nature climate change

REVIEW ARTICLE

PUBLISHED ONLINE: 24 JULY 2015 | DOI: 10.1038/NCLIMATE2649

Evaluation of CMIP5 palaeo-simulations to improve climate projections

S. P. Harrison[1,2]*, P. J. Bartlein[3], K. Izumi[3,4,5], G. Li[2], J. Annan[6], J. Hargreaves[6], P. Braconnot[4] and M. Kageyama[4]

Structural differences among models account for much of the uncertainty in projected climate changes, at least until the mid-twenty-first century. Recent observations encompass too limited a range of climate variability to provide a robust test of the ability to simulate climate changes. Past climate changes provide a unique opportunity for out-of-sample evaluation of model performance. Palaeo-evaluation has shown that the large-scale changes seen in twenty-first-century projections, including enhanced land–sea temperature contrast, latitudinal amplification, changes in temperature seasonality and scaling of precipitation with temperature, are likely to be realistic. Although models generally simulate changes in large-scale circulation sufficiently well to shift regional climates in the right direction, they often do not predict the correct magnitude of these changes. Differences in performance are only weakly related to modern-day biases or climate sensitivity, and more sophisticated models are not better at simulating climate changes. Although models correctly capture the broad patterns of climate change, improvements are required to produce reliable regional projections.

State-of-the-art climate models were used during the fifth phase of the Coupled Model Intercomparison Project (CMIP5, see Box 1) to provide information about the likely evolution of climate over the twenty-first century, with additional experiments to analyse the uncertainties inherent in these projections[1]. Evaluation of the CMIP5 historical (that is, twentieth-century; Supplementary Table 1) experiments shows that the simulation of modern climate has improved compared with simulations made as part of CMIP3. In particular, the current generation of models reproduces continental-scale surface patterns and long-term trends in temperature, and shows an improved ability to capture continental-scale precipitation patterns and reproduce the statistics of leading modes of climate variability such the North Atlantic Oscillation, El Niño–Southern Oscillation and Quasi-Biennial Oscillation[2]. Nevertheless, models that perform equally well for present-day climate produce very different responses to anthropogenic forcing (that is, in Representative Concentration Pathway (RCP) scenario simulations; Supplementary Table 1). The largest component of the uncertainty in model projections in the first part of the twenty-first century stems from differences between the response of individual models to the same forcing rather than internal variability or differences between the forcing scenarios themselves[3]. Differences between the climate forcing scenarios become more important by the end of the century[4], but intermodel differences still play a role in amplifying the scenario-related uncertainties and, indeed, still play a dominant role in explaining regional differences[3,5,6].

Past climates provide an opportunity to evaluate model performance outside the range of recent observed climate variability. Palaeoclimate simulations of the Last Glacial Maximum (LGM, 21 kyr ago) and mid-Holocene (6 kyr ago) were included in the CMIP5 simulations for this reason (Supplementary Table 1). Neither of these periods provides an analogue for the future evolution of climate — indeed, no past climate state provides a direct analogue for the future — but the change in forcing at the LGM was of similar magnitude (of the order of 3–6 W m^{-2}) to that projected for the next century[7], whereas the mid-Holocene provides an opportunity to evaluate simulations at a time of radically changed seasonality. Both periods have been foci for synthesis of palaeoclimate reconstructions[8,9]. Palaeoclimate evaluation using mid-Holocene and LGM climate reconstructions can help both to explain why the simulated mean response differs between models and to determine whether the upper or lower part of the range of response to future changes in forcing is inherently more likely to be realistic.

In a review of the potential of the CMIP5 palaeoclimate experiments to quantify uncertainties in model projections, six objectives were identified for the palaeoclimate simulations, including: (1) identification of robust features of past and future climates; (2) evaluation of model ability to simulate regional climate changes; (3) multi-parameter evaluation of overall model skill; (4) improvements in model performance between CMIP3 and CMIP5 in the simulation of large climate changes; (5) provision of well-founded constraints on climate sensitivity; and (6) evaluation of the role and magnitude of feedbacks. Analyses of many aspects of the CMIP5 palaeosimulations have now been completed and considerable progress has been made in addressing these six tasks. Our goal here is to synthesize and update these results, and to discuss their implications for the reliability (or otherwise) of future projections. We focus on the mid-Holocene and LGM simulations because these are the time periods for which there are global data sets of quantitative climate reconstructions[8,9] and because they have been examined with several generations of models[10–13], allowing us to assess the evolution of model performance. However, we also draw on

---

[1]Centre for Past Climate Change, School of Archaeology, Geography and Environmental Sciences (SAGES), University of Reading, Whiteknights, Reading RG6 6AH, UK. [2]Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia. [3]Department of Geography, University of Oregon, Eugene, Oregon 97403-1251, USA. [4]Laboratoire des Sciences du Climat et de l'Environnement/Institut Pierre Simon Laplace, unité mixte de recherches CEA-CNRS-UVSQ, Orme des Merisiers, bât 712, 91191 Gif sur Yvette Cedex, France. [5]Laboratoire de Meteorologie Dynamique (LMD/IPSL), CNRS/UPMC, Tour 45-55, 3eme etage, 4 place Jussieu, boite 99, 75252 Paris cedex 05, France. [6]BlueSkiesResearch.org.uk, The Old Chapel, Albert Hill, Settle BD24 9HE, UK. *email: s.p.harrison@reading.ac.uk

other Palaeoclimate Modelling Intercomparison Project (PMIP) experiments where appropriate.

Box 1 | The relationship between the CMIP and the PMIP.

The CMIP was set up in 1995 by the Working Group on Coupled Modelling of the World Climate Research Programme to provide a standard experimental protocol for studying climate changes using coupled atmosphere–ocean general circulation models. CMIP provides a community-based infrastructure in support of climate model diagnosis, validation and intercomparison. Initially, CMIP archived and analysed outputs from model 'control runs' in which climate forcing was constant, and idealized simulations in which atmospheric CO_2 concentration was increased either abruptly or in a transient fashion. Phase 3 of CMIP (CMIP3) included 'realistic' scenarios for historic, present and future climate. These simulations provided the basis for analyses underpinning the IPCC fourth assessment report. The current phase of CMIP (CMIP5)[1], which was initiated at the end of 2008, involves a large range of experiments for past, present and projected future climate, as well as more idealized experiments designed to explore model behaviour. Analysis of these simulations has already been used as input to the recent IPCC fifth assessment report and continues to be exploited for improved understanding of the mechanisms of climate change.

Palaeoclimate simulations were not included in CMIP prior to CMIP5. Nevertheless, the modelling community has been involved in a parallel effort to use past climate states to understand the mechanisms of climate change since the 1980s. These efforts have been coordinated by the PMIP[7,10,12,93]. The first round of PMIP intercomparisons (PMIP1) focused on atmospheric general circulation model simulations of the mid-Holocene and LGM, and was broadly parallel to the concurrent efforts of the Atmospheric Modelling Intercomparison Project[94]. PMIP2 focused on comparison of coupled atmosphere–ocean model simulations of the mid-Holocene and LGM. Although PMIP2 was broadly concurrent with CMIP3, and many of the same modelling groups were involved in both intercomparison projects, the palaeosimulations were generally run with either lower resolution or older versions of the models. The inclusion of palaeo-experiments in CMIP5 means that we now have simulations of past and future climate made with exactly the same version and at exactly the same resolution. As PMIP coordinates the analysis of the CMIP5 palaeo-experiments, these are often referred to as PMIP3 experiments (or PMIP3/CMIP5) experiments (although here, for simplicity, we refer to them as CMIP5 experiments).

The evaluation of the CMIP5 simulations is only one component of the ongoing work during PMIP3. PMIP3 is also running experiments for non-CMIP5 time periods and is coordinating the analysis and exploitation of transient simulations across intervals of rapid climate change in the past. PMIP also provides an umbrella for model intercomparison projects focusing on specific times in the past, such as the Pliocene Modelling Intercomparison Project[89], or on specific aspects of the palaeoclimate system, such as the Palaeo Carbon Modelling Intercomparison Project[95].

