sensitivity in the Community Earth System Model Version 2 (CESM2), Geophysical Research Letters, https://doi.org/10.1029/2019GL083978, 2019.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., et al.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, 2013.

Gregory, J. M., Ingram, W., Palmer, M., Jones, G., Stott, P., Thorpe, R., Lowe, J., Johns, T., and Williams, K.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, https://doi.org/10.1029/2003GL018747, 2004.

Grothe, P., Cobb, K. M., Liguori, G., Di Lorenzo, E., Capotondi, A., Lu, Y., Cheng, H., Edwards, R. L., Southon, J., Santos, G., Deocampo, D., Lynch-Stieglitz, J., Chen, T., Sayani, H. R., Townsend, K., Hagos, M., O'Connor, G., Thompson, D. M., Toth, L., Conroy, J. L., and Moore, A.: Evidence for intensification of El Niño - Southern Oscillation over the late 20th century, Geophys. Res. Lett., https://doi.org/10.1029/2019GL083906, 2019.

Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM), Geoscientific Model Development, 12, 343–362, https://doi.org/10.5194/gmd-12-343-2019, https://www.geosci-model-dev.net/12/343/2019/, 2019.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Description of the MIROC-ES2L Earth system model and evaluation of its climate–biogeochemical processes and feedbacks, Geoscientific Model Development Discussions, 2019, 1–73, https://doi.org/10.5194/gmd-2019-275, https://www.geosci-model-dev-discuss.net/gmd-2019-275/, 2019.

Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A.: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene, Climate of the Past, 9, 811–823, https://doi.org/10.5194/cp-9-811-2013, 2013.

Harrison, S., Bartlein, P., Brewer, S., Prentice, I., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Climate Dynamics, 43, 671–688, https://doi.org/10.1007/s00382-013-1922-6, 2014.

Harrison, S. P., Bartlein, P., Izumi, K., Li, G., Annan, J., Hargreaves, J., Braconnot, P., and Kageyama, M.: Evaluation of CMIP5 palaeo-
525   simulations to improve climate projections, Nature Climate Change, 5, 735–743, https://doi.org/10.1038/nclimate2649, 2015.

Harrison, S. P., Bartlein, P. J., and Prentice, I. C.: What have we learnt from palaeoclimate simulations?, Journal of Quaternary Science, 31, 363–385, https://doi.org/10.1002/jqs.2842, 2016.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Climate Dynamics, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.

530   Hazeleger, W., Wang, X., Severijns, C., Ştefănescu, S., Bintanja, R., Sterl, A., Wyser, K., Semmler, T., Yang, S., Van den Hurk, B., et al.: EC-Earth V2. 2: description and validation of a new seamless earth system prediction model, Climate dynamics, 39, 2611–2629, https://doi.org/10.1007/s00382-011-1228-5, 2012.

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

Hoogakker, B. A., Chapman, M. R., McCave, I. N., Hillaire-Marcel, C., Ellison, C. R., Hall, I. R., and Telford, R. J.: Dynamics of North
535   Atlantic deep water masses during the Holocene, Paleoceanography, 26, https://doi.org/10.1029/2011PA002155, 2011.

Hopcroft, P. O. and Valdes, P. J.: On the Role of Dust-Climate Feedbacks During the Mid-Holocene, Geophysical Research Letters, 46, 1612–1621, https://doi.org/10.1029/2018GL080483, 2019.

Ilyas, M., Brierley, C. M., and Guillas, S.: Uncertainty in regional temperatures inferred from sparse global observations: Application to a probabilistic classification of El Niño, Geophysical Research Letters, 44, 9068–9074, https://doi.org/10.1002/2017GL074596, 2017.

540   Jackson, L., Kahana, R., Graham, T., Ringer, M., Woollings, T., Mecking, J., and Wood, R.: Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM, Climate dynamics, 45, 3299–3316, https://doi.org/10.1007/s00382-015-2540-2, 2015.

Jeffrey, S., Rotstayn, L., Collier, M., Dravitzki, S., Hamalainen, C., Moeseneder, C., Wong, K., and Syktus, J.: Australia's CMIP5 submission using the CSIRO Mk3. 6 model, Aust. Meteor. Oceanogr. J, 63, 1–13, https://doi.org/10.22499/2.6301.001, 2013.

Jiang, D., Tian, Z., and Lang, X.: Mid-Holocene net precipitation changes over China: model–data comparison, Quaternary Science Reviews,
545   82, 104–120, https://doi.org/10.1016/j.quascirev.2013.10.017, 2013.

Jiang, D., Tian, Z., and Lang, X.: Mid-Holocene global monsoon area and precipitation from PMIP simulations, Climate Dynamics, 44, 2493–2512, https://doi.org/10.1007/s00382-014-2175-8, 2015.

Joussaume, S. and Braconnot, P.: Sensitivity of paleoclimate simulation results to season definitions, Journal of Geophysical Research: Atmospheres, 102, 1943–1956, https://doi.org/10.1029/96JD01989, 1997.

550   Joussaume, S., Taylor, K., Braconnot, P., Mitchell, J., Kutzbach, J., Harrison, S., Prentice, I., Broccoli, A., Abe-Ouchi, A., Bartlein, P., et al.: Monsoon changes for 6000 years ago: results of 18 simulations from the Paleoclimate Modeling Intercomparison Project (PMIP), Geophysical Research Letters, 26, 859–862, https://doi.org/10.1029/1999GL900126, 1999.

Juckes, M., Taylor, K. E., Durack, P., Lawrence, B., Mizielinski, M., Pamment, A., Peterschmitt, J.-Y., Rixen, M., and Sénésis, S.: The CMIP6 Data Request (version 01.00.31), Geoscientific Model Development Discussions, 2019, 1–35, https://doi.org/10.5194/gmd-2019-
555   219, https://www.geosci-model-dev-discuss.net/gmd-2019-219/, 2019.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., et al.: The PMIP4 contribution to CMIP6-Part 1: Overview and over-arching analysis plan, Geoscientific Model Development, 11, 1033–1057, https://doi.org/10.5194/gmd-11-1033-2018, 2018.

Kaufman, D., McKay, N., Routson, C., Erb, M., Davis, B., Heiri, O., Jaccard, S., Tierney, J., Dätwyler, C., et al.: A global database of Holocene paleo-temperature records, Scientific Data, in press.

Kaufman, D., McKay, N., Routson, C., Erb, M., Dätwyler, C., Sommer, P., Heiri, O., and Davis, B.: Holocene global mean surface temperature: a multi-method reconstruction approach, submitted.

Kohfeld, K. E. and Harrison, S. P.: How well can we simulate past climates? Evaluating the models using global palaeoenvironmental datasets, Quaternary Science Reviews, 19, 321–346, https://doi.org/10.1016/S0277-3791(99)00068-2, 2000.

Koutavas, A. and Joanides, S.: El Niño-Southern Oscillation extrema in the Holocene and Last Glacial Maximum, Paleoceanography, 27, PA4208, https://doi.org/10.1029/2012PA002378, 2012.

Koutavas, A., Lynch-Stieglitz, J., Marchitto, T. M., and Sachs, J. P.: El Nino-like pattern in ice age tropical Pacific sea surface temperature, Science, 297, 226–230, https://doi.org/10.1126/science.1072376, 2002.

