What have we learnt from palaeoclimate simulations?

Journal of Quaternary Science

What have we learnt from palaeoclimate simulations?

Journal:	*Journal of Quaternary Science*
Manuscript ID	JQS-15-0098.R2
Wiley - Manuscript type:	Special Issue Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Harrison, Sandy; University of Reading, Centre for Past Climate Change and School of Archaeology, Geography and Environmental Sciences (SAGES), Bartlein, Patrick; University of Oregon, Department of Geography Prentice, I; Imperial College, Department of Life Sciences
Keywords:	palaeoclimate modelling, palaeoenvironmental data synthesis, climate reconstruction, forward modelling, CMIP5

What have we learnt from palaeoclimate simulations?

Sandy P. Harrison[1], Patrick J. Bartlein[2], I. Colin Prentice[3]

1: Centre for Past Climate Change *and* School of Archaeology, Geography and Environmental Sciences (SAGES), University of Reading, Whiteknights, Reading, RG6 6AH, UK

2: Department of Geography, University of Oregon, Eugene, Oregon 97403-1251, USA.

3: AXA Chair Programme in Climate and Biosphere Impacts, Grand Challenges in Ecosystems and the Environment and Grantham Institute – Climate Change and the Environment, Department of Life Sciences, Silwood Park Campus, Imperial College, Ascot, UK

Journal of Quaternary Science

Corresponding author: Sandy P. Harrison, School of Archaeology, Geography and Environmental Sciences (SAGES), University of Reading, Whiteknights, Reading, RG6 6AH, UK.
Email: s.p.harrison@reading.ac.uk

Abstract

There has been a gradual evolution in the way that palaeoclimate modeling and palaeoenvironmental data are used together to understand how the Earth System works, from an initial and largely descriptive phase through explicit hypothesis testing to diagnosis of underlying mechanisms. Analyses of past climate states are now regarded as integral to the evaluation of climate models, and have become part of the toolkit used to assess the likely realism of future projections. Palaeoclimate assessment has demonstrated that changes in large-scale features of climate that are governed by the energy and water balance show consistent responses to changes in forcing in different climate states, and these consistent responses are reproduced by climate models. However, state-of-the-art models are still largely unable to reproduce observed changes in climate at a regional scale reliably. While palaeoclimate analyses of state-of-the-art climate models suggest an urgent need for model improvement, much work is also needed on extending and improving palaeoclimate reconstructions and quantifying and reducing both numerical and interpretative uncertainties.

Keywords: palaeoclimate modelling, palaeoenvironmental data synthesis, climate reconstruction, forward modelling, CMIP5

Introduction

Climate has varied continuously through Earth's history. There are several styles of climate variability, associated with different drivers and operating on characteristic time scales. For example, there are periodic climate changes, resulting from astronomical or 'orbital' forcing on seasonal and multi-millennial timescales (Berger, 1978). Examples of progressive changes include the long-term cooling through the Cenozoic in response to changes in land-sea configuration and atmospheric composition (Zachos et al., 2001; Fletcher et al., 2007), or the cooling trend of the last two millennia caused by orbitally-driven changes in incoming solar radiation (insolation). Finally, there are rapid climate shifts such as those that were caused by the re-organisation of the coupled atmosphere-ocean circulation during the Dansgaard-Oeschger cycles (Bond et al., 1993; Kageyama et al., 2010). The combination of these styles of variability gives rise to a large and diverse set of examples of the response of regional and global climates to changes in climate forcing.

The impacts of past climate change are recorded by a variety of geological, isotopic and biological records (Bradley, 2014). These records can be interpreted, either using qualitative inference or explicit statistical approaches, to provide reconstructions of past climate variables. Such reconstructions document how the climate system behaves in response to different kinds of forcing – this illustration of what responses are physically possible is the basis for the idea that "the past is the key to the future" (Masson-Delmotte et al., 2013). However, there has been an increasing emphasis in recent years on the importance of palaeoclimatic and palaeoenvironmental reconstructions for climate-model evaluation (e.g. Izumi et al., 2013; Li et al., 2013; Perez Sanz et al., 2014). This arises from recognition that meteorological records from recent decades sample a range of climate variability that is too limited to provide a robust test of how well a numerical climate model can simulate a large climate change. Past climates provide a unique opportunity for "out-of-sample" evaluation of model performance, and thus a measure of the reliability of model predictions of the future (Braconnot et al., 2012; Harrison et al., 2014; Schmidt et al., 2014; Harrison et al., 2015).

Palaeoclimate simulations and data-model comparisons have been made for many iconic events in the past, including the early Holocene (ca 9 ka: Marzin and Braconnot, 2009; Marzin et al., 2013), Younger Dryas (ca 12.9–11.7 ka: Renssen et al., 2015), Last Interglacial (ca 125 ka: Bakker et al., 2013), the mid-Pliocene warm period (ca 4.2 Ma: Haywood et al., 2010a, 2010b), and the Eocene (ca 55-50 Ma: Lunt et al., 2012). In this paper, however, we only focus on the three periods that were included in the current phase of the Coupled Modelling Intercomparison project (CMIP5), the Last Millennium (850-1850 CE), the mid-Holocene (6 ka) and the Last Glacial Maximum (21 ka). These are the experiments that were used as part of climate-model evaluation reported in the Working Group 1 report to the Intergovernmental Panel on Climate Change (Flato et al., 2013). We summarise what has been learnt from the evaluation of these three simulations and the future challenges that face palaeoclimatology.

A Brief History of Palaeoclimate Simulations

The initial focus for palaeoclimate modeling was the Last Glacial Maximum (LGM, ca 21 ka), a time when there was a large change in forcing due to the presence of large ice sheets over North America (Laurentide ice sheet) and northern Europe (Eurasian ice sheet) and a radical change in atmospheric composition compared to the pre-industrial period. The earliest experiments (Alyea, 1972; Williams et al., 1974; Gates, 1976; Manabe and Hahn, 1977; Kutzbach and Guetter, 1986) were made with atmospheric general circulation models (AGCMs), and required changes in sea-surface temperature (SST) to be specified from observations (CLIMAP, 1976, 1981). In some cases (e.g. Alyea, 1972; Williams et al., 1974), simulations were confined to a single season because of the limitations in computing power. Nevertheless, these equilibrium simulations established that the presence of large ice sheets had a major impact on northern hemisphere climates, both through the direct effect of replacing vegetated land surfaces with highly-reflecting ice on albedo, and through the displacement of atmospheric circulation patterns caused by the increase in regional elevation by the mountain-like ice masses.

The Cooperative Holocene Mapping Project: COHMAP Members, 1988; Wright et al., 1993) subsequently broadened the focus to encompass simulations of the whole of the period from the LGM to present, in order to examine the impact of changing orbital configuration on radiative forcing and climate. However, these simulations were still equilibrium simulations made with an atmosphere-only model, thus requiring SSTs to be prescribed along with changes in the ice sheet height and extent, land-sea geography, atmospheric composition, and insolation. The COHMAP experiments were particularly important because they demonstrated the role of orbital changes in the evolution of the northern hemisphere monsoon systems (Kutzbach and Street-Perrott, 1985). A key aspect of the COHMAP project was the creation of large-scale syntheses of palaeoenvironmental and palaeoclimate data in order to document regional climate changes over the last glacial-interglacial, thus creating the basis for systematic comparisons of simulated and observed regional climates (Wright et al., 1993).

The availability of large-scale data syntheses, as well as the identification of mechanisms underpinning large-scale regional climate changes, was a motivation for the choice of the mid-Holocene (MH, 6ka) and the LGM as the experimental foci for the Palaeoclimate Modelling Intercomparison Project (PMIP). The goal of PMIP is to compare the behaviour of different climate models when run using the same forcing. The first phase of PMIP (PMIP1: Joussaume and Taylor, 2000) focused on comparison of AGCMs. By the second phase of the project (PMIP2: Crucifix et al., 2005), climate models routinely included an explicit simulation of ocean circulation (coupled ocean-atmosphere models: OAGCMs) and some models also included dynamic vegetation (coupled ocean-atmosphere-vegetation models: OAVGCMs). The evaluations of MH and LGM simulations carried out by PMIP have established unequivocally that climate models can reproduce observed, first-order global or hemispheric changes in climate in response to changes in forcing (Joussaume et al., 1999; Braconnot et al., 2007a, b; Zheng et al., 2008; Otto-Bliesner et al., 2009). However, they have also shown that models differ, often quite substantially, in their predictions, and comparison with palaeoclimate reconstructions shows that models often fail to capture regional changes accurately (e.g. Joussaume et al., 1999; Coe and Harrison, 2002; Brewer et al., 2007; Perez Sanz et al., 2014). Understanding the reasons for inter-model differences, and for model-data discrepancies has become the major focus of the third phase of the PMIP project (PMIP3: Braconnot et al., 2011; Braconnot et al., 2012) – and the reason that palaeoclimate experiments were included for the first time in CMIP5 (Taylor et al., 2011), the core international project that assembled model runs for the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).

Palaeo-simulations in CMIP5

Three palaeoclimate-simulations are included in the CMIP5 set of simulations: LGM, MH, and the Last Millennium (LM: 850-1850 CE). The LGM and MH simulations are equilibrium simulations. Both the LGM and the MH represent substantially different climate states from the present day and from each other, and have large natural forcings that are relatively well known (Braconnot et al., 2012; Harrison et al., 2015). The LM is a transient simulation to examine natural climate variability under conditions more similar to those of the present day (Schmidt et al., 2011).

At the LGM, the orbital parameters were nearly the same as they are today (Table 1) so that the differences in insolation were small. The major differences in forcing were caused by the presence of large ice sheets in the northern hemisphere (and concomitant changes in sea level and palaeogeography) and the lower atmospheric concentration of greenhouse gases. The changes in greenhouse gas concentrations (CO_2: 185 ppm, CH_4: 350 ppb, N_2O: 200 ppb) are well known from ice core records (EPICA Community Members, 2004). The decrease in the greenhouse gases relative to pre-industrial alone results in a radiative forcing of the troposphere of −2.8 W m^{-2} (Braconnot et al., 2007a). The expansion of the ice sheets at the LGM resulted in a sea-level lowering of ca 130m, and associated changes in albedo had an important effect on climate, particularly in the northern hemisphere. The marginal limits of the North American (Laurentide), Greenland and European (Eurasian) ice sheets are well known (e.g. Dyke and Prest, 1987; Mickelson and Colgan, 2003; Dyke, 2004; Gyllencreutz et al., 2007; Simpson et al., 2009; Ehlers et al., 2011; Mangerud et al., 2013) but there is no direct evidence for the distribution of ice mass. The form and height of the ice sheets are therefore inferred through a combination of physical modelling and indirect observational constraints (e.g. information on relative sea-level changes). A composite ice sheet was created for the CMIP5 experiments (Abe-Ouchi et al., 2015) by combining information from three reconstructions of the distribution of ice mass (ICE-6G v2.0: Argus and Peltier, 2010; GLAC-1a: Tarasov et al., 2012; ANU: Lambeck et al., 2010). In the CMIP5 LGM simulations, calculations using a simplified shortwave radiative model of the atmosphere, perturbed by changes in individual boundary conditions separately, show that the change in the ice sheets results in an implied forcing of between −1.85 and −3.49 Wm^{-2} depending on the climate model (Abe-Ouchi et al., 2015) while the overall change in forcing varied between –3.62 and –5.20 W m^{-2}. Thus, the change in forcing due to changes in atmospheric composition and expansion of the ice sheets at the LGM is of a similar magnitude to that projected for the next century.

The CMIP5 LGM experiments do not include the additional climate forcing that results from changes in vegetation distribution (Prentice et al., 2000; Harrison and Bartlein, 2012) because the observations of LGM vegetation are too sparse (in many regions) to provide a global gridded data set to use as a model input. LGM vegetation is therefore either computed (in models which include dynamic vegetation) or prescribed to be the same as the pre-industrial control simulation in the CMIP5 simulations (Table 1). The CMIP5 simulations also ignore the potential impact of known changes in atmospheric dust loading (Kohfeld and Harrison, 2001).

The MH provides an opportunity to evaluate simulations at a time of changed seasonality, when the influence of changes in ice sheet extent and land-sea geography on global climate was negligible. The seasonal and latitudinal distribution of MH insolation was different from present because of the difference in orbital configuration (Table 1). Seasonal contrast in the northern hemisphere was enhanced (by about 60 Wm^{-2}), through an increase in summer insolation and a decrease in winter insolation, and correspondingly reduced by decreased summer and increased winter insolation in the southern hemisphere. Greenhouse gas concentrations were similar to levels in the pre-industrial era (CO_2: 280 ppm, CH_4: 650 ppb, N_2O: 270 ppb. Although there were changes in vegetation distribution (Prentice et al., 2000; Harrison and Bartlein, 2012), these were not taken into account in the CMIP5 experiments (Table 1). As is the case for the LGM, the main focus of analyses of the MH experiments is on the impact of a large change in forcing on the mean climate response.

The LM is a transient simulation, included in CMIP5 to examine natural climate variability in a climate state close to that of the present day (Schmidt et al., 2011) and as a reference for detecting and attributing observed twentieth-century changes in climate patterns and trends resulting from human activities (Hegerl et al., 2011). The LM also provides opportunities to investigate the link between volcanism and climate, including ENSO variability (Emile-Geay et al., 2008; Wilson et al., 2010), to test the stability of atmospheric modes (e.g. Yiou et al., 2012), to analyse the interaction between short-term variability and land-surface feedbacks (e.g. Acosta Navarro et al., 2014), and to explore changes in recurrence or intensity of extreme events (e.g. Fallah and Cubasch, 2015). The LM simulation is characterized by changes in orbital, solar, volcanic and land-use forcing. With the exception of the orbital changes, there are large uncertainties associated with each of these forcings (Schmidt et al., 2011). The CMIP5 protocol therefore defines a number of alternative forcing histories to take account of these large uncertainties (Table 1). There are, for example, two reconstructions of the volcanic forcing (Crowley et al., 2008; Gao et al., 2008), five reconstructions of solar forcing (Wang et al., 2005; Muscheler et al., 2007; Steinhilber et al., 2009; Delaygue and Bard, 2011; Vieira et al., 2011) and two land-use scenarios (Hurtt et al., 2006; Pongratz et al., 2008). Modelling groups have been allowed to choose which forcing “scenarios” they use. While this makes comparison between models more difficult, some modelling groups have run ensembles of simulations using different forcing scenarios (e.g. Goosse et al., 2005; Bothe et al., 2013; Otto-Bliesner et al., 2015) thus allowing the effects of uncertainty in forcing to be assessed.

Many modelling groups have run palaeoclimate simulations as part of CMIP5 (Table 2). The MH experiment is relatively simple and has a smaller perturbation than the LGM, thus requiring less time to reach equilibrium. Thus many more groups have performed the MH experiment than have performed the LGM experiment. Only a few groups have performed the LM simulation – in part because multiple forced and unforced runs are required for a complete diagnosis. Nevertheless, there are sufficient simulations for all three periods to allow comparisons of the reaction of different climate models to the same change in forcing and evaluation of the realism of the simulations through comparison with palaeodata.

A Brief History of Palaeodata Synthesis

With the exception of ice-core records of the well-mixed trace gases, individual palaeoenvironmental records document local or regional changes – although the spatial sampling scale may vary from metres up to some tens or hundreds of kilometres. The synthesis of records at a regional scale provides a way of documenting robust responses to past climate changes. Regional data syntheses are the appropriate tool for extracting information that is comparable to simulated climates, given the spatial resolution of current climate models. The highest resolution of the CMIP5 models used for palaeoclimate experiments, for example, is ca 1° x 1° latitude/longitude.

The comparison of individual records from a region, and identification of similarities in their response, is standard practice. Data synthesis, however, requires that the individual records are interpreted using a common approach. One of the earliest examples of this was the synthesis of lake records from northern Africa (Street and Grove, 1976) that led to the creation of the Global Lake Status Database (GLSDB: Street and Grove, 1979; Street-Perrott et al., 1989; Kohfeld and Harrison, 2000; Fig. 1), one of the databases used by the COHMAP project. The GLSDB had transparent rules for site selection, and used an explicit method to categorise individual records into status classes (high, intermediate, low) so that they were easily compared both within and between regions. This focus on lake status also facilitated direct comparison with model output, because lake status is sensitive to changes in the balance between precipitation and evaporation (Street-Perrott and Harrison, 1984; Cheddadi et al., 1987).