Robust features of past and future climates

There are several features of the temperature changes in future projections (and the more idealized 1pctCO2 and abrupt4xCO2 warming scenarios) that are remarkably consistent, including stronger warming over land than ocean (enhanced land–ocean contrast), stronger responses at higher than lower latitudes (latitudinal amplification) and differential responses in summer and winter leading to changes in seasonal contrast[3,14,15]. These large-scale temperature responses, which emerged in the first Intergovernmental Panel on Climate Change (IPCC) assessment 25 years ago, are present in palaeoclimate simulations as well, not only of the LGM and mid-Holocene[16–18], but also of other intervals such as the last interglaciation[19] and the mid-Pliocene[20,21]. The variations in response are proportional and nearly linear across simulations of both warm (1pctCO2, abrupt4xCO2) and cold (LGM) climate states (Fig. 1a,b), and the simulated magnitude of the relative changes between land and ocean, higher and lower latitudes, and summer and winter temperature is supported by historical and palaeoclimate observations[16] (Fig. 1c,d). Although this agreement between the simulated and observed responses is apparent on the large (hemispheric) scale, there is some evidence that it may not hold on a more regional level. It has been suggested, for example, that the CMIP5 models underestimate mid-Holocene warming in the eastern Canadian Arctic by >1 °C (ref. 22).

Several components of the surface energy balance are involved in the temperature responses in the LGM, historical and idealized warming scenario simulations, but surface downward clear-sky longwave radiation, which includes the effect of changes in CO_2, water vapour and atmospheric energy transport, is the most important component driving land–ocean contrast and high-latitude amplification in both warm (abrupt4xCO2) and cold (LGM) climates[23]. Surface albedo plays a significant but secondary role in promoting high-latitude amplification in both cold and warm climates[15], and in intensifying the land–ocean contrast in the warm climate case. Surface albedo has also been shown to contribute to latitudinal amplification in mid-Pliocene simulations[20]. Changes in seasonality are consistent in pattern but, in contrast to the relative simplicity of the mechanisms underpinning land–ocean contrast or latitudinal amplification, the genesis of the seasonality changes is different in warm and cold climates[23].

Precipitation increases as temperature increases, although at a rate that is consistently smaller than the rate of change in saturation vapour pressure, partly because of energetic constraints on evaporation and partly because of constraints in water availability over land[24,25]. Precipitation increases are characteristic of the CMIP5 future (RCP) and idealized warming (1pctCO2, abrupt4xCO2) simulations[26]. The scaling between the change in temperature and precipitation is remarkably consistent in palaeoclimate (LGM), historical and idealized warming (1pctCO2, abrupt4xCO2) simulations (Fig. 1e), both over land and ocean (Fig. 1f), and is also consistent with palaeoclimate and historical observations[27] (Fig. 1g). Analyses of precipitation changes in idealized warm (1pctCO2, abrupt4xCO2) climate states[27] also show other robust large-scale responses, including larger changes in precipitation per degree temperature change in extratropical than tropical land areas. Changes in tropical precipitation are greatest in areas that are currently wet, resulting in increased precipitation in warm climate states and decreased precipitation in cold climate states. The seasonality of precipitation in the tropics also changes in a consistent way, with increased seasonality in warm climate states and decreased seasonality in cold climate states[27]. All of these features are consistent with palaeoclimate and historical observations of large-scale precipitation changes.

Ability to simulate regional climate features

The regional response to changes in forcing has been a major focus in the evaluation of the CMIP5 mid-Holocene and LGM simulations. There can be three types of mismatch: cases where the models simulate the same robust response to a forcing but the response is

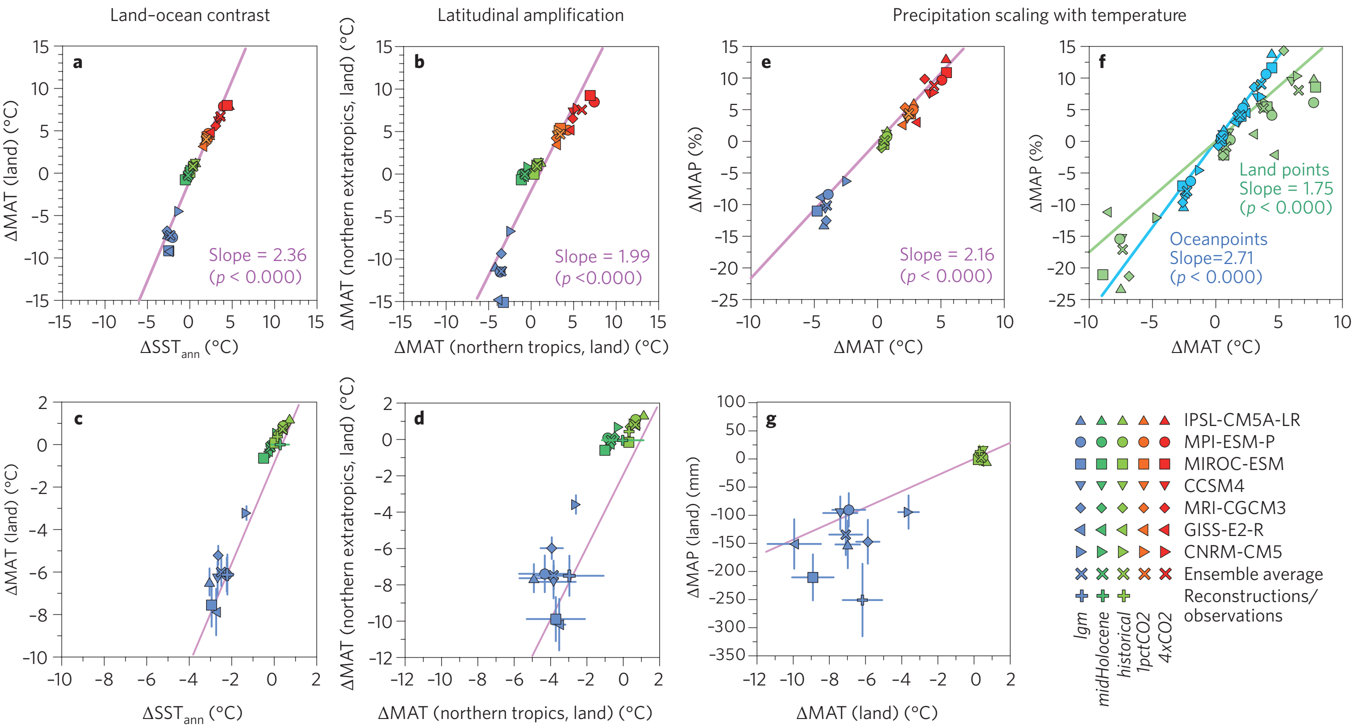

Figure 1 | Scatter plots showing temperature and precipitation changes in past, present and projected climates. The values shown are the simulated long-term mean differences (experiment minus piControl) for the seven models that have carried out all the experiments. **a,** Differences in the relative warming (or cooling) over global land and global ocean (land–ocean contrast) and **b**, over land in the northern extratropics and northern tropics (latitudinal amplification). $\mathrm{SST}_{\mathrm{ann}}$: annual sea surface temperature. **c,** Comparisons of the simulated changes in land–ocean contrast and **d**, latitudinal amplification for the twentieth-century (historical) and LGM with observed changes. The reduced major axis regression lines derived using all appropriate model grid cells are shown in magenta; the p-values test the null hypothesis that the slopes of the reduced major axis regression lines $= 1.0$. **e,** Percentage precipitation change relative to the change in global temperature and **f**, global temperature over land (green) and ocean (blue). The ordinary least-square regression with the intercept set at zero is shown in magenta; the p-values test the null hypothesis that the slope $= 0$. **g,** Comparisons of the simulated changes in precipitation scaling over land for the twentieth-century (historical) and LGM with observed changes. The ordinary least-square regression for absolute values of precipitation based on all model simulations is shown in magenta. In **c**, **d** and **g**, model output has been sampled only at the locations of respective observations. Bold crosses: area-weighted averages of twentieth-century observations and palaeoclimate reconstructions; finer lines: reconstruction uncertainties (standard deviation).

of the wrong magnitude; cases where the models simulate the same robust response to a forcing but the response is of the wrong sign; and cases where different models give different responses to the same forcing. Here we provide examples of each of these cases.