Li, L., Lin, P., Yu, Y., Wang, B., Zhou, T., Liu, L., Liu, J., Bao, Q., Xu, S., Huang, W., et al.: The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2, Advances in Atmospheric Sciences, 30, 543–560, https://doi.org/10.1007/s00376-012-2140-6, 2013.

Linsley, B. K., Rosenthal, Y., and Oppo, D. W.: Holocene evolution of the Indonesian throughflow and the western Pacific warm pool, Nature Geoscience, 3, 578–583, https://doi.org/10.1038/ngeo920, 2010.

Lippold, J., Pöppelmeier, F., Süfke, F., Gutjahr, M., Goepfert, T. J., Blaser, P., Friedrich, O., Link, J. M., Wacker, L., Rheinberger, S., et al.: Constraining the variability of the atlantic meridional overturning circulation during the holocene, Geophysical Research Letters, https://doi.org/10.1029/2019GL084988, 2019.

Liu, Z., Kutzbach, J., and Wu, L.: Modeling climate shift of El Niño variability in the Holocene, Geophysical Research Letters, 27, 2265–2268, https://doi.org/10.1029/2000GL011452, 2000.

Liu, Z., Lu, Z., Wen, X., Otto-Bliesner, B. L., Timmermann, A., and Cobb, K. M.: Evolution and forcing mechanisms of El Niño over the past 21,000 years, Nature, 515, 550–553, https://doi.org/10.1038/nature13963, 2014a.

Liu, Z., Zhu, J., Rosenthal, Y., Zhang, X., Otto-Bliesner, B. L., Timmermann, A., Smith, R. S., Lohmann, G., Zheng, W., and Timm, O. E.: The Holocene temperature conundrum, Proceedings of the National Academy of Sciences, 111, E3501–E3505, https://doi.org/10.1073/pnas.1407229111, 2014b.

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

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

Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., and Brewer, S.: Reconciling divergent trends and millennial variations in Holocene temperatures, Nature, 554, 92, https://doi.org/10.1038/nature25464, 2018.

Mauri, A., Davis, B., Collins, P., and Kaplan, J.: The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data–model comparison, Climate of the Past, 10, 1925–1938, https://doi.org/10.5194/cp-10-1925-2014, 2014.

McGregor, H. V. and Gagan, M. K.: Western Pacific coral $\delta^{18}O$ records of anomalous Holocene variability in the El Niño–Southern Oscillation, Geophysical Research Letters, 31, L11 204, https://doi.org/10.1029/2004GL019972, 2004.

McGregor, H. V., Fischer, M. J., Gagan, M. K., Fink, D., Phipps, S. J., Wong, H., and Woodroffe, C. D.: A weak El Niño-Southern Oscillation with delayed seasonal growth around 4,300 years ago, Nature Geoscience, 6, 949–953, https://doi.org/10.1038/ngeo1936, 2013.

McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D., and Brown-Leger, S.: Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes, Nature, 428, 834, https://doi.org/10.1038/nature02494, 2004.

Moffa-Sanchez, P., Hall, I. R., Thornalley, D. J., Barker, S., and Stewart, C.: Changes in the strength of the Nordic Seas Overflows over the past 3000 years, Quaternary Science Reviews, 123, 134–143, https://doi.org/10.1016/j.quascirev.2015.06.007, 2015.

600 Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017187, 2012.

Ng, H. C., Robinson, L. F., McManus, J. F., Mohamed, K. J., Jacobel, A. W., Ivanovic, R. F., Gregoire, L. J., and Chen, T.: Coherent deglacial changes in western Atlantic Ocean circulation, Nature communications, 9, https://doi.org/10.1038/s41467-018-05312-3, 2018.

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

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

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

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

615 Pausata, F. S., Messori, G., and Zhang, Q.: Impacts of dust reduction on the northward expansion of the African monsoon during the Green Sahara period, Earth and Planetary Science Letters, 434, 298–307, https://doi.org/10.1016/j.epsl.2015.11.049, 2016.

Perez-Sanz, A., Li, G., González-Sampériz, P., and Harrison, S. P.: Evaluation of modern and mid-Holocene seasonal precipitation of the Mediterranean and northern Africa in the CMIP5 simulations, Climate of the Past, 10, 551–568, https://doi.org/10.5194/cp-10-551-2014, 2014.

620 Phillips, A. S., Deser, C., and Fasullo, J.: Evaluating modes of variability in climate models, Eos, Transactions American Geophysical Union, 95, 453–455, https://doi.org/10.1002/2014EO490002, 2014.

Phipps, S., Rotstayn, L., Gordon, H., Roberts, J., Hirst, A., and Budd, W.: The CSIRO Mk3L climate system model version 1.0–Part 2: Response to external forcings, Geoscientific Model Development, 5, 649–682, https://doi.org/10.5194/gmd-5-649-2012, 2012.

Prado, L. F., Wainer, I., and Chiessi, C. M.: Mid-Holocene PMIP3/CMIP5 model results: Intercomparison for the South American monsoon
625 system, The Holocene, 23, 1915–1920, https://doi.org/10.1177/0959683613505336, 2013.

Prentice, I. C., Jolly, D., and Participants, B. .: Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa, Journal of biogeography, 27, 507–519, https://doi.org/10.1046/j.1365-2699.2000.00425.x, 2000.

Rahmstorf, S.: Ocean circulation and climate during the past 120,000 years, Nature, 419, 207, https://doi.org/10.1038/nature01090, 2002.

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J.: Exceptional twentieth-century
630 slowdown in Atlantic Ocean overturning circulation, Nature climate change, 5, 475, https://doi.org/10.1038/nclimate2554, 2015.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L.-B., and Yiou, P.:

Using palaeo-climate comparisons to constrain future projections in CMIP5, Climate of the Past, 10, 221–250, https://doi.org/10.5194/cp-10-221-2014, https://www.clim-past.net/10/221/2014/, 2014a.

635 Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., et al.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 6, 141–184, https://doi.org/10.1002/2013MS000265, 2014b.

Shi, X. and Lohmann, G.: Simulated response of the mid-Holocene Atlantic meridional overturning circulation in ECHAM6-FESOM/MPIOM, Journal of Geophysical Research: Oceans, 121, 6444–6469, https://doi.org/10.1002/2015JC011584, 2016.

640 Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Gößling, H. F., et al.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Climate Dynamics, 44, 757–780, https://doi.org/10.1007/s00382-014-2290-6, 2015.

Sueyoshi, T., Ohgaito, R., Yamamoto, A., Chikamoto, M., Hajima, T., Okajima, H., Yoshimori, M., Abe, M., O'ishi, R., Saito, F., et al.: Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM, Geoscientific Model Development, 6,
645 819–836, https://doi.org/10.5194/gmd-6-819-2013, 2013.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183–7192, https://doi.org/ 10.1029/2000JD900719, 2001.

Thornalley, D. J., Blaschek, M., Davies, F. J., Praetorius, S., Oppo, D. W., McManus, J. F., Hall, I. R., Kleiven, H., Renssen, H., and McCave, I. N.: Long-term variations in Iceland–Scotland overflow strength during the Holocene, Climate of the Past, 9, 2073–2084,
650 https://doi.org/10.5194/cp-9-2073-2013, 2013.

Thornalley, D. J., Oppo, D. W., Ortega, P., Robson, J. I., Brierley, C. M., Davis, R., Hall, I. R., Moffa-Sanchez, P., Rose, N. L., Spooner, P. T., et al.: Anomalously weak Labrador Sea convection and Atlantic overturning during the past 150 years, Nature, 556, 227, https://doi.org/10.1038/s41586-018-0007-4, 2018.