The COHMAP project used pollen data as a source of information about regional vegetation and climate, but it was not until the creation of the Palaeovegetation Mapping Project (BIOME 6000: Prentice and Webb, 1998) as part of the International Geosphere-Biosphere Programme that these data were treated in a systematic and consistent way. BIOME 6000 developed an approach to translate pollen assemblages into vegetation reconstructions, quaintly termed biomisation, which involved classification of individual pollen taxa into plant functional types (PFTs), the characterization of major vegetation types (biomes) according to their characteristic or defining PFTs, and the application of an algorithm to select the most likely biome represented at a site (Prentice et al., 1996). BIOME 6000 produced vegetation maps for the MH and LGM (Prentice et al., 2000; Bigelow et al., 2003; Pickett et al., 2004; Marchant et al., 2009), explicitly for comparison with vegetation simulations made either using OAVGCMs or by running a biogeography model driven by outputs from e.g. OAGCMs (e.g. Harrison et al., 1998; Wohlfahrt et al., 2004). The biomisation approach has also been used to produce maps for other time intervals for certain regions (Marchant et al., 2001; Williams et al., 2004).

There have been other efforts to create datasets comparable to model outputs. The Dust Indicators and Records from Terrestrial and MArine Palaeoenvironments (DIRTMAP: Kohfeld and Harrison, 2001; Maher et al., 2014) database contains estimates of aeolian accumulation rates at key time periods measured in ice cores, marine cores and at terrestrial locations. The modern and LGM dust deposition estimates from DIRTMAP have been used for evaluation of dust-cycle simulations (e.g. Werner et al., 2003; Bauer and Ganopolski, 2014). The Global Palaeofire Working Group (website) has created a global synthesis of charcoal records (Power et al., 2010), which provides a qualitative record of changes in biomass burning of the last glacial-interglacial cycle. Much of the focus on this group has been on documenting regional changes (Power et al., 2008; Marlon et al, 2008; Daniau et al., 2010; Marlon et al., 2013) or investigating the controls on fire (Daniau et al., 2012), but the data set has potential to be used for model evaluation (e.g. Brücher et al., 2014).

Palaeoenvironmental data have long been used to reconstruct climate variables quantitatively (e.g. Grichuk, 1969; Imbrie and Kipp, 1971; McIntyre et al., 1976; Hutson and Prell, 1980; Bartlein et al., 1984; Atkinson et al., 1987; Guiot, 1987; Huntley and Prentice, 1988; Guiot, 1990). The development of well-documented, quantitative global palaeodata sets portraying the spatial climatic patterns of the LGM and MH time periods is a central objective of the PMIP research programme.

The palaeoceanographic community has provided a global reconstruction of LGM SSTs (MARGO Project Members 2009), which supersedes the CLIMAP data set that was developed in the 1980s. MARGO (Multiproxy Approach for the Reconstruction of the Glacial Ocean surface) defined the LGM as the interval between 19 and 23 ka. The project compiled 696 SST reconstructions from this interval. The data set includes all available microfossil-based (transfer functions based on planktonic foraminifera, diatom, dinoflagellate cyst and radiolarian abundances) and geochemical (alkenones and planktonic foraminiferal Mg/Ca ratio) reconstructions. Each type of sensor has a different geographical coverage – the reconstructions from the Southern Ocean, for example, are largely based on diatom records, whereas most of the tropical records are derived from foraminiferal assemblages. Nevertheless, there are some regions of the world where reconstructions based on multiple sensors are available and could be compared to provide an estimate of robustness.

In the global reconstruction, the data were gridded at $5^\circ \times 5^\circ$ resolution, where each grid cell was assigned an SST estimate by averaging individual reconstructions that fall into the same cell, weighted by a mean reliability index. The resulting SST anomalies show robust spatial and seasonal changes (Fig. 2), and there is first-order agreement on the magnitude of latitudinal anomalies between geochemical and microfossil-based reconstructions with the strongest mean annual cooling in the mid-latitude North Atlantic – a feature confirmed by reconstructions from four different types of sensor.

For the MH, the only global SST product available is the Global database for alkenone-derived HOlocene Sea-surface Temperature (GHOST), which includes reconstructions based on Mg/Ca and alkenones (Kim, 2004; Leduc et al., 2010). Model comparisons using the GHOST data set have shown significant mismatches between the modelled and reconstructed SST anomalies (Schneider et al., 2010; Hargreaves et al., 2013; Lohmann et al., 2013). An attempt has been made to produce a more comprehensive data set, including reconstructions from Mg/Ca and alkenone palaeothermometry and statistical estimates obtained using planktonic foraminifera and organic-walled dinoflagellate cyst census counts (Hessler et al., 2014). However, analyses of these data show that the MH change in SST is small compared to the magnitude of known methodological uncertainties associated with SST reconstructions, and also compared to the differences between the observed modern ocean temperature datasets used as the baseline to determine the MH change in SST. Hessler et al. (2014) concluded that, unlike the LGM, where robust changes in SST patterns emerge despite the methodological uncertainties (MARGO Project Members, 2009), MH SSTs do not provide a reliable benchmark for model simulations.

Terrestrial environments are diverse and many types of geochemical, isotopic and biological data have been used to provide quantitative reconstructions for specific areas and ecosystems (see e.g. Atkinson et al. 1987; Stute et al. 1992; Heiri et al. 2003; Jones et al. 2004). The most widespread source of quantitative reconstructions is palaeovegetation (fossil pollen and plant macrofossil) records. Palaeovegetation records provide a unique combination of near-global coverage of information on several distinct aspects of climate (seasonal temperature, rainfall, soil moisture), combined with robust and well-documented methodologies to derive reconstruction uncertainties.

The terrestrial palaeoecology community has produced a unified gridded data set for the MH and the LGM based on combining all existing quantitative reconstructions, subject to availability of the primary data (i.e. the reconstructions) and a transparent screening procedure (Bartlein et al., 2011). Although the reconstructions were produced using different techniques, ranging from simple regression through analogue techniques to inverse modelling, analyses for regions with multiple reconstructions made using different methods show that the choice of method has little impact on the results. Thus, compositing reconstructions made with different methods provides robust and coherent reconstructions of the large regional climate changes at the MH and LGM (Fig. 2). Although this synthesis represents the state-of-the-art target for model evaluation and benchmarking, the coverage is poor for many important regions including Australia and South America. There are pollen records from both regions (Fig. 2) that could be used to make statistical reconstructions of climate variables; even in regions that are relatively well represented in the gridded data set, there is the potential for a much-expanded set of climate reconstructions.

Pre-industrial climate provides a baseline for the detection and attribution of recent anthropogenic impacts on the Earth system (Hegerl et al., 2011), and this provides the major motivation for the inclusion of LM simulations in CMIP5. Reconstruction of annual climate before the pre-instrumental period relies on the use of natural archives, including isotopic records from laminated sediments or corals, ice core records and tree rings. However, statistical reconstructions from tree rings provide by far the largest number of pre-instrumental records. The major focus of data synthesis to date has been on seasonal (e.g. Briffa et al., 2002; Luterbacher et al., 2004; Guiot et al., 2005; Xoplaki et al., 2005) or annual temperature. Reconstructions of regional or hemispheric temperature changes over the last millennium (e.g. Jones et al., 1998; Briffa et al., 2002; Esper et al., 2002; Moberg et al., 2005; Rutherford et al., 2005; Mann et al., 2007, 2008; Ljungqvist et al., 2012; PAGES 2k Consortium, 2013; Shi et al., 2013; Neukom et al., 2014) generally use several of these palaeodata sources, combined with historical and instrumental records when available. While there are regional reconstructions of precipitation (Pauling et al., 2006; Steinman et al., 2012; Wilson et al., 2012; Feng et al., 2013), there is currently no global synthesis of precipitation data. Although there is broad agreement on multidecadal to centennial time scales, there is considerable variability among individual hemispheric temperature reconstructions on short time scales over the last millennium (Fig. 3), depending on methodology and the selection of site-based reconstructions included in the reconstructions (Juckes et al., 2007; Fernandez-Donado et al., 2013). The range in reconstructed change in northern hemisphere average temperature between the Mediaeval Warm Anomaly and the Little Ice Age, for example, encompasses the simulated range of temperature change across different models using different combinations of forcings and including simulations made with and without volcanic forcing (Fernandez-Donado et al., 2013). Thus, the large uncertainties in the reconstructions coupled with similarly large uncertainties in the forcing currently limits the usefulness of the last millennium as a target for model evaluation *sensu stricto*.

Confronting models with observations

Palaeodata document what has actually happened in the past, but explanation of observed changes is dependent on a conceptual model of how the climate system works and is therefore rarely unequivocal. Models that incorporate current understanding of physical climate processes provide a way of making the conceptual model explicit. Thus, one of the most fruitful approaches to understanding the mechanisms of climate change is through confronting observations and model experiments in hypothesis-testing mode, where the ability of the model to reproduce observed patterns in space or time indicates the plausibility of the underlying conceptual explanation while disagreement indicates that alternative explanations are required.

The demonstration by Kutzbach and Street-Perrott (1985) that the evolution of the African monsoon over the last glacial-interglacial cycle was a direct response to orbital forcing (Kutzbach and Street-Perrott, 1985) provides the classic example of this hypothesis-testing approach. In this paper, the water balance (precipitation minus evaporation) over northern Africa (8.9 and 26.6 °N) was calculated based on a sequence of only January or only July climates (known as perpetual January or perpetual July simulations, where the mean annual climate is then calculated as the average of the two monthly simulations and the seasonal contrast as the difference between these two months) with an atmospheric general circulation model forced by changes in insolation, ice sheet extent, and sea-surface temperatures. The simulations predicted the observed temporal evolution of lake status. Analyses of the simulations (including additional sensitivity experiments) confirmed that the primary driver of the observed changes in lake status was changes in orbital forcing. Changes in boundary conditions changes associated with northern-hemisphere ice sheets or atmospheric composition had little impact on the regional water balance. The primary importance of orbitally-induced changes in insolation as a driver of the waxing and waning of the northern hemisphere monsoons has subsequently been confirmed with more advanced models and modeling protocols (Zhao et al., 2005; Braconnot et al., 2007; Marzin and Braconnot, 2009; Dallmayer et al., 2015). However, it is clear that there is considerably more complexity in the seasonal evolution of monsoon rainfall than originally thought (Fig. 4) and considerable millennial- and sub-millennial scale variability is superimposed on the orbitally-driven evolution (Otto-Bleisner et al., 2014). It is also clear that feedbacks associated with the ice sheets, ocean conditions and climate-induced changes in land-surface conditions are necessary to produce the observed temporal evolution of the northern hemisphere monsoons (e.g. Clausen and Gayler, 1977; Ganopolski et al., 1998; de Noblet-Ducoudré et al., 2000; Zhao et al., 2005; Patricola and Cook, 2007; Zhao et al., 2007; Marzin and Braconnot, 2009; Ohgaito and Abe-Ouchi, 2009; Dallmeyer et al., 2010; Zhao and Harrison, 2012; Dallmeyer et al., 2015).

This hypothesis-testing approach underpins data-model comparison of regional climate changes during the MH and LGM conducted as during the first phase of PMIP, which focus on demonstrating how far large-scale patterns are a consequence of changes in orbital and/or glacial boundary conditions. For example, comparisons with model simulations driven by the combined influence of known changes in orbital, ice sheet and greenhouse gas forcing have been used to explain observed differences in the temporal evolution of fire regimes between tropical and extratropical regions of the northern and southern hemisphere over the last glacial-interglacial cycle (Daniau et al., 2012; Fig. 5). However, the hypothesis-testing approach is much more powerful when it is used to test potential mechanisms explicitly through experiments that separate out the potential influence of individual forcings. For example, Harrison and Prentice (2003) used a simple biogeography-biogeochemistry model driven by climate-model simulations of the LGM to demonstrate the necessity of including the direct impacts of low CO_2 on productivity and water-use efficiency to explain observed changes in tropical vegetation distribution. They showed that the area of tropical forests would have increased in response to climate changes at the LGM, whereas the observed reduction of tropical forest and increase in grassland could only be achieved when CO_2 was lowered to glacial levels. A similar conclusion was reached by Bragg et al. (2013), comparing simulated and observed glacial-interglacial changes in leaf-wax $\delta^{13}C$ of terrestrial origin from a transect of marine cores recording vegetation shifts in southern Africa. Bragg et al. (2013) also discussed the general importance of atmospheric CO_2 concentration as a driver of vegetation changes, and the relative roles of climate and CO_2 changes in glacial-interglacial vegetation shifts – a topic that has suffered from some misconceptions, as the two kinds of effect are neither mutually exclusive, nor independent. The importance of the direct effects of changing CO_2 on vegetation productivity, and hence fuel load, has subsequently been demonstrated as an important control on LGM fire regimes (Martin Calvo et al., 2014).

Model evaluation and benchmarking

The importance of assessing how well state-of-the-art climate models can simulate large climate changes has led to the increasing use of palaeodata for the purposes of model evaluation. At its simplest, model evaluation can involve qualitative comparisons of spatial patterning. Such map-map comparisons can be powerful. For example, the inability of climate models to capture the spatial expansion of the northern African monsoon during the MH is readily apparent by comparing maps of observed and simulated water balance (see e.g. Perez Sanz et al., 2014). However, when the discrepancies are in the magnitude of a signal rather than spatial pattern (or sign) then quantitative comparisons are necessary.

There are many potential sources of uncertainty in using palaeodata to make climate reconstructions. Some of these uncertainties are strictly numerical but others are associated with dating, methodologies, baseline choice, or interpretation – and are much more difficult to deal with when making quantitative comparisons. Numerical uncertainties (e.g. root mean-squared errors on statistically-based climate reconstructions) are easily factored into data-model comparisons (see e.g. Hargreaves et al., 2013). The other sources of uncertainty, even when quantifiable, are often ignored.

An absolute chronology is fundamental to comparisons of palaeo-records and the construction of palaeodata syntheses. The development of reliable techniques to construct age models has been a major focus for the community (e.g. Bennett, 1994; Boreux et al., 1997; Bennett and Fuller, 2002; Blaauw et al., 2003; Heegaard, 2003; Blaauw and Christen, 2005; Bronk Ramsey, 2009; Blaauw, 2010; Blaauw and Christen, 2011 Werner and Tingley, 2015). Age models are only meaningful when created using calibrated radiocarbon dates (Bartlein et al., 1995) because of the variability in the radiocarbon calibration curve. However, the gradual refinement of the radiocarbon calibration curve (Reimer et al., 2009: Reimer et al., 2013), and increasing understanding of the need to account for reservoir ages in deriving calibrated ages on both marine (Craig, 1957; Stuiver et al., 1998; Reimer and Reimer, 1991; Franke et al., 2008) and freshwater (Godwin, 1951; Philippsen, 2013) sediments, means that even calibrated age models may need to be revisited during the construction of data syntheses. Although there has been an awareness of chronological uncertainties, most approaches to dealing with these in the context of data-model comparison have been in terms of either selecting sites with chronologies that are believed to be most reliable or through assigning some kind of quality control index (e.g. Street-Perrott et al., 1989; Wright et al., 1993; MARGO Project Members, 2009; Giesecke et al., 2014) – an approach that is difficult to combine with numerical estimates of uncertainty.

There are uncertainties caused by the different techniques used in different laboratories for measuring particular variables. Differences in the protocols used for sample cleaning, types of machine used for measurement and machine calibration, for example, have been shown to yield differences of up to 3°C in the sea-surface temperature estimates derived from Mg/Ca measurements of planktonic foraminifera (Roesenthal et al., 2004; Greaves et al., 2008). Similarly large inter-laboratory differences have been found for stable isotope analyses on bone collagen, again deriving from differences in sample cleaning and instrumentation (Pestle et al., 2014). While inter-laboratory differences in other types of measurement appear to be smaller than the actual measurement uncertainty (e.g. Foster et al., 2013), the fact that there are differences between measurements made by different groups poses difficulties for data synthesis. Again, it is difficult to know how to incorporate these uncertainties within a traditional data-model comparison framework.