The insolation-induced amplification of Northern Hemisphere monsoons during the early to mid-Holocene provides the classic example of the use of model simulations to provide a mechanistic explanation of past climate changes[28]. Monsoon amplification, expressed through an increase in both the geographic area receiving monsoon rain and the overall amount of precipitation, is a feature of atmosphere-only simulations[10,29]. Simulations show that ocean feedback increases the length of the monsoon season and amplifies the magnitude of the overall response[30–32]. The CMIP5 mid-Holocene simulations show both a substantial expansion of the Asian (Pacific) and northern Africa monsoons with an increase in total precipitation[33–35], and a corresponding reduction in area and decrease in total precipitation in the Southern Hemisphere monsoons[36]. A previous study[7] showed that the PMIP Phase II (PMIP2) mid-Holocene simulations consistently underestimated the magnitude of change in the Northern Hemisphere monsoons. The CMIP5 mid-Holocene simulations show less amplification than the PMIP2 simulations over Asia, with strengthening of the meridional wind of only 32% compared with 40% in PMIP2 (ref. 33). The discrepancy between observed and CMIP5 simulated changes in the amount of mid-Holocene precipitation over northern Africa is at least 50% in the latitude band from 15–30° N (Fig. 2)[35]. Land-surface feedbacks, associated with the climate-induced change in vegetation cover and surface water storage, have been invoked as one way to reconcile these discrepancies[37–42]. Although some vegetation-enabled models show amplification of the northern Africa monsoon in the mid-Holocene[43,44], the PMIP2 models with dynamic vegetation did not produce greater amplification of any of the Northern Hemisphere monsoons during that time[7]. Furthermore, mid-Holocene simulations with the (CMIP5) CCSM4 model[45] show that vegetation feedback produces only very small changes in seasonal temperature and has no impact on precipitation over the Pacific monsoon region. The contrast between these PMIP2/CMIP5 results and earlier studies that prescribed vegetation changes or used simpler models suggests that significant improvements to the modelling of vegetation and its coupling with the atmosphere are required to address the role of land-surface feedbacks properly[46].

The intertropical convergence zone is located too far south in the Atlantic sector in most of the CMIP5 historical and pre-industrial control (piControl) simulations, reflecting a damped meridional temperature gradient that has been related to biases in radiation and heat fluxes[47]. Analyses of the West African monsoon in a subset of the PMIP2 mid-Holocene simulations show this bias affects the meridional temperature gradient and limits the northward movement of the intertropical convergence zone[48], which is also true for some of the CMIP5 mid-Holocene experiments. Differences in the amplification of the mid-Holocene monsoon over northern Africa in the CMIP5 experiments are not consistently related

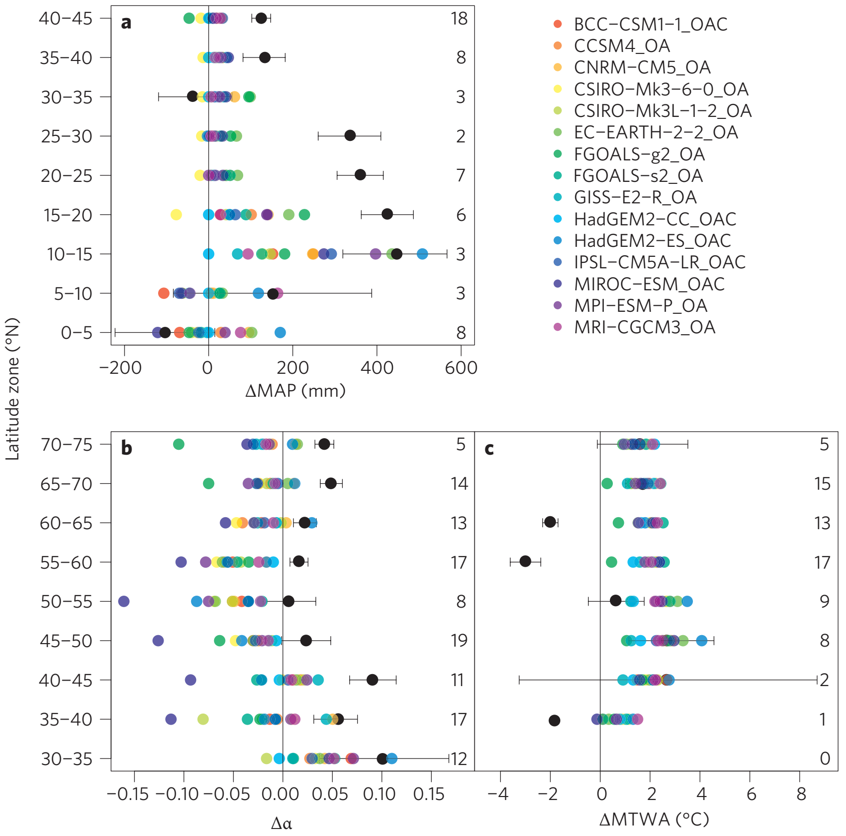

Figure 2 | Comparison of observed and simulated regional climate. **a**, Comparison of simulated and reconstructed mid-Holocene changes in mean annual precipitation for 5° latitude bands (longitude 20° W to 40° E) between 0 and 45° N across northern Africa and the circum-Mediterranean region. **b**, Ratio of actual to equilibrium evapotranspiration and **c**, mean temperature of the warmest month for 5° latitude bands between 30 and 80° N across Eurasia (longitude 60° to 180° E). The reconstructions are from the data set in ref. 9, which provides a climate reconstruction for a 2 × 2° grid cell based on averaging the individual site-based reconstructions within that grid cell. The mean and standard error of the mean of the grid cell reconstructions are shown in each latitude band. The model results are averages of model output sampled at the location of the grid cells with observations. The number of grid cells contributing to the comparison for each variable is shown on the plots.

to precipitation biases in the piControl simulations[35] because the mid-Holocene change in precipitation is driven by an increase in deep convection, so differences are largely linked to the way each model represents different convective regimes[48].

The extent of mid-continental drying in Eurasia during the mid-Holocene is another example of a persistent regional mismatch between models and observations[29,49,50]. The CMIP5 mid-Holocene simulations (Fig. 2) show drier conditions in Eurasia, particularly between 45° and 60° N, whereas observations systematically show that the mid-continental extratropics were wetter than today. At the same time, the models show a significant increase in summer temperature, whereas observations suggest cooler summers (Fig. 2). Temperature biases in the CMIP5 historical (twentieth-century) simulations are linked to systematic biases in evapotranspiration[51], with oversimulation of precipitation (and hence evapotranspiration) leading to cold temperature biases, and undersimulation of precipitation leading to warm biases. A similar mechanism seems to explain the mismatch in Eurasia in the mid-Holocene: the models do not produce a sufficient increase in regional precipitation and therefore underestimate evapotranspiration (and hence the ratio of actual to equilibrium evaporation, α) compared with observations, causing simulated summer temperatures of up to 4° warmer than observed.

The mid-Holocene climate of Europe provides a third example of a persistent mismatch between models and observations. CMIP5 mid-Holocene simulations show generalized warming over Europe in summer and fail to reproduce the observed summer cooling in southern Europe[52]. Winter temperature anomalies are not as consistent between models, but the CMIP5 mid-Holocene simulations do not show the strong winter warming in northern Europe shown by observations. These same discrepancies were present in mid-Holocene simulations with previous generations of models[53–55]. A previous study[52] suggests that these persistent discrepancies are related to the failure to simulate atmospheric circulation patterns correctly, specifically anticyclonic blocking in summer and increased dominance of the positive phase of the NAO in winter during the mid-Holocene. This study argues that atmospheric circulation patterns over Europe are also poorly simulated in modern (twentieth-century) simulations, which could explain why Europe is warming faster than projected[56].