Tudhope, A. W., Chilcott, C. P., McCulloch, M. T., Cook, E. R., Chappell, J., Ellam, R. M., Lea, D. W., Lough, J. M., and
655 Shimmield, G. B.: Variability in the El Niño-Southern Oscillation through a glacial-interglacial cycle, Science, 291, 1511–1517, https://doi.org/10.1126/science.1057969, 2001.

Voldoire, A., Sanchez-Gomez, E., y Mélia, D. S., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., et al.: The CNRM-CM5. 1 global climate model: description and basic evaluation, Climate Dynamics, 40, 2091–2121, https://doi.org/10.1007/s00382-011-1259-y, 2013.

660 Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., Iakovlev, N. G., Shestakova, A. A., et al.: Simulation of the modern climate using the INM-CM48 climate model, Russian Journal of Numerical Analysis and Mathematical Modelling, 33, 367–374, https://doi.org/10.1515/rnam-2018-0032, 2018.

Wang, B., Kim, H.-J., Kikuchi, K., and Kitoh, A.: Diagnostic metrics for evaluation of annual and diurnal cycles, Climate dynamics, 37, 941–955, https://doi.org/10.1007/s00382-010-0877-0, 2011.

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

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

Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 100, Academic press, 2011.

670   Williams, K., Copsey, D., Blockley, E., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H., Hill, R., et al.: The
        Met Office global coupled model 3.0 and 3.1 (GC3. 0 and GC3. 1) configurations, Journal of Advances in Modeling Earth Systems, 10,
        357–380, https://doi.org/10.1002/2017MS001115, 2018.

      Xiao-Ge, X., Tong-Wen, W., Jiang-Long, L., Zai-Zhi, W., Wei-Ping, L., and Fang-Hua, W.: How well does BCC_CSM1.1
        reproduce the 20th century climate change over China?, Atmospheric and Oceanic Science Letters, 6, 21–26,
675   https://doi.org/10.1080/16742834.2013.11447053, 2013.

      Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical
        model outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, https://doi.org/10.1175/2008JAMC1921.1, 1997.

      Yoshimori, M. and Suzuki, M.: The relevance of mid-Holocene Arctic warming to the future, Climate of the Past, 15, 1375–1394,
        https://doi.org/10.5194/cp-15-1375-2019, 2019.

680   Yukimoto, S., Adachi, Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Tanaka, T. Y., Shindo, E., Tsujino, H., Deushi, M., et al.: A
        new global climate model of the Meteorological Research Institute: MRI-CGCM3—model description and basic performance—, Journal
        of the Meteorological Society of Japan. Ser. II, 90, 23–64, https://doi.org/10.2151/jmsj.2012-A02, 2012.

      Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., et al.: The
        Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2. 0: Description and Basic Evaluation of the Physical
685   Component, Journal of the Meteorological Society of Japan. Ser. II, https://doi.org/10.2151/jmsj.2019-051, 2019.