The general approach to using palaeo-reconstructions for model evaluation is to use the estimated change in reconstructed climate and compare this with the simulated change between a palaeo-experiment and a control, usually a pre-industrial control. Little thought has been given to the choice of baseline climate, either for the reconstructions or for the simulations. Hessler et al. (2014) showed that the choice of baseline climate, in this case SST anomalies based on either the WOA98, WOA09 or HadiSST data sets, made an average absolute difference of 0.3-0.4°C to mid-Holocene SST reconstructions with differences of >1°C in the Mediterranean and eastern Pacific. Although the MARGO Project used a standard baseline climatology (MARGO Project Members, 2009), other SST data sets (e.g. Ruddiman and Mix, 1993; Leduc et al., 2010; Marcott et al., 2013) have different definitions of the baseline climate, and this needs to be taken into account when combining these sources to create data sets for model evaluation. Differences in temperature between the pre-industrial (PI) control and the mid-20th century in the historical simulations (1961-1990 CE, i.e. the interval most nearly corresponding to the modern observational data sets used for statistical calibrations) can also be of the order of 0.5–1.0°C locally and this may also contribute to mis-matches between simulated and reconstructed climates (e.g. Wagner et al., 2012).

The largest source of unquantifiable uncertainty in palaeoclimate reconstructions is associated with the climate interpretation of a given record. Biological assemblages contain a wealth of information, and this underpins their use to make reconstructions of multiple climate variables (e.g. Webb et al., 1993; Jackson et al., 2000; Davis et al., 2003; Cheddadi et al., 2007; Fréchette et al., 2008; Guiot et al., 2008). Derivation of a statistical relationship with a specific climate variable under modern climate conditions is based on the fact that this variable either controls or is correlated with something that controls the growth of the organism. July temperature, for example, has a limited direct impact on plant growth, but it is generally correlated in the northern hemisphere with the length and warmth of the growing season, which is the major determinant of whether plants can accumulate sufficient carbon to survive and reproduce. Very high July temperatures also tend to be associated with heat and/or moisture stress that can impact photosynthesis and strongly determine the composition and structure of vegetation (Kohfeld and Harrison, 2000; Harrison et al., 2010). Similarly, mean January temperature in the northern hemisphere is usually highly correlated with daily extreme low temperatures in winter, which determine whether a plant is killed by frost and therefore exert a strong selective pressure that differentiates plants with different overwintering mechanisms (Woodward, 1987; Harrison et al., 2010). The definition and adoption of "bioclimatic" variables, such as mean temperature of the coldest month, accumulated growing season warmth, and indices of plant-available soil moisture, for climate reconstruction was an attempt to move closer to the actual controls on plant growth (e.g. Cheddadi et al., 1997; Tarasov et al., 1999; Peyron et al., 2000) and which could therefore be expected to be invariant through time. Nevertheless, the palaeoclimate record is characterized by changes in seasonality, interannual variability and the frequency of extremes – all of which have the potential to invalidate modern-day correlations even between bioclimatic variables and species abundance. Furthermore, at least as far as terrestrial plants are concerned, statistical relationships with climate are modulated by the fact that plants respond directly to changes in atmospheric CO_2 concentration through changes in productivity and water-use efficiency (Street-Perrott et al., 1997; Cowling and Sykes, 1999; Harrison and Prentice, 2003: Prentice and Harrison, 2009). It is not possible to take this into account using statistical techniques, and this is likely a contributory cause of the breakdown of statistical relationships between climate and tree-ring width in recent years (D'Arrigo et al., 2008; Gagen et al., 2011) and could also impact reconstructions of high-CO_2 intervals such as the mid-Pliocene (e.g. Salzmann et al., 2013) and low-CO_2 intervals such as the LGM (Jolly and Haxeltine, 1997; Cowling and Sykes, 1999; Prentice and Harrison, 2009). The effect of changing CO_2 will also impact on palaeo-reconstructions of other plant properties, including leaf area index and tree cover (e.g. Gonzalez et al., 2008; Williams et al., 2011).

An alternative way of exploiting palaeoenvironmental data for climate-model evaluation is to use models that explicitly simulate the sensor – for example, vegetation (Kaplan et al., 2003; Prentice et al., 2011a), tree rings (Evans et al., 2006; Li et al., 2014), fire (Prentice et al., 2011b; Martin Calvo et al., 2014), the dust cycle (Werner et al., 2003; Mahowald et al., 2006; Takemura et al., 2009), peat growth (Charman et al., 2013), glacier mass balance (Michelmayer et al., 2008), marine biogeochemistry (Aumont et al., 2003; Bopp et al., 2003) or phytoplankton abundance (Le Quéré et al., 2005), and stable isotopes in corals (Thompson et al., 2011). This type of "forward" modelling using a simple biogeography model (BIOME4: Kaplan et al., 2003), for example, makes direct comparisons with observations possible and discriminates between the performance of different climate models (Fig. 6). Many climate models now explicitly simulate isotopic tracers (e.g. Schmidt et al., 2007; Sturm et al., 2010; Holloway et al., 2016) to facilitate model evaluation and diagnosis. Similarly, there are an increasing number of models that simulate vegetation, fire and the dust cycle (e.g. Lawrence et al., 2011; Reick et al., 2013; Kok et al., 2014), primarily in order to account for feedbacks to climate. The explicit simulation of these components of the climate system greatly facilitates comparisons with natural records (see e.g. Wasson and Claussen, 2002; Ohgaito et al., 2013).

A natural extension of the forward modelling approach is to use inversion techniques to derive quantitative climate reconstructions that are consistent with a process-based model (Guiot et al., 2000; Wu et al., 2007; Hatté et al., 2009; Garretta et al., 2010; Boucher et al., 2014). The use of process-based models in palaeoclimate reconstruction sidesteps many of the potential problems with correlation-based statistical methods. One caveat to the use of process-based modelling is the assumption that the model used is correct. Projections of future changes in vegetation (Piao et al., 2013; Friedlingstein et al., 2014) and fire (Harrison et al., 2010; Kloster et al., 2012; Kelley and Harrison, 2014) show that different process-based models produce radically different simulations for the 21st century, despite being equally good at reproducing modern day vegetation patterns and fire regimes. Thus, as with climate models, it is imperative either to use an ensemble of models or to demonstrate that the forward-model selected is reliable. It should also be noted that process-based modelling does not overcome the problem of equifinality (different climates generating similar effects), although it is possible to use this modelling approach to determine the range of potential climates and the probabilities associated with each (Garreta et al., 2010).

A particular motivation for the use of process-based models for reconstructing climate from palaeovegetation records is the strong effect of changes in atmospheric CO_2 concentration on vegetation composition (as discussed above). Manifest today in worldwide "woody thickening" (increase of tree density especially in tropical savannas), this same effect also accounts for much of the extreme reduction in forest area in the tropics and subtropics during glacial periods; and the globally lower than present terrestrial carbon storage at the LGM (Harrison et al., 2010), which was only partly counteracted by greater than present storage of inert carbon in permafrost (Ciais et al., 2013). There is no obvious way to build the CO_2 effect into statistical climate reconstruction methods because at any one time there is very limited variation in CO_2 concentration across the globe. Process-based models, including BIOME4, include a CO_2 effect on vegetation composition (a consequence of the effect of CO_2 on photosynthesis, and the differential effects on plants with the C_3 and C_4 pathways) and so inversion of such models can take known changes in CO_2 concentration into account (Guiot et al., 2000; Wu et al., 2007; Hatté et al., 2009; Garretta et al., 2010; Boucher et al., 2014). An alternative approach, which decouples the consideration of CO_2 effects from the use of a specific process-based model, involves defining a bioclimatic index that reflects "apparent" plant-available moisture, as sensed by plants responding to changes in atmospheric CO_2. Wang et al. (2013) used this approach to modify results of a statistical model to predict vegetational responses to future climate change. It could potentially be adapted to "correct" palaeoclimate reconstructions made from palaeovegetation data by any method (Prentice, 2015). The correction would generally be to increase palaeoprecipitation estimates for periods of low CO_2 and to decrease them for periods of high CO_2. The use of such a physically-based correction factor could provide a rapid method of modifying existing statistical reconstructions of palaeoprecipitation to account for the direct impact of CO_2 on plant growth.

Despite various sources of uncertainty, quantitative reconstructions can be used for model evaluation and benchmarking as long as the climate signal being examined is larger than the potential uncertainties (Lohmann et al., 2013; Harrison et al., 2014) and especially when reconstructions derived using different methods (e.g. statistical techniques, forward modelling, model inversion) show similar, spatially coherent patterns that are consistent with a single climatic explanation (Bartlein et al., 2011).

The terms 'evaluation' and 'benchmarking' are not synonymous. Benchmarking is a measurement tool, whereby model outputs are compared to a pre-defined set of observations using appropriate metrics to define the degree of agreement quantitatively (Taylor, 2001; Gleckler et al., 2008). Benchmarking serves multiple functions. It allows the performance of different models to be compared, but it can also be used to identify processes that require improvement in a particular model or to evaluate parameter choices, including ensuring that improvements to one component of a model do not compromise performance in another. Benchmarking is routinely used to assess climate-model performance under modern conditions, including investigation of parameter uncertainties (e.g. Murphy et al., 2004) and multi-model comparison (e.g. Reichler and Kim, 2008). It has been used to inform model development (e.g. Jackson et al., 2008) and to assess the reliability of projections of future climate (e.g. Hall and Qu, 2006). Globally comprehensive syntheses that include multiple climate variables now make routine benchmarking of palaeosimulations possible (e.g. Flato et al., 2013; Hargreaves et al., 2013; Harrison et al., 2014).

Evaluation of the CMIP5 simulations: what have we learnt?

There are now many papers presenting analyses and evaluations of the CMIP5 palaeosimulations. The PAGES 2k–PMIP3 group (2015) have made initial analyses of the LM simulations. A preliminary summary of the analyses related to the MH and LGM simulations was presented by Harrison et al. (2015). Here we draw on the Harrison et al. (2015) paper to outline some of the major lessons that have been learnt from comparing simulated and reconstructed climates for these two periods.

Large-scale features of climate that are governed by the energy- and water-balance show remarkably consistent simulated responses to changes in forcing in different climate states. For example, the magnitude of the temperature change over land compared to ocean is consistent in both warm and cold climate states: depending on the sign of the forcing, the land warms or cools by ca 2.36 times more than the ocean (Fig. 7). The ratio of the land-sea temperature contrast is constant over a wide range of climates, including climates with higher-than-present CO_2 levels (Izumi et al., 2013; Lunt et al., 2013; Hill et al., 2014; Schmidt et al., 2014). Land-ocean contrast is primarily driven by changes in surface downward clear-sky longwave radiation, which includes the effect of changes in CO_2, water vapour, and atmospheric energy transport (Izumi et al., 2014). The relative change in tropical temperature compared to high latitudes (often referred to as polar or high-latitude amplification) is also consistent across different climate states. Again, the major driver of this contrast is surface downward clear-sky longwave radiation, with surface albedo playing a significant but secondary role in promoting high-latitude amplification in both cold and warm climates (Izumi et al., 2014). The simulated magnitude of the relative changes in land-sea contrast and in high-latitude amplification is supported by historical and LGM observations, confirming that the simulated changes are realistic (Izumi et al., 2013). Thus, palaeo-evaluation of the CMIP5 simulations does confirm that the large-scale patterns of temperature change in future projections are believable.

Large-scale changes in precipitation scale with temperature, increasing as temperature increases and decreasing in cold climates. The change in precipitation per degree change in temperature is approximately the same in palaeoclimate, historical, and increased CO_2 simulations. The rate of change is consistently smaller than the rate of change in saturation vapour pressure (i.e. it is much less than predicted by the Clausius-Clapyeron relationship), partly because of energetic constraints on evaporation, and partly because of constraints in water availability over land (Trenberth and Shea, 2005; Allan, 2009). Geographical differences in the strength of these constraints means there are larger changes in precipitation per degree temperature change over the ocean than over land, and in extratropical than tropical land areas (Li et al., 2013). All of these large-scale features are consistent with palaeoclimate and historical observations (Li et al., 2013). Again, the palaeoclimate diagnosis of the CMIP5 simulations confirms that the large-scale patterns of precipitation change in future projections are believable.

The CMIP5 simulations of MH and LGM climates show only moderate skill in predicting observed patterns of climate change overall (Hargreaves et al., 2013; Hargreaves and Annan, 2014; Harrison et al., 2014; Harrison et al., 2015) and this arises because of persistent problems in simulating regional climates. In the MH, for example, models predict an increase in the northern hemisphere monsoons in response to the orbitally-driven increase in summer insolation. This increase in monsoons is amplified by ocean and land-surface feedbacks. Nevertheless, the models do not produce as large a change in either the amount of rainfall or the extent of the area influenced by monsoon precipitation as indicated by observations. The discrepancy between observed and CMIP5 simulated changes in MH precipitation over northern Africa between 15°-30° N is at least 50% (Perez-Sanz et al., 2014). The mismatch between simulated and observed monsoon climates has not been reduced in the CMIP5 simulations compared to simulations made with earlier generations of models (Harrison et al., 2015).

Failure to capture the magnitude of an observed change suggests that there are feedback processes that are either not included or are poorly treated in the current generation of models. However, differences in the sign of regional changes between observations and simulations are likely to indicate more fundamental problems. The CMIP5 MH simulations show drier conditions in the Eurasian mid-continent, particularly between 45°-60° N, whereas observations systematically show the region was wetter than today. The simulated drying leads to a significant warm temperature bias in this region, whereas observations indicate that the mid-continent had cooler summers than today. Discrepancies in the sign of regional climate change are also found in other extratropical regions, most notably in southern Europe where the models show warmer summers and the observations indicate cooler summers during the MH (Mauri et al., 2014). Mauri et al. (2014) suggested this mismatch was due to poor simulation of the short-term variability in atmospheric circulation, specifically the prevalence of anticyclonic blocking in summer and increased dominance of the positive phase of the North Atlantic Oscillation in winter during the MH.

There has been little assessment of how well models reproduce changes in short-term climate variability during the MH and LGM, in part because of the lack of large-scale syntheses of high-resolution palaeodata. In general, models underestimate interannual variability under modern climate conditions (Flato et al., 2013). The direct observational record is too short to know how well they capture decadal to centennial-scale variability. Oxygen isotope measurements on marine carbonates (corals, molluscs) from the tropical Pacific Ocean show a substantial reduction in the strength of the El Niño-Southern Oscillation (ENSO) through most of the Holocene, culminating between 5 and 3 ka (Emile-Geay et al., 2015). Most of the CMIP5 simulations show a reduction in ENSO in the MH compared to the pre-industrial climate, but it is very much smaller than the reduction shown by the palaeo-observations and only marginally significant. This raises the possibility that an important component of the observed changes in ENSO may result from internal variability. The contribution of internal variability to projected future climate generally decreases through the 21st century, but nevertheless remains an important contribution to the uncertainties in projections of regional precipitation e.g. in Asia and Europe throughout the century (Kirtman et al., 2013). It is also clear that a considerable part of the differences in the simulated response to forcing during the last millennium can be explained in terms of internal variability (Goosse et al., 2012; Masson-Delmotte et al., 2013).

Model responses to forcing at a regional scale are not always consistent. Various CMIP5 models show opposite changes in the location of the southern hemisphere westerlies during the LGM, for example, with half showing a equatorward shift and half showing a poleward shift in mean position compared to the pre-industrial control (Chavaillez et al., 2013; Rojas, 2013). The models that unexpectedly simulate a poleward shift of the jet stream at LGM show a strong cooling in the lower troposphere at high latitudes, which suggests that inter-model differences in the position of the westerlies may reflect different sensitivity to prescribed changes in the Antarctic ice sheet (Chavaillez et al., 2013). In the CMIP5 MH simulations, there is a consistent reduction of summer sea-ice cover in response to increases in summer insolation but some models show increased and some decreased ice thickness in winter (Berger et al., 2013). These inter-model differences appear to be related to differences in the cloud feedback. Differences in the response between models are potentially helpful, providing that the actual response is well-constrained by observations, because they offer the possibility of determining the correct sensitivity to different feedbacks.

The systematic biases in the simulation of regional climates means that models are generally better at simulating mean values of any climate variable than at simulating the spatial variability or the geographical patterning in that variable (Harrison et al., 2014: Fig. 8). Nevertheless, the benchmarking of the CMIP5 MH and LGM shows that some models consistently perform better than others, even in the prediction of spatial patterning (Harrison et al., 2014). Unfortunately, better performance in palaeo-simulations is not related to better performance under modern conditions (Harrison et al., 2015). The ability to simulate modern climate regimes and processes does not mean that a model will be good at simulating climate changes. This emphasises how important it is to test models against the palaeorecord if we are to have any confidence in their projections of future climate (Braconnot et al., 2012; Hargreaves and Annan, 2014; Schmidt et al., 2014).