Not all features of regional climates show a robust response to past changes in forcing, even when there is a consistent response in the future RCP simulations. For example, there is a consistent year-round reduction in the extent of Arctic sea ice in CMIP5 RCP simulations[3]. There is also a consistent reduction of summer sea-ice cover in response to increases in summer insolation in the CMIP5 mid-Holocene simulations[57], with the largest changes shown by those models with thicker sea ice in the piControl simulation. However, some models show increased and some decreased ice thickness in winter. An analysis of two models with similar sea-ice sensitivity in RCP scenario and abrupt4xCO2 simulations, but very different responses to mid-Holocene forcing, suggests that differences in the sign of the mid-Holocene changes in winter sea-ice extent may be related to cloud feedback. HADGEM2-ES shows a year-round decrease in sea-ice extent, whereas MIROC-ESM shows a smaller decrease in summer and a slight increase in winter. The difference in summer is attributed to differences in cloud cover between the two simulations[57]. Cloud-mediated differences in the summer response help to explain the different winter responses when the direct forcing is weak: a large reduction in summer sea-ice extent offsets the growth of sea ice in autumn and winter such that the overall extent of winter sea ice remains less than in the piControl simulation, whereas when the change in summer sea-ice extent is small it is insufficient to offset any orbitally induced winter increase. Similar analyses of the CMIP5 LGM simulations[58] confirm the relationship between sea-ice thickness in the piControl simulations and the magnitude of the change in summer sea-ice extent during the LGM, and also show that models have different responses to the change in forcing. The responses seem to be most different in the Southern Ocean, where there are also large discrepancies between simulated and observed sea-ice patterns.

The behaviour of the Southern Hemisphere westerly jet provides a second example of inconsistency in model simulations of the past that is not characteristic of future projections. The position of the Southern Hemisphere westerly jet is consistently shifted poleward in future simulations compared with the pre-industrial state because the tropospheric meridional temperature gradient is weakened. The CMIP5 models show diametrically opposed changes in the location of the Southern Hemisphere westerlies during the LGM, with half showing a equatorward shift and half showing a poleward shift in mean position compared with the piControl state[59,60]. The equatorward shift is consistent with the expected strengthening of the upper tropospheric temperature gradient. However, the models that unexpectedly simulate a poleward shift of the jet stream during the LGM compared with the pre-industrial state in fact show a strong LGM lower tropospheric cooling at high latitudes. This implicates different sensitivity to prescribed changes in the Antarctic ice sheet and to the simulated sea-ice extent in influencing the location of the Southern Hemisphere westerlies during the LGM[59]. Situations in which there is a consistent response in the future but different responses in the past thus provide an opportunity to explore model sensitivity to a wider diversity of feedbacks, such as the evolution of the ice sheets, than are currently included in simulations of the future.

A final example of inconsistent behaviour among models is provided by an analysis of hydroclimate in the tropical Pacific during the LGM[61]. This study provides an analysis of seven PMIP2 and five CMIP5 simulations, and shows contrasting responses of change in precipitation over the maritime continent (Southeast Asia, Indonesia, New Guinea and the Philippines): some models show widespread drying whereas others show a modest increase in precipitation. These different behaviours are, at least in part, due to simulated differences in the Walker circulation. The model (HadCM3) that most accurately reflects the pattern of the observed change in precipitation and ocean salinity, with strong and widespread drying over the maritime continent associated with freshening of the Arabian Sea and the western Pacific, is the sole model to produce a sufficiently weakened Walker circulation over the Indian Ocean. Only one of the CMIP5 models (MPI-ESM-P) shows weakening of the Walker circulation, but the change is not large enough to reproduce the observations.

Multi-parameter evaluation of model skill

Multi-parameter evaluation of simulations using global data sets is a routine measure of model performance under modern conditions[2]. Evaluations of the CMIP5 mid-Holocene and LGM simulations based on ten different seasonal or annual climate variables show that no model performs equally well for all variables[2,13]. In general, models are better at simulating mean (or median) values of any climate variable than at simulating the spatial variability or the geographic patterning in that variable. Although the CMIP5 models seem to have some skill in predicting mean annual temperature (MAT) and mean annual precipitation (MAP) during the LGM (Fig. 3), they have no skill in predicting summer temperature (mean temperature of the warmest month, MTWA) in the mid-Holocene, a result that confirms earlier analyses of the PMIP2 models[62]. Precipitation (as represented by MAP) is somewhat better simulated than temperature (as represented by MTWA) in the mid-Holocene but the reverse is true in the LGM simulations, where temperature (as represented by MAT) is better simulated than MAP (Fig. 3). Nevertheless, some models are better than others at capturing mid-Holocene and LGM climate change, and indeed perform better than the ensemble mean model[13]. The ensemble mean model usually provides the best estimate of the modern climate[63]. This may be because the ensemble mean filters out the impact of outliers in a collection of models that essentially have been 'tuned' to modern climate, but could also reflect the fact that the ensemble is too small and, physically speaking, the models are too closely related to characterize the underlying distribution effectively[64]. The fact that this is not the case in palaeo-experiments challenges the prevailing approach of using future projections to examine climate impacts — in which all available simulations are averaged to derive an ensemble response, with the spread of the experiments considered as a measure of uncertainty.

The spatial coverage of palaeoclimate reconstructions is more limited than modern observations and some key regions are undersampled[9]. Model evaluation has therefore often focused on data-rich regions and/or a limited number of climate parameters. Nevertheless, even taking account of the limited data and occasionally large uncertainties of palaeoclimate reconstructions, it is clear that the LGM and mid-Holocene simulations provide a rigorous test of model performance. Although there is reasonable agreement in the overall magnitude of the cooling during the LGM, primarily because the large-scale changes are dominated by land–ocean contrast and latitudinal gradients, the CMIP5 models show only limited skill in capturing sub-continental-scale patterns of temperature change. The mid-Holocene lacks the strong annual mean forcing that is present during the LGM and in future simulations, so the poor performance with respect to mean annual signals (Fig. 3) is therefore unsurprising. Our ability to evaluate these simulations is somewhat compromised by uncertainties about the seasonal attribution of sea-surface temperature reconstructions[65]. However, comparisons of continental seasonal climates confirm that the mid-Holocene still presents a challenge for the models. Some of the discrepancy between simulated and reconstructed mid-Holocene and LGM climate may reflect the simplified design of the experiments and, in particular, the omission of known feedbacks (for example, dust forcing during the LGM, land-surface characteristics in both simulations[18]). However, these feedbacks are also not included in future projections.

Improvement in ability to simulate climate change

The evidence of modest overall model skill during the mid-Holocene and LGM, and for substantial misrepresentation of past regional climates, clearly raises serious questions about state-of-the-art models. The current generation of models has been shown to be better at simulating some aspects of the modern climate[2]. Individual models are incorporating more complex treatments of key processes and feedbacks, and for individual models these improvements translate into better simulations of key aspects of past climate[66,67]. However, relative to previous generations of models, these developments apparently do not translate into an improved ability to simulate climate change. At the ensemble level, the differences between the CMIP5 simulations and earlier CMIP3/PMIP2 simulations are small and statistically unimportant, both for the past and for the future (Fig. 4). There is growing feeling that future analyses of climate change and its impacts should be based on cross-generational