      Zheng, W., Braconnot, P., Guilyardi, E., Merkel, U., and Yu, Y.: ENSO at 6ka and 21ka from ocean–atmosphere coupled model simulations,
        Clim. Dyn., 30, 745–762, https://doi.org/10.1007/s00382-007-0320-3, 2008.
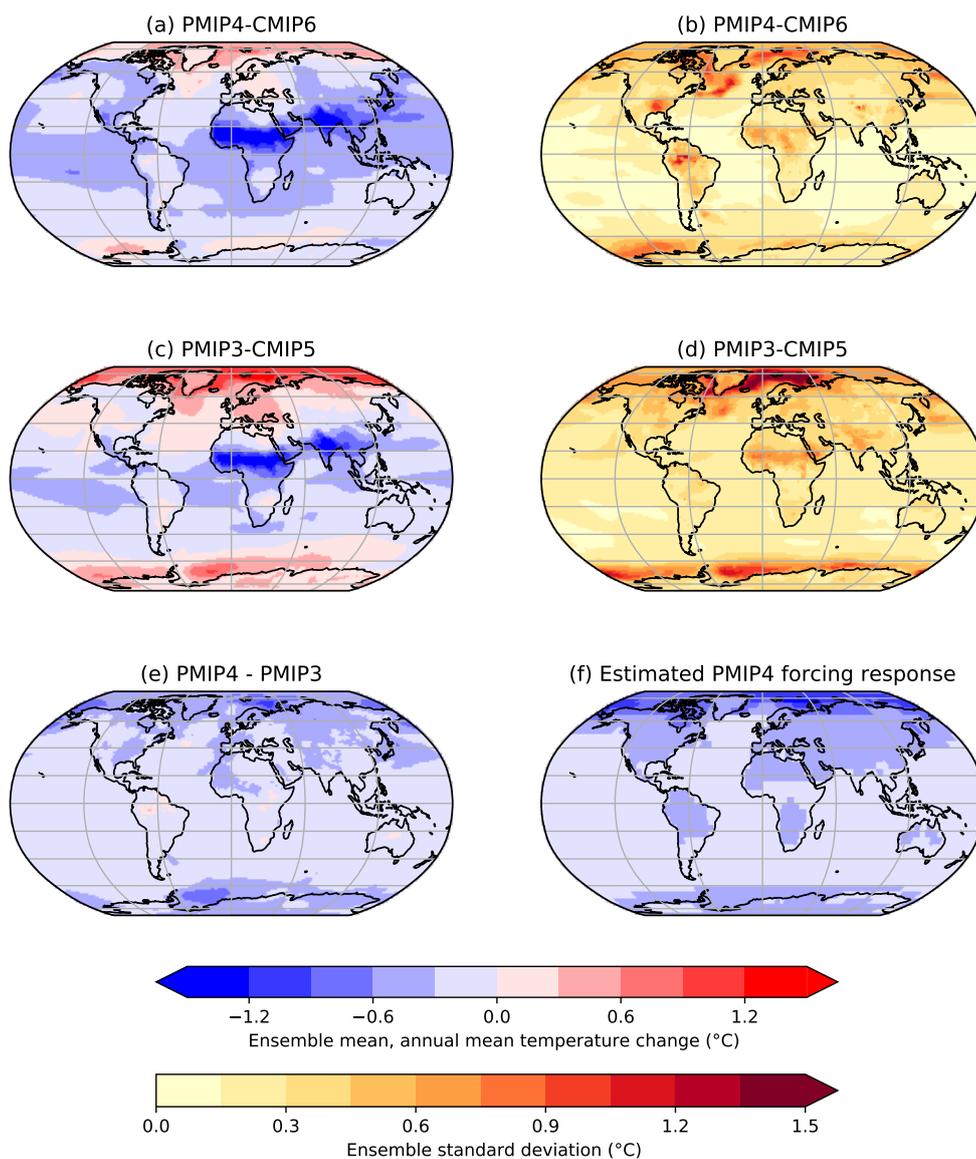
Figure 1. Annual mean surface temperature change in the *midHolocene* simulations (°C). (a) The ensemble mean, annual mean temperature changes in PMIP4-CMIP6 (*midHolocene - piControl*) and (b) its standard devation. (c) The ensemble mean, annual mean temperature change in PMIP3-CMIP5 and (d) its standard devation. (e) The difference in temperature between the two ensembles. (f) The estimated response to the greenhouse gas concentration reductions in the experimental protocol.
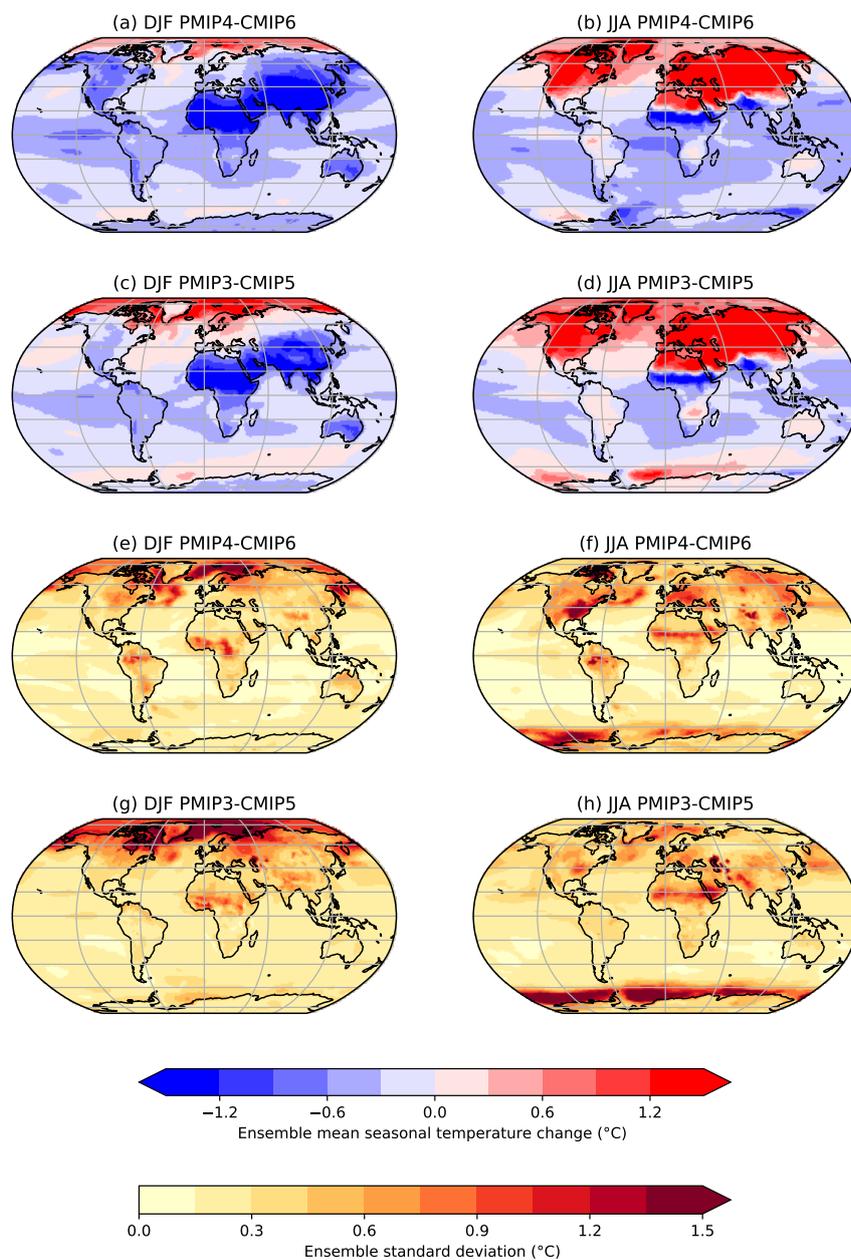
Figure 2. Seasonal surface temperature changes in the *midHolocene* simulations (°C). (a,b) The ensemble mean temperature changes in PMIP4-CMIP6 (*midHolocene - piControl*) in DJF and JJA. (c,d) The ensemble mean temperature changes in PMIP3-CMIP5 in DJF and JJA. The standard devations in seasonal temperature changes seen across the ensembles: (e) DJF in PMIP4-CMIP6, (f) JJA in PMIP4-CMIP6, (g) DJF in PMIP3-CMIP5 and (h) JJA in PMIP3-CMIP6.
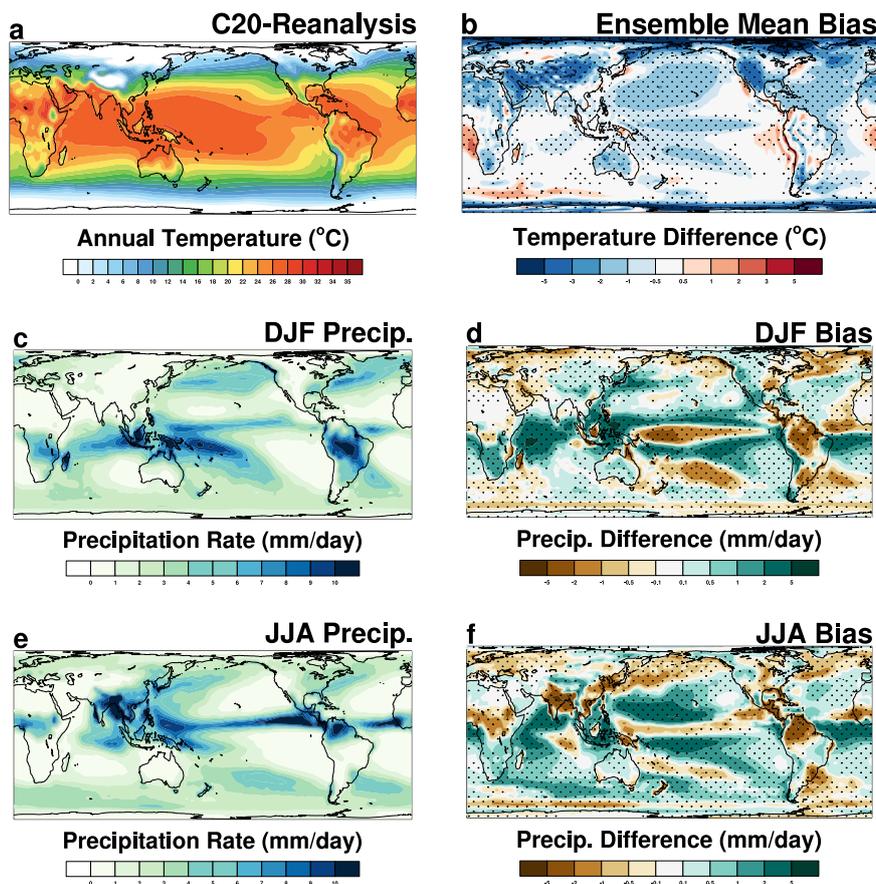
Figure 3. Comparison of the CMIP6 ensemble to observations. (a) The annual mean surface temperatures in the C20 Reanalysis (Compo et al., 2011) between 1981-1900. (b) The ensemble mean difference in annual surface air temperature from the C20 Reanalysis within the *piControl* simulations. Ability of the ensemble to simulate the seasonal cycle of precipitation for the present-day. (c,e) The precipitation climatology seen in the GPCP (Adler et al., 2003) observational dataset between 1971-2000 for DJF and JJA respectively. (d,f) The ensemble mean difference in seasonal precipitation from GPCP within the *piControl* simulations for DJF and JJA respectively. Stippling indicates that two-thirds of the models agree on the sign of the bias.