The future

Evaluation of the CMIP5 palaeo-simulations demonstrates the value of using past climates as data targets in model intercomparisons. It has been shown that the broad-scale temperature and precipitation simulated responses to past changes in forcing are correctly represented, and this suggests they are features of the actual response of the climate system to changes in forcing rather than model artefacts. Projected changes in land-sea temperature contrast, high-latitude amplification, temperature seasonality and the scaling of precipitation with temperature are therefore likely to be reliable. But models are much less reliable at predicting regional climate changes. 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 or less reliable. Efforts to improve the skill of climate models based on evaluation using modern climate states are having a declining impact (Knutti, 2010; Rausser et al., 2014), pointing to a need for innovation (Stevens and Bony, 2013; Palmer, 2014). We are therefore at a key moment for the climate modelling enterprise to benefit from insights gained from the study of past climates. There are a number of areas that have been identified as potential sources of error in the simulation of regional palaeoclimates, including the balance between deep and shallow convection in monsoon regimes (Zheng and Braconnot, 2013), incorrect representation of water- and energy-exchanges between the land and the atmosphere (Harrison et al., 2015), poor understanding of the relationship between mean climate state and short-term climate variability (Emile-Geay et al., 2015) and failure to capture the short-term variability in atmospheric circulation (Mauri et al., 2014). Further investigation of these issues in the radically different climate regimes of the past could provide clues to improve state-of-the-art models.

The need for renewed effort is not confined to the modeling community. For example, our ability to evaluate model performance in the southern hemisphere is currently limited by a lack of coherent and consistent syntheses of the available palaeoenvironmental data. There is an urgent need for quantitative climate reconstructions covering South America and Australia. The use of existing quantitative reconstructions could also be improved, in particular through the development of standardized measures of uncertainty and exploitation of probabilistic approaches to comparison. Our ability to explore the linkages between forced changes in the mean climate and short-term climate variability is limited by the lack of global-scale syntheses of high-resolution records that extends beyond the past two millennia. Again, a community focus on producing such syntheses would be worthwhile. But the likely complexity of the seasonal changes in climate in the geological past, coupled with the known complexity of the controls on biological systems, means there will always be large uncertainties associated with statistical reconstructions. An emphasis on developing and using process-based models of a range of palaeoenvironmental sensors is also required to improve climate-model evaluation. The application of process-based models will facilitate a more systematic exploitation of existing syntheses of qualitative data. Many of the synthetic products are out-of-date and do not include sites published in the last decade or so, and thus a community effort to update these data sets would be useful.

Equilibrium time-slice simulations have been the focus of climate modeling for many decades, and this type of simulation will still be a focus of the next phase of the Climate Model Intercomparison Project (CMIP6: Meehl et al., 2014). However, many features of the climate system cannot be examined using equilibrium simulations. As the CMIP5 Last Millennium experiment has demonstrated, long transient simulations are now possible; and indeed transient simulations of both the deglaciation and Holocene using the same models that are used for future climate projections are planned within PMIP. Evaluation of transient simulations poses new and as yet unexplored issues for data synthesis and data-model comparison.

Acknowledgements

Alayne Street-Perrott has been a pioneer in developing ways to bring palaeoenvironmental observations and palaeoclimate models together, thus starting a continuing process for palaeoclimate science to transition from description and story-telling to hypothesis testing, model evaluation and diagnosis, and ultimately policy relevance. Alayne was one of the first to create and advocate large-scale data syntheses, and to recognize the importance of plant physiological processes (including the effects of atmospheric CO_2 concentration) on plants and ecosystems in the past. SPH acknowledges Alaynes's profound influence on her research career, and in particular thanks Alayne for inducting her into the COHMAP project. We thank our PMIP colleagues for contributing to the PMIP simulation archive and to benchmark data syntheses, as well as for discussions of the PMIP analyses. This paper is a contribution to the AXA Chair Programme in Biosphere and Climate Impacts and the Imperial College initiative on Grand Challenges in Ecosystems and the Environment.

References

Abe-Ouchi A, Saito F, Kageyama M, Braconnot PB, Harrison SP, Lambeck K, Otto-Bleisner BL, Peltier WR, Tarasov L, Peterschmidt J-Y, Takahashi K. 2015. Ice-sheet configuration in the PMIP3/CMIP5 Last Glacial Maximum experiments. *Geoscientific Model Development Discussions* **8** : 4293-4336.

Acosta Navarro JC, Smolander S, Struthers H, Zorita E, Ekman AML, Kaplan JO, Guenther A, Arneth A, Riipinen I. 2014. Global emissions of terpenoid VOCs from terrestrial vegetation in the last millennium. *Journal of Geophysical Research Atmospheres* **119** : 6867-6885.

Allan RP. 2009. Examination of relationships between clear-sky longwave radiation and aspects of the atmospheric hydrological cycle in climate models, reanalysis, and observations. *Journal of Climate* **22** : 3127–3145.

Alyea FN. 1972. Numerical simulation of an ice age paleoclimate. *Atmospheric Science Paper* **193**, Colorado State University.

Argus DF, Peltier WR. 2010. Constraining models of postglacial rebound using space geodesy: a detailed assessment of model ICE-5G (VM2) and its relatives. *Geophysical Journal International* **181** : 697–723.

Atkinson TC, Briffa KR, Coope GR. 1987. Seasonal temperatures in Britain during the last 22,000 years, reconstructed using beetle remains. *Nature* **325** : 587–593.

Aumont O, Maier-Reimer E, Blain S, Monfray P. 2003. An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles* **17** : 1060.

Bakker P, Stone EJ, Charbit S, Gröger M, Krebs-Kanzow U, Ritz SP, Varma V, Khon V, Lunt DJ, Mikolajewicz U, Prange M, Renssen H, Schneider B, Schulz M. 2013. Last interglacial temperature evolution – a model inter-comparison. *Climate of the Past* **9** : 605-619.

Bartlein PJ, Webb T III, Fleri E. 1984. Holocene climatic change in the northern Midwest: pollen-derived estimates. *Quaternary Research* **22** : 361–374.

Bartlein PJ, Edwards MD, Shafer SL, Barker ED. 1995. Calibration of radiocarbon ages and the interpretation of paleoenvironmental records. *Quaternary Research* **44** : 417–424.

Bartlein PJ, Harrison SP, Brewer S, Connor S, Davis BAS, Gajeweski K, Guiot J, Harrison-Prentice TI, Henderson A, Peyron O, Prentice IC, Scholze M, Seppa H, Shuman B, Sugita S, Thompson RS, Viau AE, Williams J, Wu H. 2011. Pollen- based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Climate Dynamics* **37** : 775–802.

Bauer E, Ganopolski A. 2014. Sensitivity simulations with direct shortwave radiative forcing by aeolian dust during glacial cycles. *Climate of the Past* **10** : 1333–1348.

Bennett KD. 1994. Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. *The Holocene* **4**: 337–348.

Bennett KD, Fuller JL. 2002. Determining the age of the mid-Holocene *Tsuga canadensis* (hemlock) decline, eastern North America. *The Holocene* **14** : 421–429.

Berger AL. 1978: Long-term variations of daily insolation and Quaternary climatic changes. *Journal of Atmospheric Sciences* **35** : 2362–2367.

Berger A, Brandefelt J, Nilsson J. 2013. 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.

Bigelow NH, Brubaker LB, Edwards ME, Harrison SP, Prentice IC, Anderson PM, Andreev AA, Bartlein PJ, Christensen TR, Cramer W, Kaplan JO, Lozhkin AV, Matveyeva NV, Murray DF, McGuire AD, Razzhivin VY, Ritchie JC, Smith B, Walker DA, Gayewski K, Wolf V, Holmqvist BH, Igarashi Y, Kremenetskii K, Paus A, Pisaric MFJ, Volkova VS. 2003. Climatic change and Arctic ecosystems I. Vegetation changes north of 55°N between the last glacial maximum, mid-Holocene, and present. *Journal of Geophysical Research-Atmosphere* **108** : D19, 8170.

Blaauw M. 2010. Methods and code for 'classical' age-modelling of radiocarbon sequences. *Quaternary Geochronology* **5** : 512-518.

Blaauw M, Christen JA. 2005. Radiocarbon peat chronologies and environmental change. *Applied Statistics* **54** : 805-816.

Blaauw M, Christen JA. 2011. Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis* **6** : 457-474.

Blaauw M, Heuvelink GBM, Mauquoy D, van der Plicht J, van Geel B. 2003. A numerical approach to ^{14}C wiggle-match dating of organic deposits: best fits and confidence intervals. *Quaternary Science Reviews* **22** : 1485–1500.

Bond G, Broecker W, Johnsen S, McManus J, Labeyrie L, Jouzel J, Bonani G. 1993. Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365** : 143–147.

Bopp L, Kohfeld KE, Le Quere C, Aumont O. 2003. Dust impact on marine biota and atmospheric CO_2 during glacial periods. *Paleoceanography* **18 :** 1046.

Boreux J-J, Pesti G, Duckstein L, Nicolas J. 1997. Age model estimation in paleoclimatic research: fuzzy regression and radiocarbon uncertainties. *Palaeogeography, Palaeoclimatology, Palaeoecology* **128** : 29–37.

Bothe O, Jungclaus JH, Zanchettin D, Zorita E. 2013. Climate of the last millennium: ensemble consistency of simulations and reconstructions. *Climate of the Past* **9** : 1089-1110.

Boucher E, Guiot J, Hatté C, Daux V, Danis P-A, Dussouillez P. 2014. An inverse modeling approach for tree-ring-based climate reconstructions under changing atmospheric CO_2 concentrations. *Biogeosciences* **11** : 3245–3258.

Braconnot P, Harrison SP, Otto-Bliesner B, Abe-Ouchi A, Jungclaus J, Peterschmitt J-Y. 2011. The Paleoclimate Modeling Intercomparison Project contribution to CMIP5. *CLIVAR Exchanges Newsletter* **56** : 15-19.

Braconnot P, Otto-Bliesner B, Harrison SP, Joussaume S, Peterschmitt J-Y, Abe-Ouchi A, Crucifix M, Driesschaert E, Fichefet T, Hewitt CD, Kageyama M, Kitoh A, Laîné A, Loutre M-F, Marti O, Merkel U, Ramstein G, Valdes P, Weber L, Yu Y, Zhao Y. 2007a. 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.

Braconnot P, Otto-Bliesner B, Harrison SP, Joussaume S, Peterschmitt J-Y, Abe-Ouchi A, Crucifix M, Driesschaert E, Fichefet T, Hewitt CD, Kageyama M, Kitoh A, Laîné A, Loutre M-F, Marti O, Merkel U, Ramstein G, Valdes P, Weber L, Yu Y, Zhao Y. 2007b. Results of PMIP2 coupled simulations of the mid-Holocene and Last Glacial Maximum - Part 2: feedbacks with emphasis on the location of the ITCZ and mid- and high latitudes heat budget. *Climate of the Past* **3** : 279–296.

Braconnot P, Harrison SP, Kageyama M, Bartlein PJ, Masson-Delmotte V, Abe-Ouchi A, Otto-Bliesner B, Zhao Y. 2012. Evaluation of climate models using palaeoclimatic data. *Nature Climate Change* **2** : 417-424.

Bradley RS. 2014. *Paleoclimatology: Reconstructing Climates of the Quaternary*. Academic Press/Elsevier, Amsterdam, 3rd Edition, 675pp.

Bragg F, Prentice IC, Harrison SP, Foster PN, Eglinton G, Rommerskirchen F, Rullkötter J. 2013. n-Alkane stable isotope evidence for CO_2 as a driver of vegetation change. *Biogeosciences* **10** : 2001–2010.

Brewer S, Guiot J, Torre F. 2007. Mid-Holocene climate change in Europe: a data-model comparison. *Climate of the Past* **3** : 499-512.

Briffa KR, Osborn TJ, Schweingruber FH, Jones PD, Shiyatov SG, Vaganov EA. 2002. Tree-ring width and density data around the Northern Hemisphere: part 2, spatio-temporal variability and associated climate patterns. *The Holocene* **12** : 759-789.

Bronk Ramsey C. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51** : 337-360.

Brücher T, Brovkin V, Kloster S, Marlon J.R, Power MJ. 2014. Comparing modelled fire dynamics with charcoal records for the Holocene. *Climate of the Past* **10** : 811-824.

Charman DJ, Beilman DW, Blaauw M, Booth RK, Brewer S, Chambers FM, Christen JA, Gallego-Sala A, Harrison SP, Hughes PDM, Jackson ST, Korhola A, Mauquoy D, Mitchell FJG, Prentice IC, van der Linden M, de Vleeschouwer F, Yu ZC, Alm J, Bauer IE, Corish YMC, Garneau M, Hohl V, Huang Y, Karofeld E, Le Roux G, Moschen R, Nichols JE, Nieminen T, McDonald GM, Phadtare NR, Rausch N, Shotyk W, Sillasoo U, Swindles GT, Tuittila ES, Ukonmaanaho L, Väliranta M, van Bellen S, van Geel B, Vitt DH, Zhao Y. 2013. Climate-driven changes in peatland carbon accumulation during the last millennium. *Biogeosciences* **10** : 929-944.

Chavaillaz Y, Codron F, Kageyama M. 2013. Southern westerlies in LGM and future (RCP4.5) climates. *Climate of the Past* **9** : 517-524.

Cheddadi R, Yu G, Guiot J, Harrison SP, Prentice IC. 1997. The climate of Europe 6000 years ago. *Climate Dynamics* **13** : 1-9.

Christen JA, Clymo RS, Litton CD. 1995. A Bayesian approach to the use of ^{14}C dates in the estimation of the age of peat. *Radiocarbon* **37** : 431–442.

Ciais P, Tagliabue A, Cuntz M, Bopp L, Scholze M, Hoffmann G, Lourantou A, Harrison SP, Prentice IC, Kelley DI, Koven C, Piao SL. 2011. Large inert carbon pool in terrestrial ecosystems during the last glacial maximum. *Nature Geoscience* **5** : 74-79.

Claussen M, Gayler V. 1997. The greening of the Sahara during the mid-Holocene: results of an interactive atmosphere-biome model. *Global Ecology and Biogeography Letters* **6** : 369-377.

CLIMAP Project Members.1976. The surface of the Ice-age Earth. *Science* **191** : 1131–1137.

CLIMAP Project Members. 1981. Seasonal reconstructions of the of the Earth's surface at the Last Glacial Maximum. *Geological Society of America Map and Chart Series* **MC-36**, 18 pp.

Coe MT, Harrison SP. 2002. The water balance of northern Africa during the mid-Holocene: an evaluation of the 6ka BP PMIP experiments. *Climate Dynamics* **19** : 155-166.

COHMAP Members. 1988. Climatic changes of the last 18,000 years: observations and model simulations. *Science* **241** : 1043-1052.

Cowling SA, Sykes MT. 1999. Physiological significance of low atmospheric CO_2 for plant–climate interactions. *Quaternary Research* **52** : 237–242.

Craig H. 1957. The natural distribution of radiocarbon and the exchange time of carbon dioxide between atmosphere and sea. *Tellus* **9** : 1-17.

Crowley TJ, Zielinski G, Vinther B, Udisti R, Kreutz K, J. Cole-Dai J, Castellano E. 2008. Volcanism and the Little Ice Age. *PAGES Newsletter* **16** : 22-23.

Crucifix M, Braconnot P, Harrison SP, Otto-Bliesner B. 2005. PMIP2 fosters evaluation of state-of-the-art climate models with palaeoclimatic data. *EOS* **86** : 264-265.

Dallmeyer A, Claussen M, Otto J. 2010. Contribution of oceanic and vegetation feedbacks to Holocene climate change in Central and Eastern Asia. *Climate of the Past* **6** : 195-218.

Dallmeyer A, Claussen M, Fischer N, Haberkorn K, Wagner S, Pfeiffer M, Jin L, Khon V, Wang Y, Herzschuh U. 2015. The evolution of sub-monsoon systems in the Afro-Asian monsoon region during the Holocene – comparison of different transient climate model simulations. *Climate of the Past* **11** : 305-326.

Davis BAS, Brewer S, Stevenson AC, Guiot J, Juggins S. 2003. The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews* **22** : 1701–1716.