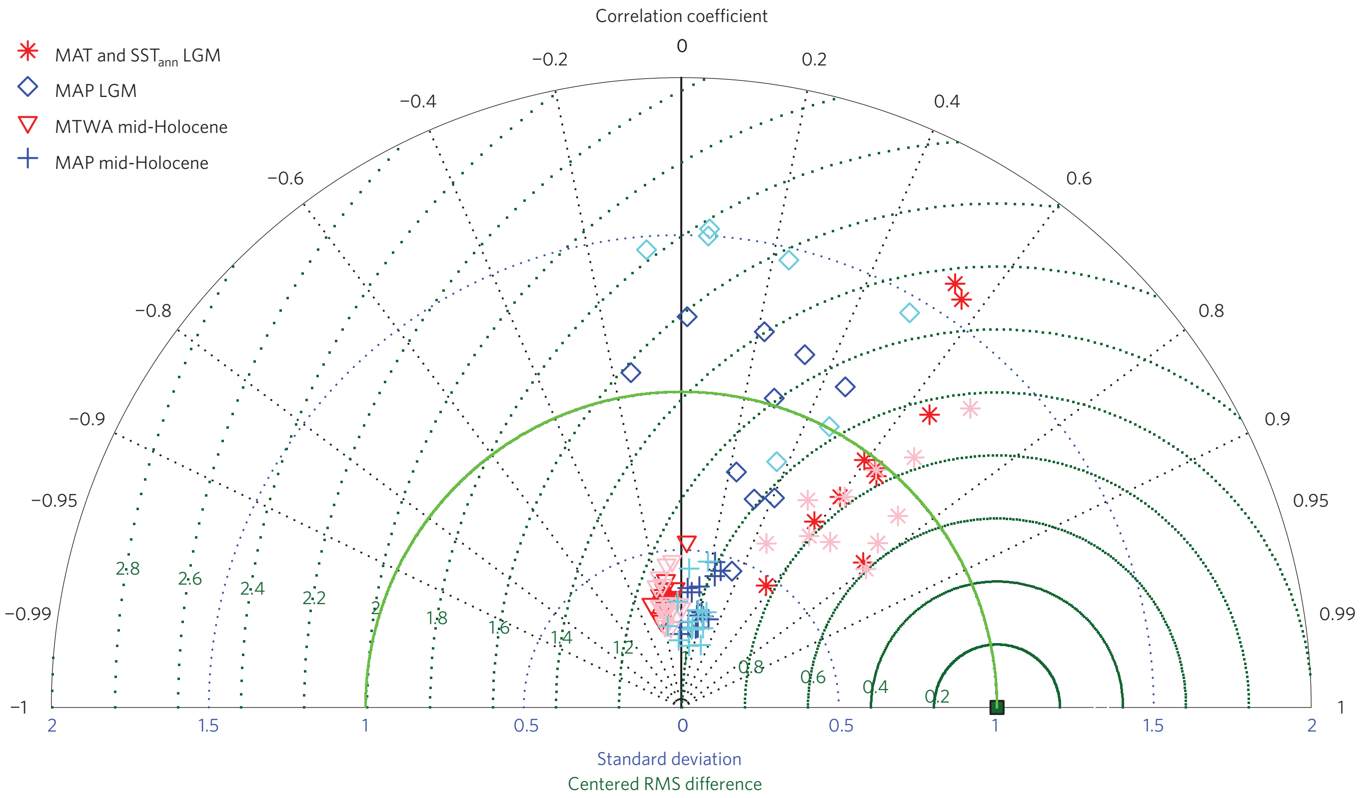

Figure 3 | Taylor diagram[90] for the LGM and mid-Holocene precipitation and temperature anomalies. The distance of any model point from the origin indicates standard deviation of field, the distance of any model point from the green reference point indicates the centred root mean square (RMS) difference between model and data. Pattern correlation between model and observations is given by the azimuthal coordinate. Temperature is represented by MAT over land and SST_{ann} at the LGM, and by MTWA for the mid-Holocene where the change in forcing is seasonal. Precipitation is always represented by MAP. Model statistics are corrected to account for observational uncertainties, by subtracting the estimated contributions made by observational errors as in ref. 62. Models from the CMIP5 ensemble are in red (temperature) and blue (precipitation), whereas models from the PMIP2 ensemble are in pink and pale blue, respectively.

ensembles of model output constrained by observations[68,69], and palaeo-evaluations support this approach, not only because there seems to be little improvement between different generations of models, but also because some models are better than others at reproducing the magnitude and patterns of large climate changes.

Providing well-founded climate sensitivity constraints

Differences in climate sensitivity, conventionally defined as the change in global average temperature for a doubling of CO_2, are realized as intermodel differences in the projections of future global warming. It has proved difficult to evaluate model sensitivity using historical observations, and this has motivated attempts to use past climate states as a constraint[70]. The LGM has been a focus for such attempts because of the large difference in climate from present[71–75]. Many (though not all) energy-balance mechanisms operate similarly in simulations of the LGM and of future (warm) climates across the ensemble of CMIP5 models[23], although there is asymmetry in the strengths of different feedbacks[76]. One study[74] found a significant correlation in the previous generation of climate models (PMIP2) between tropical temperature change during the LGM and equilibrium sensitivity, but this relationship is not evident in the CMIP5 LGM simulations. We have re-examined this finding by combining the CMIP5 and PMIP2 ensembles (following the approach suggested in ref. 69), taking the mean of the outputs where more than one integration was carried out by closely related models. This gives a total of 11 simulations and a weak correlation between tropical temperature during the LGM and equilibrium climate sensitivity, which is barely significant at the 90% level. This provides an estimate of climate sensitivity in the range of 1.4–4.4 °C, but the tenuous nature of the correlation cannot be ignored when assessing the credibility of this result. The presence of strong and consistent spatial patterns in temperature changes, as evidenced by land–ocean contrast and high-latitude amplification, suggest that tropical temperature may be an insufficient constraint on climate sensitivity. Another study[13] adopted an alternative approach, by comparing the CMIP5 and PMIP2 model ensemble with all available LGM temperature reconstructions and estimating climate sensitivity from the regression as the temperature at which global bias is zero. They obtained an estimate of 2.7 °C, but again argued that the result was only barely significant ($p = 0.12$) even after the removal of a marked warm-bias outlier. Thus, although the LGM provides a useful check on model performance, it remains a challenge to generate well-founded quantitative constraints on climate sensitivity from these simulations.

Palaeosimulations and future projections

Evaluation of the CMIP5 palaeosimulations demonstrates the value of including past climate states as targets for model intercomparison. Systematic examination of features that are characteristic of future climate simulations in palaeoclimate experiments and palaeoclimate reconstructions provides an opportunity to determine whether these features are robust characteristics of the climate system, and whether they are features of the actual response of the climate system to changes in forcing rather than model artefacts[7,16,18]. The broad-scale temperature and precipitation responses seen in future simulations are present in palaeosimulations and correctly represented in both LGM and historical simulations. This gives us confidence that the projected changes in land–sea temperature contrast, high-latitude amplification, temperature seasonality, the scaling of precipitation with temperature and the differential precipitation–temperature scaling over land and ocean are reliable. Similarly, the fact that models produce large-scale changes in climate consistent with palaeo-reconstructions for multiple different climate states enhances our confidence in the simulated changes shown in future projections. The palaeo-record has the ability to discriminate between models where they show differences in the response to forcing, and again this provides a way of determining which models are more reliable.

Nevertheless, the modest overall skill of the CMIP5 models for the mid-Holocene and LGM shows the limitations of the current generation of models. Specifically, the models are unable to reproduce the magnitude of changes in regional climates, even when taking into account the uncertainties inherent in the palaeo-reconstructions. The amplification of the Northern Hemisphere monsoons is a robust feature of future (RCP) climate simulations[3,4]. Although the underlying cause differs (increased greenhouse gases rather than a change in insolation), the antecedent condition of continental warming in the subtropics leading to increased land–ocean contrast is the same in future and mid-Holocene simulations. Thus, the fact that models persistently underestimate the magnitude of regional precipitation changes over Africa and Asia during the mid-Holocene suggests that the future predictions could be similarly affected. Given that these monsoon systems influence the livelihood of more than half of the world's population, this is a situation that needs to be rectified. Addressing the causes of persistent mismatches, both for the monsoon regions and for other regions identified by palaeo-comparison, should be a research priority.

It is possible that discrepancies between simulated and observed regional climates in the mid-Holocene and LGM are due to uncertainties in the specification of prescribed boundary conditions or the failure to include additional potential forcings[7]. Simulated LGM climates are indeed sensitive to the form of the prescribed ice sheet[7,77]. However, the latest reconstructions of the size and form of the LGM ice sheets are more similar to one another than to previous attempts at reconstruction. Furthermore, the impact of uncertainties in ice-sheet prescription is small and highly localized compared with the other, well-constrained forcings. Similarly, some of the model simulations prescribe vegetation to be the same as present in both the mid-Holocene and the LGM. However, inclusion of dynamic vegetation does not seem to improve the simulation of mid-Holocene regional climates. Furthermore, Earth system models do not seem to perform better overall than models that do not include a dynamic carbon cycle and/or dynamic vegetation. Inclusion of dust forcing has been shown to improve the simulations of LGM climate[78], for example, but again the impact of dust is small compared with the impact of the changes in the ice sheet or atmospheric composition[78,79]. Thus, although uncertainties in the experimental protocol could contribute somewhat to the poor performance of the CMIP5 models, the large discrepancies between observations and simulations cannot be explained away by invoking the experimental design.