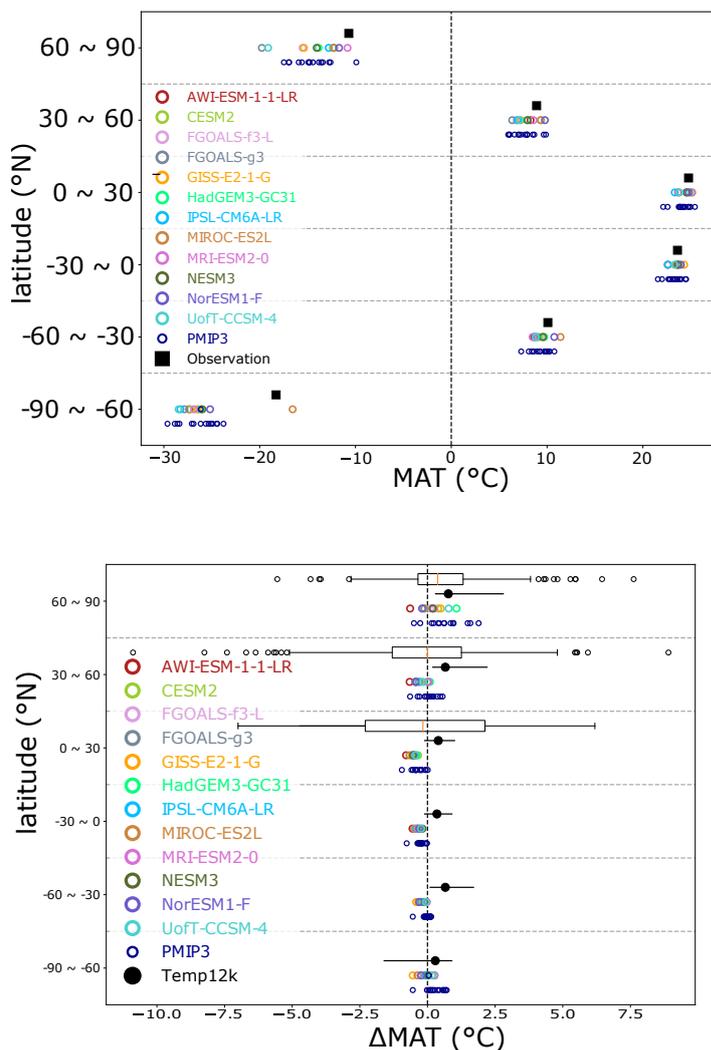
Figure 4. Zonal averaged temperatures within the PMIP4-CMIP6 ensemble. (a) Comparison of the *piControl* zonal mean temperature profile of individual climate models to the 1850-1900 observations. The area-averaged, annual mean surface air temperature for 30°latitude bands in the CMIP6 models (identified), CMIP5 models (blue circles) and a spatially complete compilation of instrumental observations over 1850-1900 (black, Ilyas et al., 2017; Morice et al., 2012). (b) The simulated annual mean temperature change averaged over 30° zonal bands for each of the individual CMIP6 models. The equivalent changes estimated from the Temperature 12k compilation (Kaufman et al., in press) via a multi-method approach along with their 80% confidence interval. The distribution of Bartlein et al. (2011) reconstructed temperatures within each latitude bands are shown in the NH.
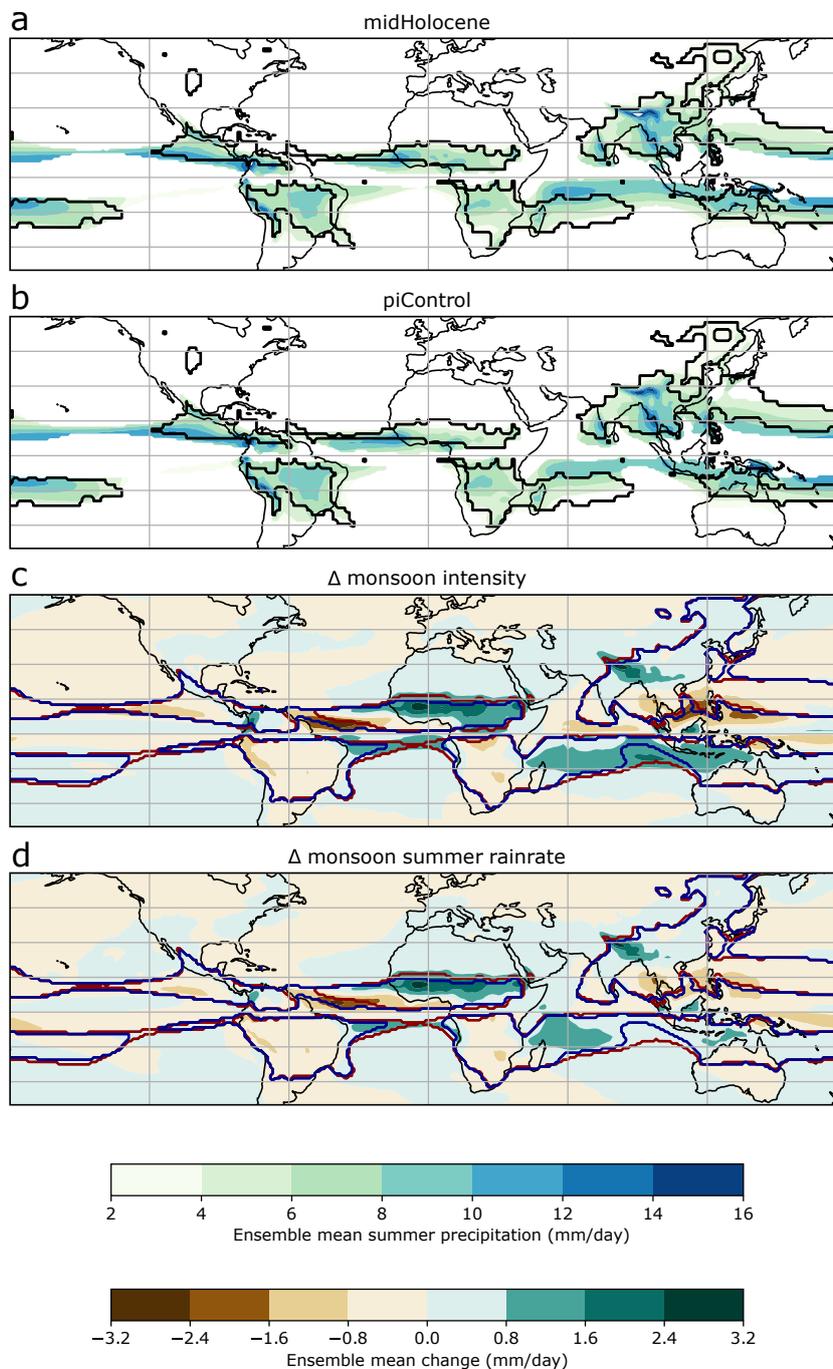
Figure 5. PMIP4-CMIP6 ensemble mean global monsoon domain (mm/day). The monsoon domain for each simulation is identified applying the definitions of Wang et al. (2011) and in sect. 2.2 to the PMIP4-CMIP6 ensemble mean of both (a) the *midHolocene* and (b) the *piControl* simulations. The black contour in (a,b) shows the boundary of the domain derived from present-day observations (Adler et al., 2003). The simulated changes in the monsoon domain are determined by changes in both (c) the monsoon intensity and (d) the summer rain rate. In (c,d) the red and blue contours show the boundary of *midHolocene* and *piControl* global monsoon domains respectively.
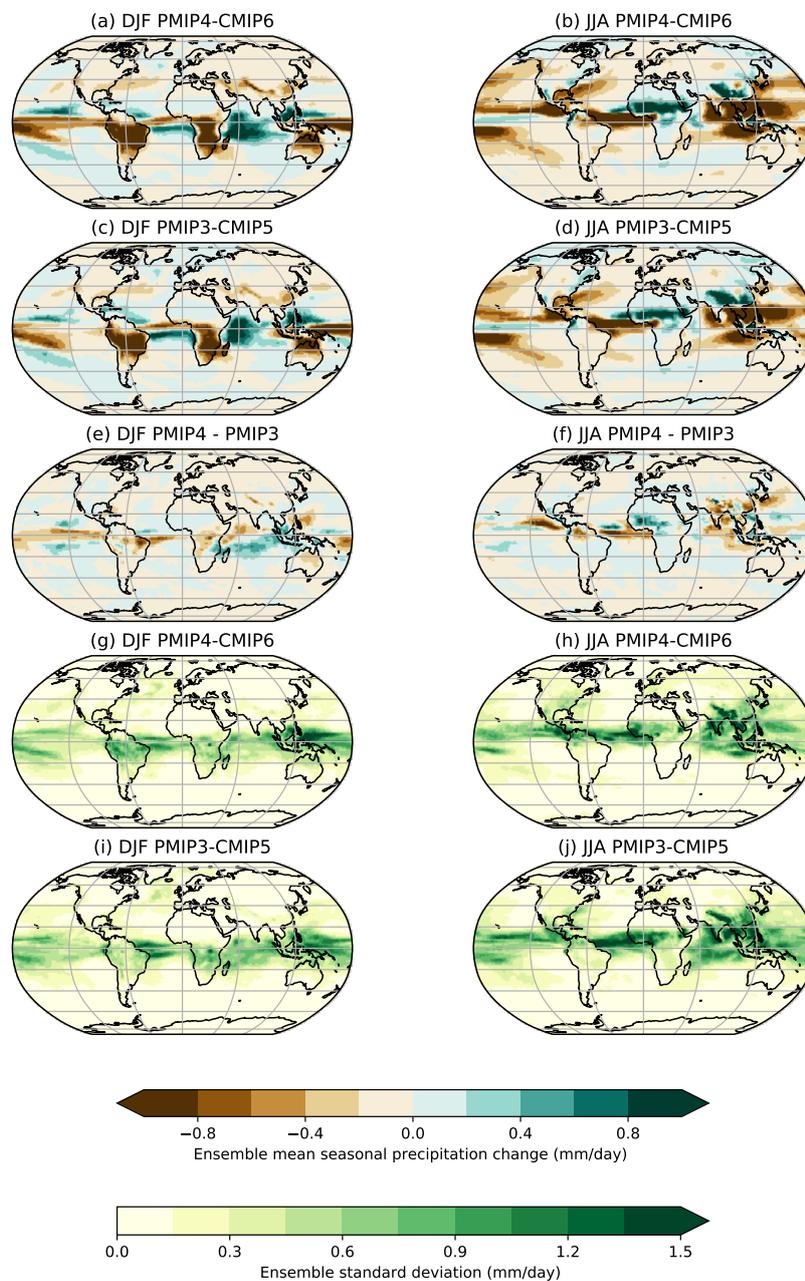
Figure 6. *midHolocene* **seasonal changes in precipitation (mm/day).** (a,b) The ensemble mean precipitation changes in PMIP4-CMIP6 (*midHolocene - piControl*) in DJF and JJA. (c,d) The ensemble mean precipitation changes in PMIP3-CMIP5 in DJF and JJA. (e,f) The differences in DJF and JJA precipitation between the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles. The standard devations in seasonal precipitation changes seen across the ensembles: (g) DJF in PMIP4-CMIP6, (h) JJA in PMIP4-CMIP6, (i) DJF in PMIP3-CMIP5 and (j) JJA in PMIP3-CMIP6.