Delaygue G, Bard E. 2011. An Antarctic view of Beryllium-10 and solar activity for the past millennium. *Climate Dynamics* **36** : 2201-2218.

Daniau A-L, Harrison SP, Bartlein PJ. 2010. Fire regimes during the last glacial. *Quaternary Science Reviews* **29** : 2918-2930.

Daniau A-L, Bartlein PJ, Harrison SP, Prentice IC, Brewer S, Friedlingstein P, Harrison-Prentice TI, Inoue J, Marlon JR, Mooney S, Power MJ, Stevenson J, Tinner W, Andrič M, Atanassova J, Behling H, Black M, Blarquez O, Brown KJ, Carcaillet C, Colhoun E, Colombaroli D, Davis BAS, D'Costa D, Dodson J, Dupont L, Eshetu Z, Gavin DG, Genries A, Gebru T, Haberle S, Hallett DJ, Horn S, Hope G, Katamura F, Kennedy L, Kershaw P, Krivonogov S, Long C, Magri D, Marinova E, McKenzie GM, Moreno PI, Moss P, Neumann FH, Norström E, Paitre C, Rius D, Roberts N, Robinson G, Sasaki N, Scott L, Takahara H, Terwilliger V, Thevenon F, Turner RB, Valsecchi VG, Vannière B, Walsh M, Williams N, Zhang Y. 2012. Predictability of biomass burning in response to climate changes. *Global Biogeochemical Cycles* **26** : doi:10.1029/2011GB004249.

D'Arrigo R, Wilson R, Liepert B, Cherubini P. 2008. On the "divergence problem" in northern forests: a review of the tree-ring evidence and possible causes. *Global and Planetary Change* **60** : 289–305.

de Noblet-Ducoudré N, Claussen M, Prentice C. 2000. Mid-Holocene greening of the Sahara: first results of the GAIM 6000 year BP Experiment with two asynchronously coupled atmosphere/biome models. *Climate Dynamics* **16** : 643-659.

Dyke AS. 2004. An outline of North American deglaciation with emphasis on central and northern Canada. *Developments in Quaternary Science* **26** : 374–424.

Dyke AS, Prest VK. 1987. Late Wisconsinan and Holocene history of the Laurentide Sheet. *Géographie physique et Quaternaire* **41** : 237–263.

Ehlers J, Gibbard PL, Hughes PD (Eds). 2011. *Quaternary Glaciations – Extent and Chronology: A Closer Look.* Elsevier, Amsterdam.

Emile-Geay J, Seager R, Cane MA, Cook ER, Haug GH. 2008. Volcanoes and ENSO over the past millennium. *Journal of Climate* **21 :** 3134–3148.

Emile-Geay J, Cobb KM, Carré M, Braconnot P, Leloup J, Zhou Y, Harrison SP, Corrège T, Collins M, Driscoll R, Elliot M, McGregor HV, Schneider B, Tudhope A. 2015. Holocene constraints on tropical Pacific dynamics. *Nature Geoscience* doi:10.1038/ngeo2608.

EPICA Community Members. 2004. Eight glacial cycles from an Antarctic ice core. *Nature* **429** : 623-628.

Esper J, Cook ER, Schweingruber FH. 2002. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* **295** : 2250-2253.

Evans MN, Reichert BK, Kaplan A, Anchukaitis KJ, Vaganov EA, Hughes MK, Cane MA. 2006. A forward modeling approach to paleoclimatic interpretation of tree-ring data. *Journal of Geophysical Research* **111 :** G03008.

Fallah B, Cubasch U. 2015. A comparison of model simulations of Asian mega-droughts during the past millennium with proxy reconstructions. *Climate of the Past* **11** : 253–263.

Feng S, Hu Q, Wu Q, Mann ME. 2013. A gridded reconstruction of warm season precipitation for Asia spanning the past half millennium. *Journal of Climate* **26 :** 2192–2204.

Fernández-Donado L, González-Rouco JF, Raible CC, Ammann CM, Barriopedro D, García-Bustamante E, Jungclaus JH, Lorenz SJ, Luterbacher J, Phipps SJ, Servonnat J, Swingedouw D, Tett SFB, Wagner S, Yiou P, Zorita E. 2013. Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium *Climate of the Past* **9** : 393-421.

Flato G, Marotzke J, Abiodun B, Braconnot P, Chou SC, Collins W, Cox P, Driouech F, Emori S, Eyring V, Forest C, Gleckler P, Guilyardi E, Jakob C, Kattsov V, Reason C, Rummukainen M. 2013. 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.* Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA; 741-866.

Fletcher BJ, Brentnall SJ, Anderson CW, Berner RA, Beerling DJ. 2007. Atmospheric carbon dioxide linked with Mesozoic and early Cenozoic climate change. *Nature Geosciences* **1** :

43-48.

Foster GL, Hönisch B, Paris G, Dwyer GS, Rae JWB, Elliott T, Gaillardet J, Hemming NG, Louvat P, Vengosh A. 2013. Interlaboratory comparison of boron isotope analyses of boric acid, seawater and marine $CaCO_3$ by MC-ICPMS and NTIMS. *Chemical Geology* **358** : 1–14.

Franke J, Paul A, Schulz M. 2008. Modeling variations of marine reservoir ages during the last 45 000 years. *Climate of the Past* **4** : 125–136.

Fréchette B, de Vernal A, Guiot J, Wolfe AP, Miller GH, Fredskild B, Kerwin MW, Richard PJH. 2008. Methodological basis for quantitative reconstruction of air temperature and sunshine from pollen assemblages in Arctic Canada and Greenland. *Quaternary Science Reviews* **27** : 1197-1216.

Friedlingstein P, Andrew RM, Rogelj J, Peters GP, Canadell JG, Knutti R, Luderer G, Raupach MR, Schaeffer M, van Vuuren DP. 2014. Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience* **7** : 709-715.

Gagen M, Finsinger W, Wagner R, McCarrol D, Loader N, Robertson I, Jalkanen R, Young G, Kirchhefer A. 2011. Evidence of changing intrinsic water use efficiency under rising atmospheric CO_2 concentrations in boreal Fennoscandia from subfossil leaves and tree ring $\delta^{13}C$ ratios. *Global Change Biology* **17** : 1064–1072.

Garreta V, Miller PA, Guiot J, Hély C, Brewer S, Sykes MT, Litt T. 2010. A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model. *Climate Dynamics* **35** : 371–389.

Gates WL. 1976. Modelling the ice-age climate. *Science* **191** : 1138– 1144.

Ganopolski A, Kubatzki C, Claussen M, Brovkin V, Petoukhov V. 1998. The influence of vegetation-atmosphere-ocean interaction on climate during the mid-Holocene. *Science* **280** : 1916-1919.

Gao C, Robock A, Ammann C. 2008. Volcanic forcing of climate over the last 1500 years: An improved ice-core based index for climate models. *Journal of Geophysical Research* **113** : D2311.

Giesecke T, Davis B, Brewer S, Finsinger W, Wolters S, Blaauw M, de Beaulieu J-L, Binney H, Fyfe RM, Gaillard M-J, Gil-Romera G, van der Knaap WO, Kuneš P, Kühl N, van Leeuwen JFN, Leydet M, Lotter AF, Ortu E, Semmler M, Bradshaw RHW. 2014. Towards mapping the late Quaternary vegetation change of Europe. *Journal of Vegetation History and Archaeobotany* **23** : 75-86.

Gleckler PJ, Taylor KE, Doutriaux, C. 2008. Performance metrics for climate models. *Journal of Geophysical Research* **113** : D06104.

Godwin H. 1951. Comments on radiocarbon dating samples from the British Isles. *American Journal of Science* **249** : 301-307.

Gonzales LM, Williams JW, Kaplan JO. 2008. Variations in leaf area index in northern and eastern North America over the past 21,000 years: a data-model comparison. *Quaternary Science Reviews* **27** : 1453–1466.

Goosse H, Crowley TJ, Zorita E, Amman CM, Renssen H, Dreisschaert E. 2005. Modelling the climate of the last millennium: What causes the differences between simulations? *Geophysical Research Letters* **32** : L06710.

Goosse, H., Crespin E, Dubinkina S, Loutre MF, Renssen H, Sallaz-Damaz Y, Shindell D. 2012. The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly". *Climate Dynamics* **39** : 2847–2866.

Greaves M, Caillon N, Rebaubier H, Bartoli G, Cacho I, Clarke L, Delaney M, de Menocal P, Dutton A, Eggins S, Elderfield H, Goddard E, Green D, Groeneveld J, Hastings D, Hathorne E, Kimoto K, Klinkhammer G, Labeyrie L, Lea DW, Marchitto T, Martínez-Botí MA, Mortyn PG, Ni Y, Nuernberg D, Paradis G, Pena L, Quinn T, Rosenthal Y, Russell A, Sagawa T, Sosdian S, Stott L, Tachikawa K, Tappa E, Thunell R, Wilson PA. 2008. Interlaboratory comparison study of calibration standards for foraminiferal Mg/Ca thermometry. *Geochemistry, Geophysics, Geosystems* **9** : Q08010.

Grichuk VP. 1969. An attempt to reconstruct certain elements of the climate of the northern hemisphere in the Atlantic period of the Holocene. In *Golotsen,* Neishtadt MI (ed). Nauka, Moscow; 41–57.

Guiot J. 1987. Late Quaternary climatic change in France estimated from multivariate pollen time series. *Quaternary Research* **28** : 100–118.

Guiot J. 1990. Methodology of the last climatic cycle reconstruction in France from pollen data. *Palaeogeography, Palaeoclimatology, Palaeoecology* **80** : 49–69.

Guiot J, Torre F, Jolly D, Peyron O, Boreux JJ, Cheddadi R. 2000. Inverse vegetation modeling by Monte Carlo sampling to reconstruct palaeoclimates under changed precipitation seasonality and CO_2 conditions: application to glacial climate in Mediterranean region. *Ecological Modelling* **127** : 119–140.

Guiot J, Nicault A, Rathgeber C, Edouard JL, Guibal F, Pichard G, Till C. 2005. Last-millennium summer-temperature variations in western Europe based on proxy data. *The Holocene* **15** : 489-500.

Guiot J, Wu H, Jiang WY, Luo YL. 2008. East Asian Monsoon and paleoclimatic data analysis: a vegetation point of view. *Climate of the Past* **4** : 137–145.

Gyllencreutz G, Mangerud J, Svendsen J-I, Lohne Ø. 2007. DATED - a GIS-based reconstruction and dating database of the Eurasian deglaciation. In *Applied Quaternary Research in the Central Part of Glaciated Terrain,* Johansson P, Sarala P (eds). Geological Survey of Finland; 113–120.

Hall A, Qu X. 2006. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters* **33** : L03502.

Hargreaves JC, Annan JD. 2014. Can we trust climate models? *WIREs Climate Change.* doi: 10.1002/wcc.288.

Hargreaves JC, Annan JD, Ohgaito R, Paul A, Abe-Ouchi A. 2013. Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene. *Climate of the Past* **9** : 811– 823.

Harrison SP, Bartlein PJ. 2012. Records from the past, lessons for the future: what the palaeo-record implies about mechanisms of global change. In *The Future of the World's Climates,* Henderson-Sellers A, McGuffie K (eds). Elsevier, Amsterdam; 403-436.

Harrison SP, Prentice IC. 2003. Climate and CO_2 controls on global vegetation distribution at the last glacial maximum: analysis based on palaeovegetation data, biome modelling and palaeoclimate simulations *Global Change Biology* **9** : 983-1004.

Harrison SP, Jolly D, Laarif F, Abe-Ouchi A, Dong B, Herterich K, Hewitt C, Joussaume S, Kutzbach JE, Mitchell J, de Noblet N, Valdes P. 1998. Intercomparison of simulated global vegetation distribution in response to 6 kyr BP orbital forcing. *Journal of Climate* **11** : 2721-2742.

Harrison SP, Marlon J, Bartlein PJ. 2010. Fire in the Earth System. In *Changing Climates, Earth Systems and Society*, Dodson J (ed.). Springer-Verlag; Amsterdam; 21-48.

Harrison SP, Prentice IC, Sutra J-P, Barboni D, Kohfeld KE, Ni J. 2010. Ecophysiological and bioclimatic foundations for a global plant functional classification. *Journal of Vegetation Science* **21** : 300-317.

Harrison SP, Bartlein PJ, Brewer S, Prentice IC, Boyd M, Hessler I, Holmgren K, Izumi K, Willis K. 2014. Model benchmarking with glacial and mid-Holocene climates. *Climate Dynamics* **43** : 671-688.

Harrison SP, Bartlein PJ, Izumi K, Li G, Annan J, Hargreaves J, Braconnot PB, Kageyama M. 2015. Implications of evaluation of CMIP5 palaeosimulations for climate projections. *Nature Climate Change* **5** : 735-743.

Hatté C, Rousseau D, Guiot J. 2009. Climate reconstruction from pollen and d13C records using inverse vegetation modelling–implication for past and future climates. *Climate of the Past* **5**

: 147–156.

Haywood AM, Dowsett HJ, Otto-Bliesner B, Chandler MA, Dolan AM, Hill DJ, Lunt DJ, Robinson MM, Rosenbloom N, Salzmann U, Sohl LE. 2010. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 1). *Geoscientific Model Development* **3** : 227-242.

Haywood AM, Dowsett HJ, Robinson MM, Stoll DK, Dolan AM, Lunt DJ, Otto-Bliesner B, Chandler MA. 2011. Pliocene Model Intercomparison Project (PlioMIP): experimental design and boundary conditions (Experiment 2). *Geoscientific Model Development* **4** : 571-577.

Heegaard E. 2003. Age–depth routine for R. http://www.bio.uu.nl/Bpalaeo/Congressen/Holivar/Literature Holivar2003.htm

Hegerl GC, Luterbacher J, Gonzalez-Ruoco F, Tett SFB, Xoplaki E. 2011.Influence of human and natural forcing on European seasonal temperatures *Nature Geosciences* **4** : 99–103.

Heiri O, Lotter AF, Hausmann S, Kienast F. 2003. A chironomid- based Holocene summer air temperature reconstruction from the Swiss Alps. *The Holocene* **13** : 477–484.

Hessler I, Harrison SP, Kuchera M, Waelbroeck C, Chen M-T, Anderson C, de Vernal A, Fréchette B, Cloke-Hayes A, Londeix L. 2014. Implication of methodological uncertainties for mid-Holocene sea surface temperature reconstructions. *Climate of the Past* **10** : 2237-2252.

Hill DJ, Haywood AM, Lunt DJ, Hunter SJ, Bragg FJ, Contoux C, Stepanek C, Sohl L, Rosenbloom NA, Chan WL, Kamae Y, Zhang Z, Abe-Ouchi A, Chandler MA, Jost A, Lohmann G, Otto-Bliesner BL, Ramstein G, Ueda H. 2014. Evaluating the dominant components of warming in Pliocene climate simulations. *Climate of the Past* **10** : 79-90.

Holloway MD, Sime LC, Singarayer JS, Tindall JC, Valdes PJ. 2016. Reconstructing paleosalinity from $\delta^{18}O$: Coupled model simulations of the Last Glacial Maximum, Last Interglacial and Late Holocene. *Quaternary Science Reviews* **131** : 350-364.

Huntley B, Prentice IC. 1988. July temperatures in Europe from pollen data 6000 years before present. *Science* **241** : 687–690.

Hurtt GC, Frolking S, Fearon MG, Moore B, Shevliakova E, Malyshev S, Pacala SW, Houghton RA. 2006. The underpinnings of land-use history: three centuries of global gridded landuse transitions, wood harvest activity, and resulting secondary lands. *Global Change Biology* **12** : 1–22.

Hutson WH, Prell WL. 1980. A paleoecological transfer function, FI-2, for Indian Ocean planktonic foraminifera. *Journal of Paleontology* **54** : 381-399.

Imbrie J, Kipp NG. 1971. A new micropaleontological method for quantitative paleoclimatology:

Application to a Late Pleistocene Caribbean core. In *The Late Cenozoic Glacial Ages,* Turekian K (ed). Yale University Press, New Haven, Connecticut; 71-181.

Izumi K, Bartlein PJ, Harrison SP. 2013. Consistent behaviour of the climate system in response to past and future forcing. *Geophysical Research Letters* **40** : 1-7.