It is of concern that the current generation of climate models does not perform better overall than previous generations of models, in terms of either modern climate or palaeoclimate changes[13,69]. On the positive side, this opens up the possibility of using cross-generational ensembles for projections of climate and climate impacts, which would provide a larger ensemble and more robust measurements of uncertainties. However, there is a need to screen the models used in constructing such ensembles, because palaeo-evaluation shows that some models are consistently better than others at reproducing the magnitude and patterns of large climate changes.

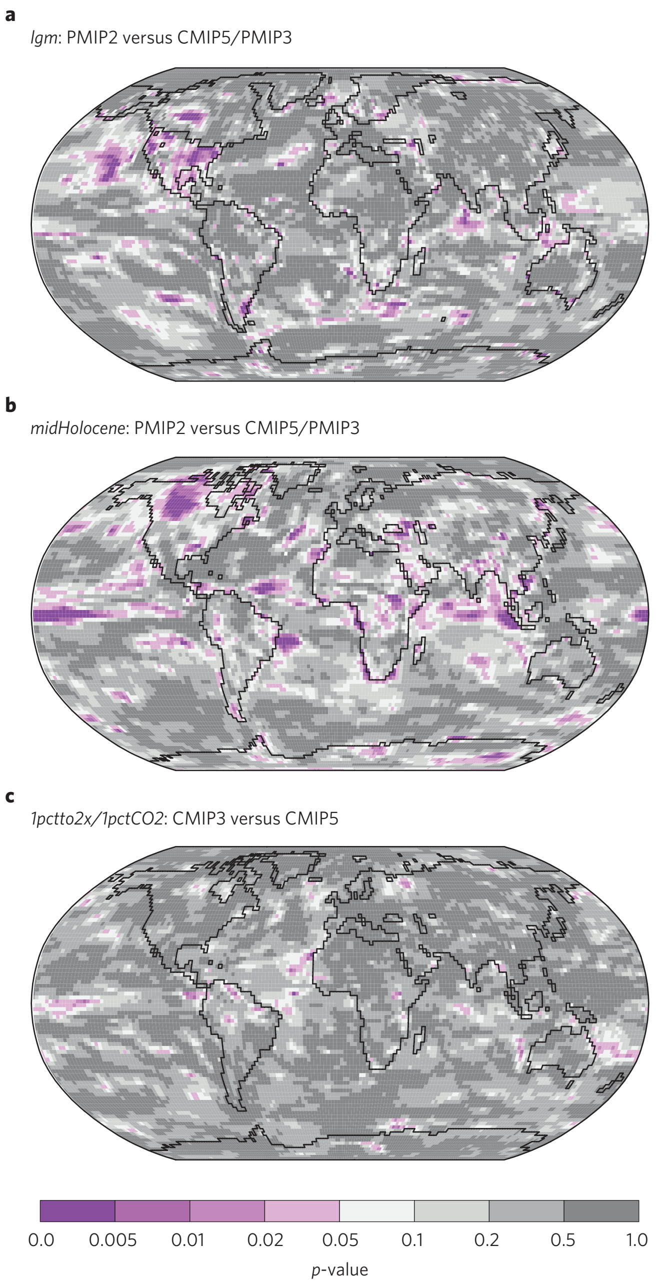

Figure 4 | Maps of the p-values of Hotelling's T2 test[91] comparing the CMIP3 plus PMIP2 versus CMIP5 ensembles. The plots show the *p*-values for the test of the hypothesis of equality of the (multivariate) ensemble means of MAT, mean temperature of the coldest month, MTWA and MAP for the LGM (*lgm*), mid-Holocene (*midHolocene*) and the 1pctCO2 simulations in CMIP3 and CMIP5 (*1pctto2x* and *1pctCO2*, respectively). The number of significant statistics (that is, $p < 0.05$, shown in pink) do not exceed that expected by chance[92]. A previous study[13] has shown that the results obtained using conventional meteorological variables are virtually identical.

Palaeo-simulations have not delivered on the promise to provide a well-founded additional constraint on climate sensitivity. This is partly because of the limited size of the ensemble, even when including LGM experiments with the previous CMIP3/PMIP2 generation of models. However, a second issue is associated with the limited amount of palaeoclimate data, and particularly quantitative reconstructions, from the Southern Hemisphere. It is difficult to constrain a global average based on unevenly distributed data points. The continued expansion of palaeoclimate data sets will also allow evaluation of other regional climate changes. However, the robust nature of the spatial variations of climate change in the past (and future) calls into question whether a focus on global average responses is sensible.

Much of our knowledge about regional climate changes is based on qualitative inferences from geologic, biological or archaeologic records, which provide a more detailed picture of the geographic areas affected than currently possible using quantitative climate reconstructions. Qualitative records are useful because they confirm that the more limited quantitative information is realistic. For example, although there are only 58 grid cells with quantitative reconstructions of mid-Holocene MAP for northern Africa, we are confident that the reconstructed increase in monsoon precipitation is reasonable because of the extensive information on the widespread occurrence of lakes[80,81], profound changes in vegetation cover[82,83] and abundant human settlements[84,85]. However, qualitative data of this sort cannot be used explicitly in model evaluation. Although some of these records could be used for quantitative reconstruction using statistical techniques, generally the exploitation of most of these data relies on the use of forward models, for example, of vegetation[86] or lake water balance[87,88]. We suggest that increased emphasis on climate reconstruction and greater exploitation of forward modelling is urgently required to improve climate model evaluation.

The CMIP community is currently defining the suite of experiments that will constitute the basis for the next IPCC assessment report. CMIP5 was the first explicit inclusion of palaeo-experiments in the CMIP suite of simulations but they have already shown their usefulness. We urge all members of the CMIP community to run palaeosimulations and to use them in model diagnosis. Demonstrating which features of the simulated climate change are likely to be realistic, and which are not, will do much to increase confidence in future projections.

Received 24 December 2014; accepted 10 April 2015;