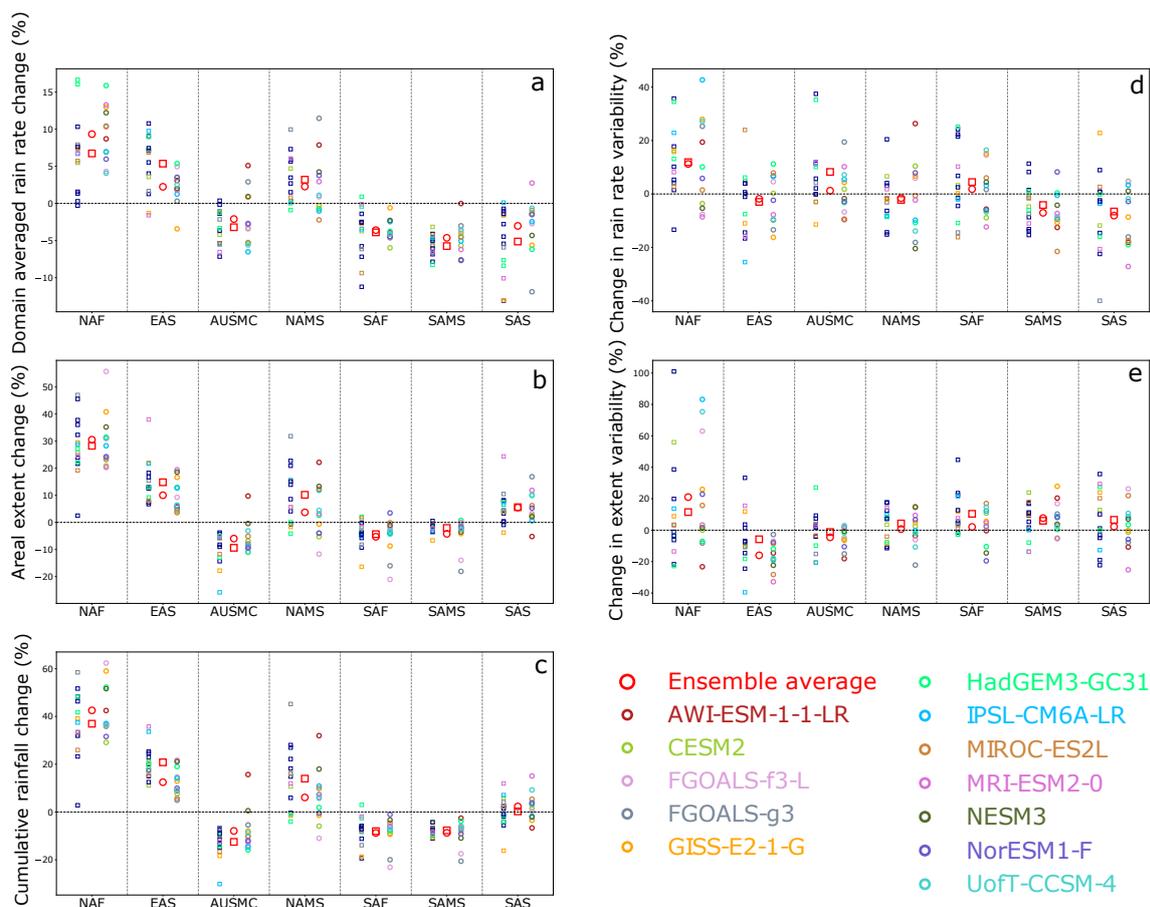
Figure 7. Relative changes in individual *midHolocene* monsoons. Five different monsoon diagnostics (see sect. 2.2) are computed for each of seven different regional domains (Christensen et al., 2013). (a) The change in area-averaged precipitation rate during the monsoon season (MJJAS) for each individual monsoon system. (b) The change in the areal extent of the regional monsoon domains. (c) The percentage change in the total amount of water precipitated in each monsoon season (computed as the precipitation rate multiplied by the areal extent). (d) Change in the standard deviation of interannual variability in the area-averaged precipitation rate. (e) The change in standard deviation of the year-to-year variations in the areal extent of the regional monsoon domain. The abbreviations used to identify each regional domain are: North America Monsoon System (NAMS), North Africa (NAF), Southern Asia (SAS) and East Asia summer (EAS) in the Northern Hemisphere and South America Monsoon System (SAMS), South Africa (SAF) and Australian-Maritime Continent (AUSMC) in the Southern Hemisphere.
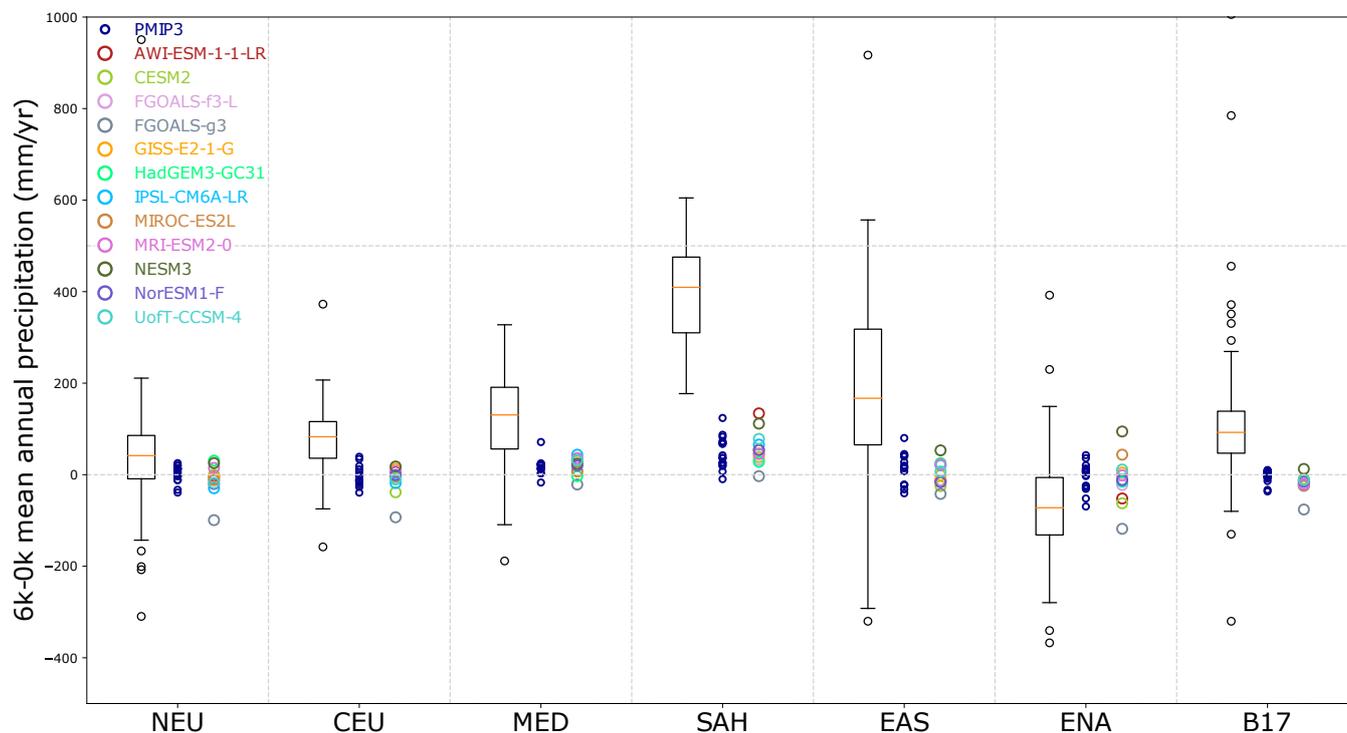
Figure 8. Comparison between simulated annual precipitation changes and pollen-based reconstructions (from Bartlein et al., 2011). Seven regions where multiple quantitative reconstructions exist are chosen. Six of them are defined after Christensen et al. (2013), and are Northern Europe (NEU), Central Europe (CEU), the Mediterranean (MED), the Sahara/Sahel (SAH), East Asia (EAS) and Eastern North America (ENA). Mid-continental Eurasia (B17) is specified by Bartlein et al. (2017) as 40–60°N, 30-120°E. The distribution of reconstructions within the region are shown by boxes and whiskers. The area-averaged change in mean annual precipitation simulated by CMIP6 (individually identifiable) and CMIP5 (blue) within each region is shown for comparison. (After Flato et al., 2013)