Izumi K, Bartlein PJ, Harrison SP. 2014. Energy-balance mechanisms underlying consistent large-scale temperature responses in warm and cold climates. *Climate Dynamics* **44** : 3111-3127.

Jackson CS, Sen MK, Huerta G, Deng Y, Bowman KP. 2008. Error reduction and convergence in climate prediction. *Journal of Climate* **21** : 6698–6709.

Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb T III, Williams J, Hansen BCS. 2000. Vegetation and environment in unglaciated eastern North America during the last glacial maximum. *Quaternary Science Reviews* **19** : 489–508.

Jolly D, Haxeltine A, 1997. Effect of low glacial atmospheric CO_2 on tropical African montane vegetation. *Science* **276** : 786–788.

Jones PD, Briffa KR, Barnett TP, Tett SFB. 1998. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with General Circulation Model control-run temperatures. *The Holocene* **8** : 455–471.

Jones VJ, Leng MJ, Solovieva N, Sloane HJ, Tarasov P. 2004. Holocene climate of the Kola Peninsula; evidence from the oxygen isotope record of diatom silica. *Quaternary Science Reviews* **23** : 833–839.

Joussaume S, Taylor KE. 2000. The Paleoclimate Modeling Intercomparison Project. In *Proceedings of the third PMIP Workshop*, Braconnot P (ed). WCRP-111, WMO/TD-1007.

Joussaume S, Taylor KE, Braconnot P, Mitchell JFB, Kutzbach JE, Harrison SP, Prentice IC, Broccoli AJ, Abe-Ouchi A, Bartlein PJ, Bonfils C, Dong B, Guiot J, Herterich K, Hewitt CD, Jolly D, Kim JW, Kislov A, Kitoh A, Loutre MF, Masson V, McAvaney B, McFarlane N, de Noblet N, Peltier WR, Peterschmitt JY, Pollard D, Rind D, Royer JF, Schlesinger ME, Syktus J, Thompson S, Valdes P, Vettoretti G, Webb RS, Wyputta U. 1999. Monsoon changes for 6000 years ago: results of 18 simulations from the Paleoclimate Modeling Intercomparison Project (PMIP). *Geophysical Research Letters* **26** : 859-862.

Juckes MN, Allen MR, Briffa KR, Esper J, Hegerl GC, Moberg A, Osborn TJ, Weber SL, 2007. Millennial temperature reconstruction intercomparison and evaluation. *Climate of the Past* **3** : 591.

Kageyama M, Paul A, Roche DM, Van Meerbeeck CJ. 2010. Modelling glacial climatic millennial-scale variability related to changes in the Atlantic meridional overturning circulation: a review. *Quaternary Science Reviews* **29** : 2931-2956.

Kaplan JO, Bigelow NH, Bartlein PJ, Christensen TR, Cramer W, Harrison SP, Matveyeva NV, McGuire AD, Murray DF, Prentice IC, Razzhivin VY, Smith B, Walker DA, Anderson PM, Andreev AA, Brubaker LB, Edwards ME, Lozhkin AV. 2003. Climate change and Arctic ecosystems II: Modeling, palaeodata-model comparisons, and future projections. *Journal of Geophysical Research-Atmosphere* **108** : No. D19, 8171.

Kelley DI, Harrison SP. 2014. Enhanced Australian carbon sink despite increased wildfire during the 21st century. *Environmental Research Letters* **9** : 104015.

Kim J.H. 2004. GHOST global database for alkenone-derived Holocene sea-surface temperature records. http://www.pangaea. de/Projects/GHOST/.

Kirtman B, Power SB, Adedoyin JA, Boer GJ, Bojariu R, Camilloni I, Doblas-Reves FJ, Fiore AM, Kimoto M, Meehl GA, Prather M, Sarr A, Schar C, Sutton R, van Oldenborgh GJ, Vecchi G, Wang HJ. 2013. Near-term climate change: projections and predictability. 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*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA; 953-1028.

Kloster S, Mahowald NM, Randerson JT, Lawrence PJ. 2012. The impacts of climate, land use, and demography on fires during the 21st century simulated by CLM-CN. *Biogeosciences* **9** : 509-525.

Knutti R. 2010. The end of model democracy? *Climate Change* **102** : 395–404.

Kohfeld KE, Harrison SP. 2000. How well can we simulate past climates? Evaluating the models using global palaeoenvironmental data sets. *Quaternary Science Reviews* **19** : 321-346.

Kohfeld KE, Harrison SP. 2001. DIRTMAP: The geological record of dust. *Earth Science Reviews* **54** : 81-114.

Kok JF, Albani S, Mahowald NM, Ward DS. 2014. An improved dust emission model – Part 2: Evaluation in the Community Earth System Model, with implications for the use of dust source functions. *Atmospheric Chemistry and Physics* **14** : 13043-13061.

Kutzbach JE, Street-Perrott FA. 1985. Milankovitch forcingof fluctuations in the level of tropical lakes from 18 to 0 kya BP. *Nature* **317** : 130–134.

Kutzbach JE, Guetter PJ. 1986. The influence of changing orbital parameters and surface boundary conditions for the past 18,000 years. *Journal of Atmospheric Sciences* **43** : 1726–1759.

Lambeck K, Purcell A, Zhao J, Svensson N-O. 2010. The Scandinavian Ice Sheet: from MIS 4 to the end of the Last Glacial Maximum. *Boreas* **39** : 410–435.

Lawrence DM, Oleson KW, Flanner MG, Thornton PE, Swenson SC, Lawrence PJ, Zeng X, Yang ZL, Levis S, Sakaguchi K, Bonan GB, Slater AG. 2011. Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *Journal of Advances in Modelling Earth Systems* **3**, DOI: 10.1029/2011MS000045.

Leduc G, Schneider R, Kim JH, Lohmann G. 2010. Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quaternary Science Reviews* **29** : 989–1004.

Le Quéré C, Harrison SP, Prentice IC, Buitenhuis ET, Aumont O, Bopp L, Claustre H, Da Cunha LC, Geider R, Giraud X, Klaas C, Kohfeld KE, Legendre L, Manizza M, Platt T, Rivkin RB, Sathyendranath S, Uitz J, Watson AJ, Wolf-Gladow D. 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* **11** : 2016–2040.

Li G, Harrison SP, Bartlein PJ, Izumi K, Prentice IC. 2013. Precipitation scaling with temperature in warm and cold climates: an analysis of CMIP5 simulations. *Geophysical Research Letters* **40 :** 4018-4024.

Li G, Harrison SP, Prentice IC, Falster D. 2014. Interpretation of tree-ring data with a model for primary production, carbon allocation and growth. *Biogeosciences* **11** : 6711-6724.

Ljungqvist FC, Krusic PJ, Brattström G, Sundqvist HS. 2012. Northern Hemisphere temperature patterns in the last 12 centuries. *Climate of the Past* **8** : 227.

Lohmann G, Pfeiffer M, Laepple T, Leduc G, Kim J-H. 2013. A model-data comparison of the Holocene global sea surface temperature evolution. *Climate of the Past* **9** : 1807–1839.

Liu Z, Otto-Bliesner BL, He F, Brady EC, Tomas R, Clark PU, Carlson AE, Lynch-Stieglitz J, Curry W, Brook E, Erickson D, Jacob R, Kutzbach J, Cheng J. 2009. Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming. Science **325** : 310-314.

Lunt DJ, Abe-Ouchi A, Bakker P, Berger A, Braconnot P, Charbit S, Fischer N, Herold N, Jungclaus JH, Khon VC, Krebs-Kanzow U, Langebroek PM, Lohmann G, Nisancioglu KH, Otto-Bliesner BL, Park W, Pfeiffer M, Phipps SJ, Prange M, Rachmayani R, Renssen H, Rosenbloom N, Schneider B, Stone EJ, Takahashi K, Wei W, Yin Q, Zhang ZS. 2013. A multi-model assessment of last interglacial temperatures. *Climate of the Past* **9** : 699-717.

Lunt DJ, Dunkley Jones T, Heinemann M, Huber M, LeGrande A, Winguth A, Loptson C, Marotzke J, Roberts CD, Tindall J, Valdes P, Winguth C. 2012. A model–data comparison for a multi-model ensemble of early Eocene atmosphere–ocean simulations: EoMIP. *Climate of the Past* **8** : 1717–1736.

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H. 2004 European seasonal and annual temperature variability, trends and extremes since 1500. *Science* **303 :** 1499-1503.

Maher BA, Kohfeld K, Leedal DT. 2014. 'DIRTMAP' Version 4, LGM and late Holocene Aeolian Fluxes from Ice Cores, Marine Sediment Traps, Marine Sediments and Loess Deposits. http://www.lancaster.ac.uk/lec/sites/dirtmap/hw.html

Mahowald NM, Yoshioka M, Collins WD, Conley AJ, Fillmore DW, Danielle B. Coleman DB. 2006. Climate response and radiative forcing from mineral aerosols during the last glacial maximum, pre-industrial, current and doubled-carbon dioxide climates. *Geophysical Research Letters* **33** : L20705.

Manabe S, Hahn DG. 1977. Simulation of the tropical climate of an ice age. *Journal of Geophysical Research* **82** : 3889–3911.

Mangerud J, Goehring BM, Lohne ØS, Svendsen JI, Gyllencreutz R. 2013. Collapse of marine-based outlet glaciers from the Scandinavian Ice Sheet. *Quaternary Science Reviews* **67** : 8e16.

Mann ME, Rutherford S, Wahl E, Ammann C. 2007. Robustness of proxy-based climate field reconstruction methods. *Journal of Geophysical Research* **112** : doi: 10.1029/2006JD008272.

Mann ME, Zhang Z, Hughes MK, Bradley RS, Miller SK, Rutherford S, Ni F. 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences* **105** : 13252-13257.

Marchant R, Behling H, Berrio JC, Cleef A, Duivenvoorden J, Hooghiemstra H, Kuhry P, Melief B, Van Geel B, Van der Hammen T, Van Reenen G, Wille M. 2001. Mid- to Late-Holocene pollen-based biome reconstructions for Colombia. *Quaternary Science Reviews* **20** : 1289-1308.

Marchant RA, Harrison SP, Hooghiemstra H, Markgraf V, Boxel JH, Ager T, Almeida L, Anderson R, Baied C, Behling H, Berrio JC, Burbridge R, Björck S, Byrne R, Bush MB, Cleef AM, Duivenvoorden JF, Flenley JR, de Oliveira PE, van Geel B, Graf KJ, Gosling WD, Harbele S, van der Hammen T, Hansen BCS, Horn SP, Islebe GA, Kuhry P, Ledru M-P, Mayle FE, Leyden BW, Lozano-Garcia MS, Lozano-Garcia S, Melief ABM, Moreno P, Moar NT, Prieto A, van Reenan GB, Salgado-Labouriau ML, Schäbitz F, Schreve-Brinkamn EJ, Wille M. 2009. Pollen-based biome reconstructions for Latin America at 0, 6000 and 18 000 radiocarbon years. *Climate of the Past* **5** : 725-767.

Marcott SA, Shakun JD, Clark PU, Mix AC. 2013. A reconstruction of regional and global temperature for the past 11,300 years. *Science* **339** : 1198–1201.

MARGO Project Members. 2009 Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum. *Nature Geoscience* **2** : 127–132.

Marlon J, Bartlein PJ, Carcaillet C, Gavin DG, Harrison SP, Higuera PE, Joos F, Power M, Prentice IC. 2008. Climate and human influences on global biomass burning over the past two millennia. *Nature Geoscience* **1** : 697-702.

Marlon JR, Bartlein PJ, Daniau A-L, Harrison SP, Maezumi SY, Power MJ, Tinner W, Vannière B. 2013. Global biomass burning: a synthesis and review of Holocene paleofire records and their controls. *Quaternary Science Reviews* **65** : 5-25.

Martin Calvo M, Prentice IC, Harrison SP. 2014. Climate versus carbon dioxide controls on biomass burning: a model analysis of the glacial-interglacial contrast. *Biogeosciences* **11** : 6017–6027.

Marzin C, Braconnot P. 2009. Variations of Indian and African monsoons induced by insolation changes at 6 and 9.5 kyr BP. *Climate Dynamics* **33** : 215–231.

Marzin C, Braconnot P, Kageyama M. 2013. Relative impacts of insolation changes, meltwater fluxes and ice sheets on African and Asian monsoons during the Holocene. *Climate Dynamics* **41** : 2267-2286.

Masson-Delmotte V, Schulz M, Abe-Ouchi A, Beer J, Ganopolski A, González Rouco JF, Jansen E, Lambeck K, Luterbacher J, Naish T, Osborn T, Otto-Bliesner B, Quinn T, Ramesh R, Rojas M, Shao X, Timmermann A. 2013. Information from Paleoclimate Archives. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Chang*e, Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA; 383-464.

Mauri A, Davis BAS, Collins PM, Kaplan JO. 2014. The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data–model comparison. *Climate of the Past* **10** : 1925–1938.

McIntyre A, Kipp NG, Be AWH, Crowley T, Kellogg T, Gardner JV, Prell W, Ruddiman WF. 1976. Glacial North Atlantic 18,000 years ago: A CLIMAP reconstruction. *Memoires of the Geological Society of America* **145** : 43-76.

Meehl GA, Moss R, Taylor KE, Eyring V, Stouffer RJ, Bony S, Stevens B. 2014. Climate model intercomparison: preparing for the next phase. *EOS* **95** : 77.

Michlmayr G, Lehning M, Koboltschnig G, Holzmann H, Zappa M, Mott R, Schöner W. 2008. Application of the Alpine 3D model for glacier mass balance and glacier runoff studies at Goldbergkees, Austria. *Hydrological Processes* **22 :** 3941–3949.

Mickelson DM, Colgan PM. 2003. The southern Laurentide Ice Sheet in the United States: What have we learned in the last 40 years?. In *Glacial Landsystems*, Evans DA, Rea BR (eds). Edwin Arnold, London; 111-142.

Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlén W, Lauritzen SE. 2005. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* **433** : 613–617.

Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature* **430** : 768–772.

Muscheler R, Joos F, Beer J, Müller SA, Vonmoos M, Snowball I. 2007. Solar activity during the last 1000 yr inferred from radionuclide records. *Quaternary Science Reviews* **26** : 82-97.

Neukom R, Gergis J, Karoly DJ, Wanner H, Curran M, Elbert J, Gonzalez-Rouco F, Linsley BK, Moy AD, Mundo I, Raible CC, Steig EJ, van Ommen T, Vance T, Villalba R, Zinke J, Frank D. 2014. Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change* **4** : 362-367.

Ohgaito R, Abe-Ouchi A. 2009. The effect of sea-surface temperature bias in the PMIP2 AOGCMs on mid-Holocene Asian monsoon enhancement. *Climate Dynamics* **33** : 975-983.

Ohgaito R, Sueyoshi T, Abe-Ouchi A, Hajimi T, Watanabe S, Kim HJ, Yamamoto A, Kawamiya M. 2013. 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 Pa*st **9** : 1519–1542.

Otto-Bliesner BL, Hewitt CD, Marchitto TM, Brady E, Abe-Ouchi A, Crucifix M, Murakami S, Weber SL. 2007. Last Glacial Maximum ocean thermohaline circulation: PMIP2 model intercomparisons and data constraints. *Geophysical Research Letters* **34** : L12706.

Otto-Bleisner BL, Schneider R, Brady EC, Kucera M, Abe-Ouchi A, Braconnot P, Crucifix M, Hewitt C, Kageyama M, Marti O, Paul A, Rosell-Mele A, Weber SL, Weinelt M, Yu Y. 2009. A comparison of PMIP2 model simulations and the MARGO proxy reconstruction for tropical sea surface temperatures at last glacial maximum. *Climate Dynamics* **32** : 799–815.

Otto-Bliesner BL., Russell JM, Clark PU, Liu Z, Overpeck JT, Konecky B, deMenocal P, Nicholson SE, He F, Lu Z. 2014. Coherent changes of southeastern equatorial and northern African rainfall during the last deglaciation. *Science* **346** : 1223-1227.

Otto-Bliesner BL, Brady EC, Fasullo J, Jahn A, Landrum L, Stevenson S, Rosenbloom N, Mai A, Strand G. 2015. Climate variability and change since 850 C.E.: An ensemble approach with the Community Earth System Model (CESM). *Bulletin of the American Meteorological Society*, doi: 10.1175/BAMS-D-14-00233.1

PAGES 2k Consortium. 2013. Continental-scale temperature variability during the past two millennia. *Nature Geoscience* **6** : 339-346.