References

1. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93,** 485–498 (2012).
2. Flato, G. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 741–866 (IPCC, Cambridge Univ. Press, 2013).
3. Kirtman, B. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 953–1028 (IPCC, Cambridge Univ. Press, 2013).
4. Collins, M. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 1029–1136 (IPCC, Cambridge Univ. Press, 2013).
5. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90,** 1095–1107 (2009).
6. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in projections of regional precipitation change. *Clim. Dynam.* **37,** 407–418 (2011).
7. Braconnot, P. *et al.* Evaluation of climate models using palaeoclimatic data. *Nature Clim. Change* **2,** 417–424 (2012).
8. MARGO Project Members Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geosci.* **2,** 127–132 (2009).
9. Bartlein, P. J. *et al.* Pollen-based continental climate reconstructions at 6 and 21 ka: A global synthesis. *Clim. Dynam.* **37,** 775–802 (2011).
10. Joussaume, S. *et al.* Monsoon changes for 6000 years ago: Results of 18 simulations from the Paleoclimate Modelling Intercomparison Project (PMIP). *Geophys. Res. Lett.* **26,** 859–862 (1999).
11. Pinot, S. *et al.* Tropical palaeoclimates at the Last Glacial Maximum: Comparison of Paleoclimate Modelling Intercomparison Project (PMIP) simulations and paleodata. *Clim. Dynam.* **15,** 857–874 (1999).
12. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the mid-Holocene and Last Glacial Maximum—Part 1: Experiments and large-scale features. *Clim. Past* **3,** 261–277 (2007).
13. Harrison, S. P. *et al.* Model benchmarking with glacial and mid-Holocene climates. *Clim. Dynam.* **43,** 671–688 (2014).
14. Joshi, M. M., Lambert, F. H. & Webb, M. J. An explanation for the difference between twentieth and twenty-first century land–sea warming ratio in climate models. *Clim. Dynam.* **41,** 1853–1869 (2013).
15. Pithan, F. & Mauritsen, T. Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geosci.* **7,** 181–184 (2014).
16. Izumi, K., Bartlein, P. J. & Harrison, S. P. Consistent behaviour of the climate system in response to past and future forcing. *Geophys. Res. Lett.* **40,** 1–7 (2013).
17. Hargreaves, J. C. & Annan, J. D. Can we trust climate models? *WIREs Clim. Change* **5**, 435–440 (2014).
18. Schmidt, G. A. *et al.* Using paleo-climate comparisons to constrain future projections in CMIP5. *Clim. Past* **10,** 221–250 (2014).
19. Lunt, D. J. *et al.* A multi-model assessment of last interglacial temperatures. *Clim. Past* **9,** 699–717 (2013).
20. Hill, D. J. *et al.* Evaluating the dominant components of warming in Pliocene climate simulations. *Clim. Past* **10,** 79–90 (2014).
21. Salzmann, U. *et al.* Challenges in quantifying Pliocene terrestrial warming revealed by data–model discord. *Nature Clim. Change* **3**, 969–974 (2013).
22. Miller, G. H., Lehman, S. J., Refsnider, K. A., Southon, J. R. & Zhong, Y. Unprecedented recent summer warmth in Arctic Canada. *Geophys. Res. Lett.* **40,** 5745–5751 (2013).
23. Izumi, K., Bartlein, P. J. & Harrison, S. P. Energy-balance mechanisms underlying consistent large-scale temperature responses in warm and cold climates. *Clim. Dynam.* **44**, 3111–3127 (2015).
24. Allan, R. P. Examination of relationships between clear-sky longwave radiation and aspects of the atmospheric hydrological cycle in climate models, reanalysis, and observations. *J. Climate* **22,** 3127–3145 (2009).
25. Trenberth, K. E. & Shea, D. J. Relationships between precipitation and surface temperature. *Geophys. Res. Lett.* **32,** L14703 (2005).
26. Lau, W. K. M., Wu, H-T. & Kim, K-M. A canonical response of precipitation characteristics to global warming from CMIP5 models. *Geophys. Res. Lett.* **40,** 3163–3169 (2013).
27. Li, G., Harrison, S. P., Bartlein, P. J., Izumi, K. & Prentice, I. C. Precipitation scaling with temperature in warm and cold climates: An analysis of CMIP5 simulations. *Geophs. Res. Lett.* **40,** 4018–4024 (2013).
28. Kutzbach, J. E. & Street-Perrott, F. A. Milankovitch forcing of fluctuations in the level of tropical lakes from 18 to 0 kyr BP. *Nature* **317,** 130–134 (1985).
29. Wohlfahrt, J., Harrison, S. P. & Braconnot, P. Synergistic feedbacks between ocean and vegetation on mid- and high-latitude climates during the mid-Holocene. *Clim. Dynam.* **22,** 223–238 (2004).
30. Marzin, C. & Braconnot, P. The role of the ocean feedback on Asian and African monsoon variations at 6 kyr and 9.5 kyr BP. *C. R. Geosci.* **341,** 643–655 (2009).
31. Wang, T., Wang, H. J. & Jiang, D. B. Mid-Holocene East Asian summer climate as simulated by the PMIP2 models. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **288,** 93–102 (2010).
32. Zhao, Y. & Harrison, S. P. Mid-Holocene monsoons: A multi-model analysis of the inter-hemispheric differences in the responses to orbital forcing and ocean feedbacks. *Clim. Dynam.* **39,** 1457–1487 (2012).
33. Jiang, D., Lang, X., Tian, Z. & Ju, L. Mid-Holocene East Asian summer monsoon strengthening: Insights from Paleoclimate Modeling Intercomparison Project (PMIP) simulations. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **369,** 422–429 (2013).
34. Jiang, D., Tian, Z. & Land, X. Mid-Holocene net precipitation changes over China: Model–data comparison. *Quat. Sci. Rev.* **82,** 104–120 (2013).
35. Perez-Sanz, A., Li, G., Gonzalez, P. & Harrison, S. P. Evaluation of seasonal climates of northern Africa and the Mediterranean in the CMIP5 simulations. *Clim. Past* **10,** 551–568 (2014).
36. Prado, L. F., Wainer, I. & Chiessi, C. M. Mid-Holocene PMIP3/CMIP5 model results: Intercomparison for the South American monsoon system. *Holocene* **21,** 1915–1920 (2013).
37. Kutzbach, J. E., Bonan, G. B., Foley, J. A. & Harrison, S. P. Vegetation and soils feedbacks on the response of the African monsoon to orbital forcing in the early to middle Holocene. *Nature* **384,** 623–626 (1996).
38. Claussen, M. & Gayler, V. The greening of the Sahara during the mid-Holocene: Results of an interactive atmosphere–biome model. *Glob. Ecol. Biogeogr. Lett.* **6,** 369–377 (1997).
39. Broström, A. *et al.* Land surface feedbacks and palaeomonsoons in northern Africa. *Geophys. Res. Lett.* **25,** 3615–3618 (1998).
40. Braconnot, P., Joussaume, S., Marti, O. & de Noblet, N. Synergistic feedbacks from ocean and vegetation on the African monsoon response to mid-Holocene insolation. *Geophys. Res. Lett.* **26,** 2481–2484 (1999).
41. Otto, J., Raddatz, T., Claussen, M., Brovkin, V. & Gayler, V. Separation of atmosphere–ocean–vegetation feedbacks and synergies for mid-Holocene climate. *Glob. Biogeochem. Cycles* **23,** L09701 (2009).
42. Claussen, M., Bathiany, S., Brovkin, V. & Kleinen, T. Simulated climate–vegetation interaction in semi-arid regions affected by plant diversity. *Nature Geosci.* **6,** 954–958 (2013).
43. Levis, S., Bonan, G. B. & Bonfils, C. Soil feedback drives the mid-Holocene North African monsoon northward in fully coupled CCSM2 simulations with a dynamic vegetation model. *Clim. Dynam.* **23,** 791–802 (2004).