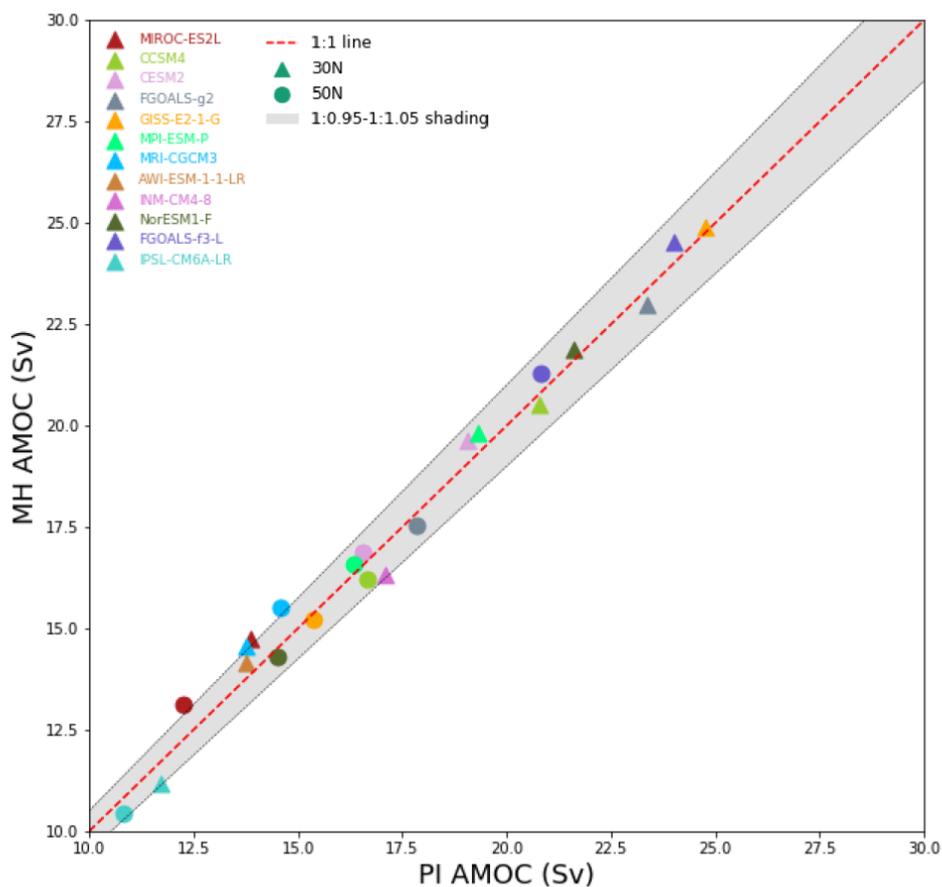
Figure 9. Atlantic Meridional Overturning Circulation in the simulations. The strength of the AMOC is defined as the maximum of the mean meridional mass overturning streamfunction below 500m at 30°and 50°N in the Atlantic. The strength in the *piControl* simulation provides the horizontal axis, whilst the vertical location is given by the strength in the *midHolocene* simulation. Data points lying on the 1:1 line demonstrate no change between the two simulations.
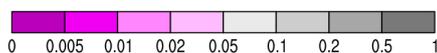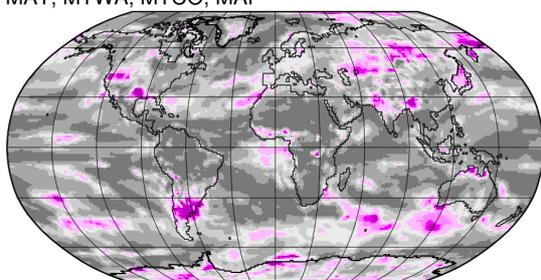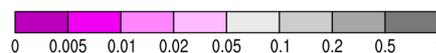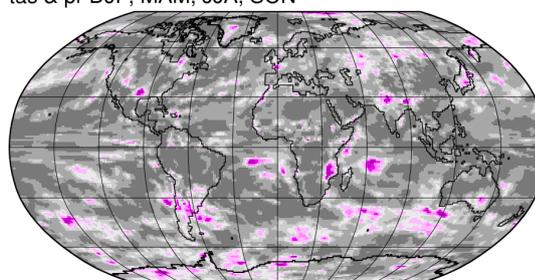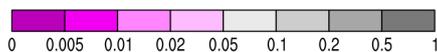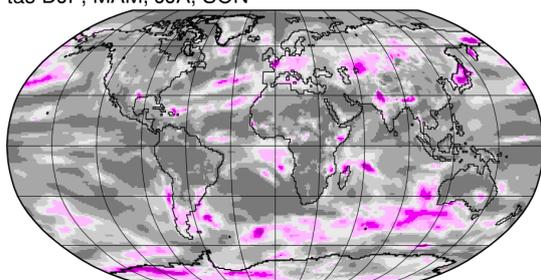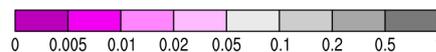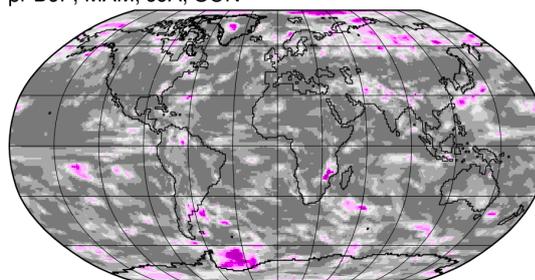
Figure 10. Maps of the p-values of Hotelling's T^2 **test** (Wilks, 2011) comparing the PMIP4-CMIP6 and PMIP3-CMIP5 ensembles. Four different combinations of the key variables analysed here are assessed (given in the top left above the panels). Values less than 0.05 would ordinarily be considered to be significant, but the total number of such values on each individual map does not exceed the false discovery rate. Harrison et al. (2015) presents equivalent analysis comparing PMIP3-CMIP5 with PMIP2-CMIP3 (using the variables in the top left panel).
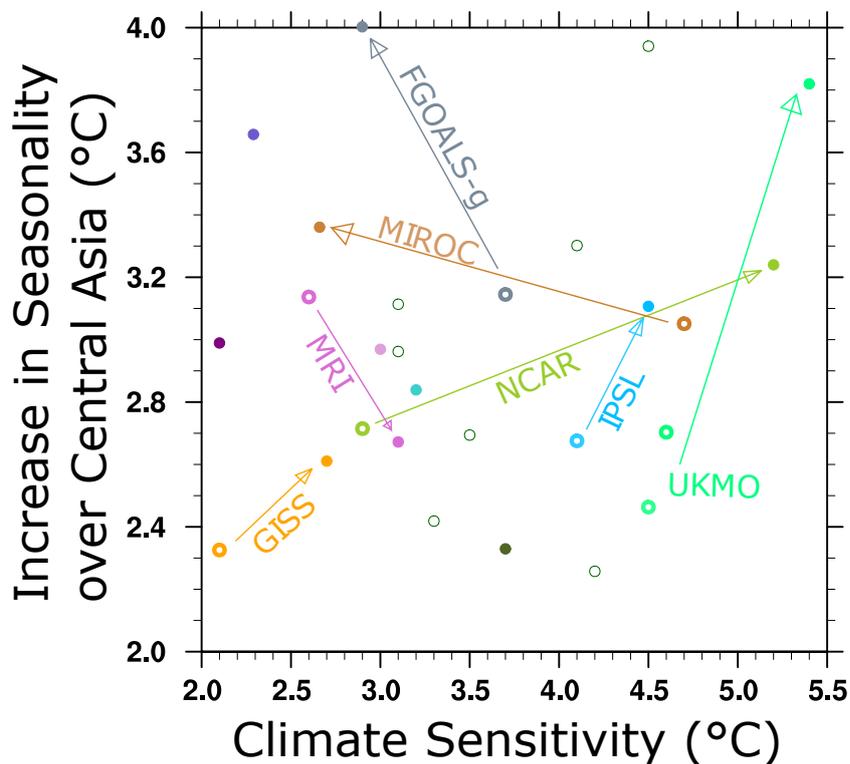
Figure 11. The relationship between equilibrium climate sensitivity and increasing seasonality over Central Asia. The seasonality is computed as the mean temperature of the warmest month minus the mean temperature of the coldest month, averaged over 30–50°N, 60–75°E (Christensen et al., 2013). The shifts between different generations of models are highlighted.