PAGES 2k–PMIP3 Group. 2015. Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium. *Climate of the Past* **11** : 1673–1699.

Palmer TN. 2014. More reliable forecasts with less precise computations: a fast-track route to cloud-resolved weather and climate simulators? *Philosophical Transactions of the Royal Society A* : 20130391.

Patricola CM, Cook KH. 2007. Dynamics of the West African Monsoon under mid-Holocene precessional forcing: Regional climate model simulations. *Journal of Climate* **20** : 694–716.

Pauling A, Luterbacher J, Casty C, Wanner H. 2006. 500 years of gridded high-resolution precipitation reconstructions over Europe and the connection to large-scale circulation. *Climate Dynamics* **26** : 387-405.

Perez-Sanz A, Li G, Gonzalez P, Harrison SP. 2014. Evaluation of seasonal climates of northern Africa and the Mediterranean in the CMIP5 simulations. *Climate of the Past* **10** : 551-568.

Pestle WJ, Crowley BE, Weirauch MT. 2014. Quantifying inter-laboratory variability in stable isotope analysis of ancient skeletal remains. *PLOSOne* DOI: 10.1371/journal.pone.0102844.

Peyron O, Jolly D, Bonnefille R, Vincens A, Guiot J. 2000. Climate of East Africa 6000 14C Yr B.P. as inferred from pollen data. *Quaternary Research* **54** : 90–101.

Philippsen B. 2013. The freshwater reservoir effect in radiocarbon dating. *Heritage Science* **2013** : 1:24 .

Piao S, Sitch S, Ciais P, Friedlingstein P, Peylin P, Wang X, Ahlstrom A, Anav A, Canadell JG, Cong N. 2013. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. *Global Change Biology* **19** : 2117-2132.

Pickett EJ, Harrison SP, Hope G, Harle K, Dodson JR, Kershaw AP, Prentice IC, Backhouse J, Colhoun EA, D'Costa D, Flenley J, Grindrod J, Haberle S, Hassell C, Kenyon C, Macphail M, Martin H, Martin AH, McKenzie M, Newsome JC, Penny D, Powell J, Raine JI, Southern W, Stevenson J, Sutra JP, Thomas I, van der Kaars S, Ward J. 2004. Pollen-based reconstructions of biome distributions for Australia, Southeast Asia and the Pacific

(SEAPAC region) at 0, 6000 and 18,000 ^{14}C yr B.P. *Journal of Biogeography* **31** : 1381-1444.

Pongratz J, Reick C, Raddatz T, Claussen M. 2008. A reconstruction of global agricultural areas and land cover for the last millennium. *Global Biogeochemical Cycles* **22** : GB3018.

Power MJ, Ortiz N, Marlon J, Bartlein PJ, Harrison SP, Mayle F, Ballouche A, Bradshaw R, Carcaillet C, Cordova C, Mooney S, Moreno P, Prentice IC, Thonicke K, Tinner W, Whitlock C, Zhang Y, Zhao Y, Anderson RS, Beer R, Behling H, Briles C, Brown K, Brunelle A, Bush M, Clark J, Colombaroli D, Chu CQ, Daniels M, Dodson J, Edwards ME, Fisinger W, Gavin DG, Gobet E, Hallett DJ, Higuera P, Horn S, Inoue J, Kaltenrieder P, Kennedy L, Kong ZC, Long C, Lynch J, Lynch B, McGlone M, Meeks S, Meyer G, Minckley T, Mohr J, Noti R, Pierce J, Richard P, Shuman BJ, Takahara H, Toney J, Turney C, Umbanhower C, Vandergoes M, Vanniere B, Vescovi E, Walsh M, Wang X, Williams N, Wilmshurst J, Zhang JH. 2008. Changes in fire activity since the LGM: an assessment based on a global synthesis and analysis of charcoal data. *Climate Dynamics* **30** : 887-907.

Power MJ, Marlon JR, Bartlein PJ, Harrison SP. 2010. Fire history and the Global Charcoal Database: a new tool for hypothesis testing and data exploration. *Palaeogeography, Palaeoclimatology, Palaeoecology* **291** : 52-59.

Prentice IC. 2015. Palaeoclimate reconstruction by model inversion taking into account CO_2 effects on plant water use efficiency, and its application to a new pollen data compilation for Australia. *INQUA Congress Abstracts* **P02-15**.

Prentice IC, Harrison SP. 2009**.** Ecosystem effects of CO_2 concentration: evidence from past climates. *Climates of the Past* **5** : 297-307.

Prentice IC, Webb III T. 1998. BIOME 6000: reconstructing global mid-Holocene vegetation patterns from palaeoecological records. *Journal of Biogeography* **25** : 997-1005.

Prentice IC, Guiot J, Huntley B, Jolly D, Cheddadi R. 1996. Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics* **12** : 185-194.

Prentice IC, Jolly D, BIOME 6000 Participants. 2000. Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography* **27** : 507-519.

Prentice IC, Harrison SP, Bartlein, PJ. 2011a. Global vegetation and terrestrial carbon cycle changes after the last ice age. *New Phytologist* **189** : 988–998.

Prentice IC, Kelley DI, Foster PN, Friedlingstein P, Harrison SP, Bartlein PJ. 2011b. Modeling fire and the terrestrial carbon balance. *Global Biogeochemical Cycles* **25** : GB3005.

Rauser F, Gleckler P, Marotzke J. 2014. Rethinking the default construction of multi-model climate ensembles. *Bulletin of the American Meteorological Society*: BAMS-D-13-00181.1.

Reichler T, Kim J. 2008. How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society* **89** : 303–311.

Reick C, Raddatz T, Brovkin V, Gayler V. 2013. Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems* **5** : 459-482.

Reimer PJ, Reimer RW. 1991. A marine reservoir correction database and on-line interface. *Radiocarbon* **43** : 461–463.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Burr GS, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kaiser KF, Kromer B, McCormac FG, Manning SW, Reimer RW, Richards DA, Southon JR, Talamo S, Turney CSM, van der Plicht J, Weyhenmeyer CE. 2009. IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51** : 1111-1150.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson, TP, Haflidason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer, RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der Plicht J. 2013. INTCAL13 and MARINE13 radiocarbon age calibration curves 0–50,000 years cal. *Radiocarbon* **55** : 1869–1887.

Renssen H, Mairesse A, Goosse H, Mathiot P, Heiri O, Roche DM, Nisancioglu KH, Valdes PJ. 2015. Multiple causes of the Younger Dryas cold period. *Nature Geosciences* **8** : 946–949.

Rojas M. 2013. Sensitivity of Southern Hemisphere circulation to LGM and 4 × CO2 climates. *Geophysical Research Letters* **40** : 965–970.

Rosenthal Y, Perron-Cashman S, Lear CH, Bard E, Barker S, Billups K, Bryan M, Delaney ML, deMenocal PB, Dwyer GS, Elderfield H, German CR, Greaves M, Lea DW, Marchitto TM, Pak DK, Paradis GL, Russell AD, Schneider RR, Scheiderich K, Stott L, Tachikawa K, Tappa E, Thunell R, Wara M, Weldeab S, Wilson PA. 2004. Interlaboratory comparison study of Mg/Ca and Sr/Ca measurements in planktonic foraminifera for paleoceanographic research. *Geochemistry, Geophysics, Geosystems* **5** : Q04D09.

Ruddiman WF, Mix AC. 1993. The North and Equatorial Atlantic at 9000 and 6000 yr B.P. In *Global Climates since the Last Glacial Maximum*, Wright HE, Kutzbach JE,Webb III T, Ruddiman WF, Street-Perrott FA, Bartlein PJ (eds). University of Minnesota Press, Minneapolis; 94–124.

Rutherford S, Mann ME, Osborn TJ, Bradley RS, Briffa KR, Hughes MK, Jones PD. 2005. Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to methodology, predictor network, target season and target domain. *Journal of Climate* **18** : 2308-2329.

Salzmann U, Dolan AM, Haywood AM, Chan WL, Voss J, Hill DJ, Abe-Ouchi A, Otto-Bliesner B, Bragg FJ, Chandler MA, Contoux C, Dowsett HJ, Jost A, Kamae Y, Lohmann G, Lunt DJ, Pickering SJ, Pound MJ, Ramstein G, Rosenbloom NA, Sohl I, Stepanek C, Ueda H, Zhang Z. 2013. Challenges in quantifying Pliocene terrestrial warming revealed by data–model discord. *Nature Climate Change* **3** : 969–974

Schmidt GA, Jungclaus JH, Ammann CM, Bard E, Braconnot P, Crowley TJ, Delaygue G, Joos F, Krivova NA, Muscheler R, Otto-Bliesner BL, Pongratz J, Shindell DT, Solanki SK, Steinhilber F, Vieira LEA. 2011. Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geoscientific Model Development* **4 :** 33–45.

Schmidt GA, Annan JD, Bartlein PJ, Cook BI, Guilyardi E, Hargreaves JC, Harrison SP, Kageyama M, LeGrande AN, Konecky B, Lovejoy S, Mann ME, Masson-Delmotte V, Risi C, Thompson D, Timmermann A, Tremblay L-B, Yiou Y. 2014. Using paleo-climate comparisons to constrain future projections in CMIP5. *Climate of the Past* **10** : 221-250.

Schmidt GA, LeGrande AN, Hoffmann G. 2007. Water isotope expressions of intrinsic and forced variability in a coupled ocean-atmosphere model. *Journal of Geophysical Research* **112** : D10103.

Schneider B, Leduc G, Park W. 2010. Disentangling seasonal signals in Holocene climate trends by satellite-model-proxy integration. *Paleoceanography* **25** : PA4217.

Shi F, Yang B, Mairesse A, von Gunten L, Li J, Bräuning A, Yang F, Xiao X. 2013. Northern Hemisphere temperature reconstruction during the last millennium using multiple annual proxies. *Climate Research* **56** : 231.

Simpson MJR, Milne GA, Huybrechts P, Long AJ. 2009. Calibrating a glaciological model of the Greenland ice sheet from the Last Glacial Maximum to present-day using field observations of relative sea level and ice extent. *Quaternary Science Reviews* **28** : 1631–1657.

Steinhilber F, Beer J, Frohlich C. 2009. Total solar irradiance during the Holocene. *Geophysical Research Letters* **36** : L19704.

Steinman BA, Abbott MB, Mann ME, Stansell ND, Finney BP. 2012. 1,500 year quantitative reconstruction of winter precipitation in the Pacific Northwest. *Proceedings of the National Academy of Sciences* **109** : 11619–11623.

Stevens B, Bony S. 2013. What are climate models missing? *Science* **340** : 1053-1054.

Street FA, Grove AT. 1976. Environmental and climatic implications of late Quaternary lake-level fluctuations in Africa. *Nature* **261** : 385-390.

Street FA, Grove AT. 1979. Global maps of lake-level fluctuations since 30,000 yr B.P. *Quaternary Research* **12** : 83-l 18.

Street-Perrott FA, Harrison SP. 1984. Temporal variations in lake levels since 30,000 yr BP - an index of the global hydrological cycle. *American Geophysical Union, Maurice Ewing Series* **5** : 118-129.

Street-Perrott FA, Marchand DS, Roberts N, Harrison SP. 1989. Global lake-level variations from 18,000 to 0 years ago: a palaeoclimatic analysis. *U.S. DOE/ER/60304-H1 TR046. U.S. Department of Energy, Technical Report*, 213pp.

Street-Perrott FA, Huang Y, Perrott RA, Eglinton G, Barker P, Khelifa LB, Harkness DD, Olago DO. 1997. Impact of lower atmospheric carbon dioxide on tropical mountain ecosystems. *Science* **278** : 1422-1426.

Stuiver M, Reimer PL, Braziunas TF. 1998. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* **40** : 1127-1154.

Sturm C, Zhang Q, Noone D. 2010. An introduction to stable water isotopes in climate models: benefits of forward proxy modelling for paleoclimatology. *Climate of the Past* **6** : 115-129.

Stute M, Schlosser P, Clark JF, Broecker WS. 1992. Paleotemperatures in the Southwestern United States derived from noble gas measurements in groundwater. *Science* **256** :1000–1003.

Takemura T, Egashira M, Matsuzawa K, Ichijo H, O'ishi R, Abe-Ouchi A. 2009. A simulation of the global distribution and radiative forcing of soil dust aerosols at the Last Glacial Maximum. *Atmospheric Chemistry and Physics* **9** : 3061-3073.

Tarasov L, Dyke AS, Neal RM, Peltier WR. 2012. A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth and Planetary Science Letters* **315-316** : 30–40.

Tarasov PE, Guiot J, Cheddadi R, Andreev AA, Bezusko LG, Blyakharchuk TA, Dorofeyuk NI, Ludmila V, Filimonova LV, Volkova VS, Zernitskaya VP. 1999. Climate in northern Eurasia 6000 years ago reconstructed from pollen data. *Earth and Planetary Science Letters* **171** : 635–645.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* **106** : 7183–7192.

Taylor KE, Stouffer RJ, Meehl GA. 2011. An overview of CMIP5 and the experiment design. *Bulletin of the Amercian Meteorological Society:* BAMS-D-11-00094.1.

Thompson DM, Ault TR, Evans MN, Cole JE, Emile-Geay J. 2011. Comparison of observed and simulated tropical climate trends using a forward model of coral $\delta^{18}O$. *Geophysical Research Letters* **38** : L14706.

Timm O, Timmermann A. 2007. Simulation of the last 21,000 years using accelerated transient boundary conditions. *Journal of Climate* **20** : 4377–4401.

Trenberth KE, Shea DJ. 2005. Relationships between precipitation and surface temperature. *Geophysical Research Letters* **32** : L14703.

Vieira LEA, Solanki SK, Krivova NA, Usoskin I. 2011. Evolution of the solar irradiance during the Holocene. *Astronomy and Astrophysics* **531** : A6.

Wang H, Prentice IC, Ni J. 2013. Data-based modelling and environmental sensitivity of vegetation in China. *Biogeosciences* **10** : 5817–5830.

Wang Y-M, Lean JL, Sheeley NR. 2005. Modeling the Sun's magnetic field and irradiance since 1713. *The Astrophysical Journal* **625** : 522–538.

Wasson RJ, Claussen M. 2002. Earth system models: a test using the mid-Holocene in the Southern Hemisphere. *Quaternary Science Reviews* **21** : 819-824.

Webb T III, Bartlein PJ, Harrison SP, Anderson KH. 1993. Vegetation, lake levels, and climate in eastern North America for the past 18,000 years. In *Global Climates Since the Last Glacial Maximum*, Wright HE, Kutzbach JE, Webb T, Ruddiman WF, Street-Perrott FA, Bartlein PJ (eds), University of Minnesota Press, Minneapolis; 415–467.

Werner JP, Tingley MP. 2015. Technical Note: Probabilistically constraining proxy age–depth models within a Bayesian hierarchical reconstruction model. *Climate of the Past* **11** : 533–545.

Werner M, Tegen I, Harrison SP, Kohfeld KE, Prentice IC, Balkanski Y, Rodhe H, Roelandt C. 2003. Seasonal and interannual variability of the mineral dust cycle under present and glacial climate conditions. *Journal of Geophysical Research - Atmospheres* **108** : 47744.

Williams J, Barry RG, Washington WW. 1974. Simulation of the atmospheric circulation using the NCAR global circulation model with ice age boundary conditions. *Journal of Applied Meteorology* **13** : 305-317.

Williams JW, Shuman BN, Webb III T, Bartlein PJ, Leduc PL. 2004. Late Quaternary vegetation dynamics in North America: scaling from taxa to biomes. *Ecological Monographs* **74** : 309–334.

Williams JW, Tarasov P, Brewer S, Notaro M. 2011. Late-Quaternary variations in tree cover at the northern forest-tundra ecotone. *Journal of Geophysical Research – Biogeosciences* **116** : G01017.

Wilson R, Cook E, D'Arrigo R, Riedwyl N, Evans MN, Tudhope A, Allan R. 2010. Reconstructing ENSO: The influence of method, proxy data, climate forcing and teleconnections. *Journal of Quaternary Science* **25** : 62–78.