44. Wang, Y. *et al*. Detecting vegetation–precipitation feedbacks in mid-Holocene North Africa from two climate models. *Clim. Past* **4,** 59–67 (2008).
45. Tian, Z. & Jiang, D. Mid-Holocene ocean and vegetation feedbacks over East Asia. *Clim. Past* **9,** 2153–2171 (2013).
46. Prentice, I. C., Liang, X., Medlyn, B. & Wang, Y. Reliable, robust and realistic: The three R's of next-generation land-surface modelling. *Atmos. Chem. Phys. Discuss.* **14,** 24811–24861 (2014).
47. Roehrig, R., Bouniol, D., Guichard, F., Hourdin, F. & Redelsperger, J-L. The present and future of the West African monsoon: A process-oriented assessment of CMIP5 simulations along the AMMA transect. *J. Climate* **26,** 6471–6505 (2013).
48. Zheng, W. & Braconnot, P. Characterization of model spread in PMIP2 mid-Holocene simulations of the African monsoon. *J. Climate* **26,** 1192–1210 (2013).
49. Yu, G. & Harrison, S. P. An evaluation of the simulated water balance of Eurasia and northern Africa at 6000 yr BP using lake status data. *Clim. Dynam.* **12,** 723–735 (1996).
50. Wohlfahrt, J. *et al*. Evaluation of coupled ocean–atmosphere simulations of Northern Hemisphere extratropical climates in the mid-Holocene. *Clim. Dynam.* **31**, 871–890 (2008).
51. Mueller, B. & Seneviratne, S. I. Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.* **41**, 128–134 (2014).
52. Mauri, A., Davis, B. A. S., Collins, P. M. & Kaplan, J. O. The influence of atmospheric circulation on the mid-Holocene climate of Europe: A data–model comparison. *Clim. Past* **10,** 1925–1938 (2014).
53. Masson, V. *et al*. Mid-Holocene climate in Europe: What can we infer from PMIP model data comparisons? *Clim. Dynam*. **15,** 163–182 (1999).
54. Brewer, S., Guiot, J. & Torre, F. Mid-Holocene climate change in Europe: A data–model comparison. *Clim. Past* **3,** 499–512 (2007).
55. Davis, B. A. S. & Brewer, S. Orbital forcing and the role of the latitudinal temperature/insolation gradient. *Clim. Dynam*. **32,** 143–165 (2009).
56. van Oldenborgh, G. J. *et al*. Western Europe is warming much faster than expected. *Clim. Past* **5,** 1–12 (2009).
57. Berger, A., Brandefelt, J. & 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. *Clim. Past* **9,** 969–982 (2013).
58. Goosse, H., Roche, D. M., Mairesse, A. & Berger, M. Modelling past sea ice changes. *Quat. Sci. Rev*. **79,** 191–206 (2013).
59. Chavaillaz, Y., Codron, F. & Kageyama, M. Southern westerlies in LGM and future (RCP4.5) climates. *Clim. Past* **9,** 517–524 (2013).
60. Rojas, M. Sensitivity of Southern Hemisphere circulation to LGM and $4\times CO_2$ climates. *Geophys. Res. Lett*. **40,** 965–970 (2013).
61. DiNezio, P. N. & Tierney, J. E. The effect of sea level on glacial Indo-Pacific climate. *Nature Geosci.* **6,** 485–491 (2013).
62. Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A. & Abe-Ouchi, A. Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene. *Clim. Past* **9**, 811–823 (2013).
63. Gleckler, P. J., Taylor, K. E. & Doutriaux, C, Performance metrics for climate models. *J. Geophys. Res.* **113**, D06104 (2008).
64. Annan, J. D. & Hargreaves, J. C. Understanding the CMIP3 multimodel ensemble. *J. Climate* **24,** 4529–4538 (2011).
65. Hessler, I. *et al*. Implication of methodological uncertainties for mid-Holocene sea surface temperature reconstructions. *Clim. Past* **10,** 2237–2252 (2014).
66. Kageyama, M. *et al*. Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model – Part I: Comparing IPSL CM5A to IPSL CM4. *Clim. Dynam*. **40,** 2447–2468 (2013).
67. Kageyama, M. *et al*. Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model: Part II: Model–data comparisons. *Clim. Dynam.* **40,** 2469–2495 (2013).
68. Knutti, R. The end of model democracy? *Clim. Change* **102,** 395–404 (2010).
69. Rauser, F., Gleckler, P. & Marotzke, J. Rethinking the default construction of multi-model climate ensembles. *Bull. Am. Meteorol. Soc.* http://dx.doi.org/10.1175/BAMS-D-13-00181.1 (2014).
70. PALAEOSENS project members Making sense of palaeoclimate sensitivity. *Nature* **491,** 683–691 (2012).
71. Annan, J. D., Hargreaves, J. C., Ohgaito, R., Abe-Ouchi, A. & Emori, S. Efficiently constraining climate sensitivity with ensembles of paleoclimate simulations. *SOLA* **1,** 181–184 (2005).
72. Crucifix, M. Does the Last Glacial Maximum constrain climate sensitivity? *Geophys. Res. Lett.* **33**, L18701 (2006).
73. Schneider von Deimling, T., Held, H., Ganolpolski, A. & Rahmstorf, S. Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim. Dynam.* **27,** 149–163 (2006).
74. Hargreaves, J. C., Annan, J. D., Yoshimori, M. & Abe-Ouchi, A. Can the Last Glacial Maximum constrain climate sensitivity? *Geophys. Res. Lett.* **39**, L24702 (2012).
75. Tripati, A. K. *et al*. Modern and glacial tropical snowlines controlled by sea surface temperature and atmospheric mixing. *Nature Geosci.* **7,** 205–209 (2014).
76. Yoshimori, M., Hargreaves, J. C., Annan, J. D., Yokohata, T. & Abe-Ouchi, A. Dependency of feedbacks on forcing and climate state in physics parameter ensembles. *J. Climate* **24,** 6440–6455 (2011).
77. Otto-Bliesner, B. L. *et al*. Last Glacial Maximum and Holocene climate in CCSM3. *J. Climate* **19,** 2526–2544 (2006).
78. Schmittner, A. *et al*. Climate sensitivity estimated from temperature reconstructions of the Last Glacial Maximum. *Science* **334,** 1385–1388 (2011).
79. Claquin, T. *et al.* Radiative forcing effect of ice-age dust. *Clim. Dynam.* **20,** 193–202 (2003).
80. Hoelzmann, P. *et al.* Mid-Holocene land-surface conditions in northern Africa and the Arabian peninsula: A data set for the analysis of biogeophysical feedbacks in the climate system. *Glob. Biogeochem. Cycles* **12,** 35–51 (1998).
81. Lézine, A-M., Hély, C., Grenier, C., Braconnot, P. & Krinner, G. Sahara and Sahel vulnerability to climate changes, lessons from Holocene hydrological data. *Quat. Sci. Rev.* **30,** 3001– 3012 (2011).
82. Elenga, H. *et al*. Pollen-based reconstruction for Southern Europe and Africa 18,000 years ago. *J. Biogeogr*. **27,** 621–634 (2000).
83. Watrin, J., Lézine, A-M. & Hély, C. Plant migration and ecosystems at the time of the "green Sahara". *C. R. Geosci.* **341,** 656–670 (2009).
84. Kuper, R. & Kropelin, S. Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* **313,** 803–807 (2006).
85. Manning, K. & Timpson, A. The demographic response to Holocene climate change in the Sahara. *Quat. Sci. Rev.* **101,** 28–35 (2014).
86. Prentice, I. C., Harrison, S. P. & Bartlein, P. J. Global vegetation and terrestrial carbon cycle changes after the last ice age. *New Phytol*. **189,** 988–998 (2011).
87. Coe, M. T. & Harrison, S. P. The water balance of northern Africa during the mid-Holocene: An evaluation of the 6ka BP PMIP experiments. *Clim. Dynam*. **19,** 155–166 (2002).
88. Ward, P. J., Aerts, J. C. J. H., de Moel, H. & Renssen, H. Verification of a coupled climate–hydrological model against Holocene palaeohydrological records. *Glob. Planet. Change* **57,** 283–300 (2007).
89. Haywood, A. M. *et al*. Pliocene Model Intercomparison Project (PlioMIP): Experimental design and boundary conditions (Experiment 1). *Geosci. Model Dev.* **3,** 227–242 (2010).
90. Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106,** 7183–7192 (2001).
91. Wilks, D. S. *Statistical Methods in the Atmospheric Sciences* 3rd edn (International Geophysics Series 100, Academic, 2011).
92. Wilks, D. S. On "field significance" and the false discovery rate. *J. Appl. Meteorol. Climatol.* **45,** 1181–1189 (2006).
93. Joussaume, S. & Taylor, K. E. in *Proc. 1st Int. AMIP Sci. Conf.* 425–430 (WCRP Series Report 92, WMO, 1995).
94. Gates, W. L. AMIP: The Atmospheric Model Intercomparison Project. *Bull. Am. Meteorol. Soc.* **73,** 1962–1970 (1992).
95. Abe-Ouchi, A. & Harrison, S. P. Constraining the carbon-cycle feedback using palaeodata: The PalaeoCarbon Modelling Intercomparison Project. *Eos* **90,** 140 (2009).

Acknowledgements

This paper is a contribution to the ongoing work on the PMIP. We thank all of the modelling groups who have contributed to the CMIP5 archive. We acknowledge financial support from the Centre for Past Climate Change, University of Reading. G.L. was supported by an international postgraduate research scholarship at Macquarie University. P.J.B. and K.I. were supported by the US National Science Foundation paleoclimatology programme. M.K., P.B. and K.I. acknowledge financial support from Labex L-IPSL.

Author contributions

S.P.H. planned the paper and was responsible for drafting the text; all authors were involved in analysis and interpretation of the data, and contributed to the final version.

Additional information

Reprints and permissions information is available online at www.nature.com/reprints. Correspondence should be addressed to S.P.H.

Competing financial interests

The authors declare no competing financial interests.