Table 1. Models contributing *midHolocene* simulations under CMIP6.

model	$\Delta T^{eq}_{2xCO_2}$ (K)	*midHolocene* length† (yrs)	*piControl* length† (yrs)	Reference	Notes
AWI-ESM-1-1-LR	3.6	100	100	Sidorenko et al. (2015)	Dynamic Vegetation
CESM2	5.3	700	1200	Gettelman et al. (2019)	Potential Natural Land Cover
FGOALS-f3-L	3.0	500	561	Wang et al. (2020)	–
FGOALS-g3	2.9	500	200	He et al. (2020)	–
GISS-E2-1-G	2.7	100	851	Bauer and Tsigardis (2020)	–
HadGEM3-GC31	5.4	100	100	Williams et al. (2018)	–
INM-CM4-8	2.1	200	531	Volodin et al. (2018)	–
IPSL-CM6A-LR	4.5	550	1200	Boucher, et al. (2020)	TSI of 1361.2 W/m^2
MIROC-ES2L	2.7	100	500	Hajima et al. (2019)	–
MRI-ESM2	3.1	200	701	Yukimoto et al. (2019)	–
NESM3	3.7	100	100	Cao et al. (2018)	–
NorESM1-F	2.3	200	200	Guo et al. (2019)	–
UofT-CCSM-4	3.2	100	100	Chandan and Peltier (2017)	TSI of 1360.89 W/m^2

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

Table 2. Models contributing *midHolocene* simulations under CMIP5

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

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