Wilson R, Miles D, Loader NJ, Melvin T, Cunningham L, Cooper R, Briffa K. 2012. A millennial long March–July precipitation reconstruction for southern-central England. *Climate Dynamics* DOI 10.1007/s00382-012-1318-z

Wohlfahrt J, Harrison SP, Braconnot P. 2004. Synergistic feedbacks between ocean and vegetation on mid- and high-latitude climates during the mid-Holocene. *Climate Dynamics* **22** : 223-238.

Woodward FI. 1987. *Climate and Plant Distribution*. Cambridge University Press, Cambridge.

Wright HE,. Kutzbach JE, Webb III T, Ruddiman WF, Street-Perrott FA, Bartlein PJ (eds). 1993. *Global Climates since the Last Glacial Maximum* University of Minnesota Press, Minneapolis, 569 pp.

Wu H, Guiot J, Brewer S, Guo Z. 2007. Climatic changes in Eurasia and Africa at the Last Glacial Maximum and mid-Holocene: reconstruction from pollen data using inverse vegetation modeling. *Climate Dynamics* **29** : 211–229.

Xoplaki E, Luterbacher J, Paeth H, Dietrich D, Steiner N, Grosjean M, Wanner H. 2005. European spring and autumn temperature variability and change of extremes over the last half millennium. *Geophysical Research Letters* **32** : L15713.

Yiou P, Servonnat J, Yoshimori M, Swingedouw D, Khodri M, Abe-Ouchi A. 2012. Stability of weather regimes during the last millennium from climate simulations. *Geophysical Research Letters* **39** : L08703.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292** : 686-693.

Zhao Y, Harrison SP. 2012. Mid-Holocene monsoons: a multi-model analysis of the inter-hemispheric differences in the responses to orbital forcing and ocean feedbacks. *Climate Dynamics* **39** : 1457-1487.

Zhao Y, Braconnot P, Marti O, Harrison SP, Hewitt C, Kitoh A, Liu Z, Mikolajewicz U, Otto-Bliesner B, Weber SL. 2005. A multi-model analysis of the role of the ocean on the African and Indian monsoon during mid-Holocene. *Climate Dynamics* **25** : 777-800.

Zhao Y, Braconnot P, Harrison SP, Yiou P, Marti O. 2007. Simulated changes in the relationship between tropical ocean temperatures and the western African monsoon during the mid-Holocene. *Climate Dynamics* **28** : 533-551.

Zheng W, Braconnot P. 2013. Characterization of model spread in PMIP2 mid-Holocene simulations of the African monsoon. *Journal of Climate* **26** : 1192-1210.

Zheng W, Braconnot P, Guilyardi E, Merkel U, Yu Y. 2008. ENSO at 6ka and 21ka from ocean-atmopshere coupled model simulations. *Climate Dynamics* **30** : 745-762.

Figure and Table Captions

Figure 1: Changes in lake status (a) in the mid-Holocene (MH, 6 ka) and (b) at the Last Glacial Maximum (LGM, 21 ka) compared to present day. Data from the Global Lake Status Database (Kohfeld and Harrison, 2000; data available from the PMIP2 website: https://pmip2.lsce.ipsl.fr/).

Figure 2: Reconstructed changes in mean annual temperature (MAT) (a) in the mid-Holocene (MH, 6ka) and (b) at the Last Glacial Maximum (LGM, 21ka) compared to present day. The reconstructions of ocean temperature are from the MARGO database (MARGO Project Members 2009) and the reconstructions of land temperature are from Bartlein et al. (2011). The original site-based reconstructions are gridded to a 2° by 2° grid for the land (https://www.ncdc.noaa.gov/paleo/study/9897) and a 5° by 5° grid for the ocean (http://www.ncdc.noaa.gov/paleo/study/12034) (Harrison et al., 2013). The significance of the temperature changes is indicated by the dot sizes: large dots show where the confidence intervals of the reconstructions do not include 0. The lower panels show (c) MH and (d) LGM sites where quantitative reconstructions exist (dark magenta) and where it would be possible to make quantitative reconstructions, although these have not been made to date (green).

Figure 3: Reconstructed northern hemisphere global annual temperatures during the last 2000 years, redrawn from Masson-Delmotte et al. (2013). All series are anomalies from the 1881–1980 CE mean (horizontal dashed line) and have been smoothed with a filter that reduces variations on time scales less than about 50 years. Curves from instrumental records are plotted in blue, and the purple lines show a locally-weighted regression curve with a 25-yr window half-width fit to the original unsmoothed series, and the 95-percent bootstrap confidence intervals for that curve that show the impact of the individual series to the overall curve.

Figure 4: Simulated and observed evolution of the hydroclimate of northern Africa during the past 21,000 years. Simulated precipitation minus evaporation (P-E) is an area-average over all land cells with centre points between 9.28° and 24.12° N, for (a) winter and spring (November, December, January, February, March, April: NDJFMA), (b) for pre-monsoon (May, June: MJ), (c) monsoon (July, August: JA) and (d) late monsoon (September, October: SO) intervals from the TraCE-21k simulation (http://www.cgd.ucar.edu/ccr/TraCE/; Liu et al., 2009). Lake status (e) in the equivalent region (7.42° to 29.69° N) is derived from data in the Global Lake Status Database (Kohfeld and Harrison, 2000; data available from the PMIP2 website: https://pmip2.lsce.ipsl.fr/), with the original

1 kyr ^{14}C reporting intervals converted to calendar ages using the *intcal13.14c* calibration curve (Reimer et al. 2013).

Figure 5: Observed and predicted zonal changes in biomass burning over the past 21 kyr. Composite charcoal influx curves for the northern extratropics (30° N– 90° N), northern tropics (0–30° N), southern tropics (0– 30° S) and southern extratropics (30° S– 90° S) with confidence intervals based on bootstrap resampling by site. The black curves and gray envelopes show locally weighted regression fitted values and confidence intervals using a window (half) width of 500 yrs, while the blue curves are fitted values for a window (half) width of 2000 yrs. The data are taken from the Global Charcoal Database v2.0 (http://www.gpwg.org/gpwgdb.html). The purple lines show values of charcoal predicted using a generalized additive model developed using zonally averaged charcoal values and zonally averaged temperature and precipitation minus evaporation (P-E) over land from a transient simulation of the ECBilt-CLIO model (Timm and Timmermann, 2007).

Figure 6: Simulated and observed vegetation changes across North America during the mid-Holocene (MH, 6 ka). The simulations were made using the BIOME4 biogeography model (Kaplan et al., 2003) driven by long-term averages of monthly mean temperature, sunshine and precipitation derived from Palaeoclimate Modelling Intercomparison Project (PMIP2) simulations made with the (a) CSIRO-Mk3L-1.0 coupled ocean-atmosphere (OA) and (c) ocean-atmosphere-vegetation (OAV) models. The observed vegetation during the MH (b) is derived from the BIOME6000 dataset (Prentice et al., 2000; Bigelow et al., 2003; data available from the PMIP2 website: https://pmip2.lsce.ipsl.fr/). The OAV model does not shows appreciably greater agreement with the observed vegetation then the less complicated OA model.

Figure 7: Scatter plots showing changes in land-ocean contrast in past, present, and projected climates. The black dots are the simulated long-term mean differences (experiment minus pre-industrial Control) in the relative warming/cooling over global land and global ocean. The red crosses show simulated changes where the model output has been sampled only at the locations for which there are temperature reconstructions for the Last Glacial Maximum (LGM, 21 ka) or mid-Holocene (MH, 6 ka), or observations for the historical (post 1850 CE) interval. Area-weighted averages of the palaeoclimate data are shown by a bold blue cross, with reconstruction uncertainties (standard deviation) shown by the finer lines. The inset shows data points for the MH and historical intervals.

Figure 8: Comparison of median and interquartile ranges (IQR) of observed and simulated growing season temperatures (as measured by growing degree days above a threshold of 5°C: GDD5) in (a) the mid-Holocene and (b) the Last Glacial Maximum. The comparisons are made using only the model land grid cells where there are observations. The reconstructed GDD5 is from the Bartlein et al. (2011) data set. The models are colour-coded to show whether they are CMIP5 simulations or from the previous generation of simulations made by the Palaeoclimate Modelling Intercomparison Project (PMIP2), and whether they are ocean–atmosphere (OA), ocean–atmosphere-vegetation (OAV) or OA carbon-cycle (OAC) models. The simulated median for each model is shown by a vertical line, the box represents the IQR.

Table 1: Description of the palaeosimulations included in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) and of the boundary conditions specified for these experiments.

Table 2: List of models and institutions contributing palaeoclimate simulations to the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The model names are the codes used to identify each model in the CMIP5 archive.

Table 1: Description of the palaeosimulations included in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) and of the boundary conditions specified for these experiments.

Abbreviation (in this paper)	**Name of Experiment (in ESGF database)**	**Description**	**Boundary conditions**
PiControl	*piControl*	Equilibrium simulation of 1850 CE, used as control for MH and LGM simulations (also used as a baseline for historical simulations by groups that did not run the palaeosimulations)	**Orbital parameters**: eccentricity = 0.016724, obliquity = 23.446°, perihelion-180° = 102.04° **Trace gases**: CO_2 = 280 ppm, CH_4 = 650 ppb, N_2O = 270 ppb, CFC = 0, O_3 = modern-10 DU **Ice sheet**: modern **Land surface**: modern or computed with dynamical vegetation model **Carbon cycle:** Interactive, with atmospheric concentration prescribed and ocean and land carbon fluxes diagnosed as recommended in CMIP5 *Note*: modelling groups that did not run palaeosimulations could have used a slightly different configuration for the PiControl
LM		Transient simulation of the last millennium, 850-1850 CE	
MH	*midHolocene*	Equilibrium simulation of 6 ka	**Orbital parameters**: eccentricity = 0.018682, obliquity = 24.105°, perihelion-180° = 0.87° **Trace gases**: CO_2 = 280 ppm, CH_4 = 650 ppb, N_2O = 270 ppb, CFC = 0, O_3 = same as in CMIP5 PI **Ice sheet**: as in CMIP5 PiControl **Land surface**: Computed using a dynamical vegetation module or prescribed as in PiControl, with phenology computed for models with active carbon cycle or prescribed from data **Carbon cycle**: Interactive, with atmospheric concentration prescribed and ocean and land carbon fluxes diagnosed as recommended in CMIP5
LGM	*lgm*	Equilibrium simulation of the Last Glacial Maximum, 21 ka	**Orbital parameters**: eccentricity = 0.018994, obliquity = 22.949°, perihelion-180° = 114.42° **Trace gases**: CO_2 = 185 ppm, CH_4 = 350

			ppb, N_2O = 200 ppb, CFC =0, O_3 = as in CMIP5 PI **Ice sheet**: Prescribed consensus ice sheet as described on PMIP3 website, with consistent changes to land-sea mask and sea level **Land surface**: Computed using a dynamical vegetation module or prescribed as in PiControl, with phenology computed for models with active carbon cycle or prescribed from data **Carbon cycle**: Interactive, with atmospheric concentration prescribed and ocean and land carbon fluxes diagnosed as recommended in CMIP5

Table 2: List of models and institutions contributing palaeoclimate simulations to the fifth phase of the Coupled Model Intercomparison Project (CMIP5). The model names are the codes used to identify each model in the CMIP5 archive.

Model name	Institution	PI control	Last Millennium	Mid-Holocene	Last Glacial Maximum
BCC-CSM1	Beijing Climate Center, China Meteorological Administration, China	✔		✔	
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France	✔		✔	✔
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence, Australia	✔		✔	
EC-EARTH	EC-Earth consortium	✔		✔	
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University, China	✔		✔	
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University, China	✔		✔	
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, US	✔		✔	
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, US	✔		✔	
GISS-E2-R	NASA Goddard Institute for Space Studies, US	✔	✔	✔	✔
HadGEM2-CC	Hadley Center, UK Met. Office, UK	✔		✔	
HadGEM2-ES	Hadley Center, UK Met. Office, UK	✔		✔	
INM-CM4	Institute for Numerical Mathematics, Russia	✔		✔	
IPSL-CM5A-LR	Institut Pierre-Simon Laplace, France	✔		✔	✔
IPSL-CM5A-MR	Institut Pierre-Simon Laplace, France	✔		✔	
MIROC-ESM	Japan Agency for Marine-Earth	✔	✔	✔	✔

	Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan				
MIROC5	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	✔		✔	
MPI-ESM-P	Max Planck Institute for Meteorology, Hamburg, Germany	✔	✔	✔	✔
MRI-CGCM3	Meteorological Research Institute, Tsukuba, Japan	✔		✔	✔
NCAR-CCSM4	National Center for Atmospheric Research, US/Dept. of Energy/NSF	✔		✔	✔
NorESM1-M	Norwegian Climate Centre, Norway	✔		✔	

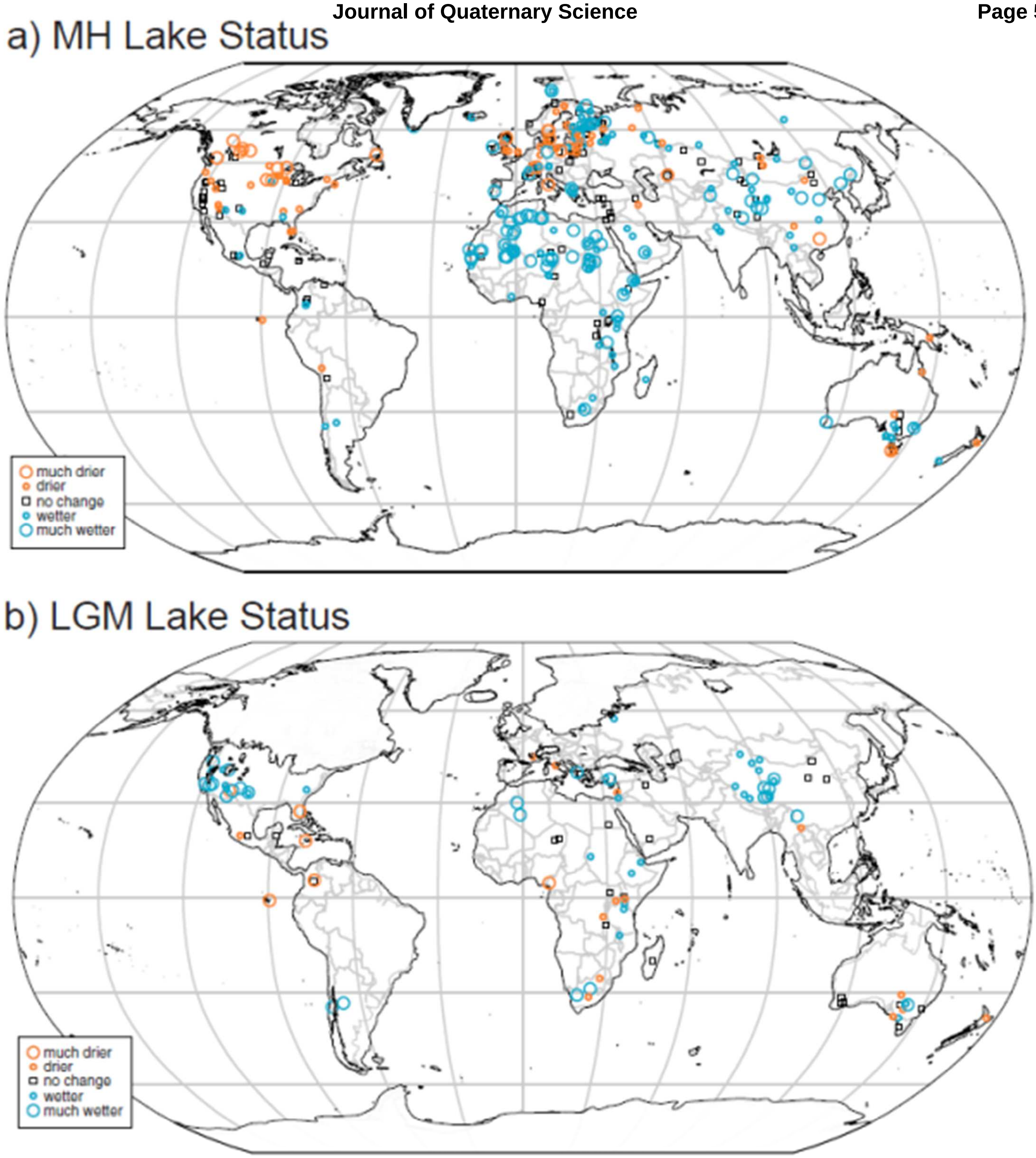


Figure 1: Changes in lake status (a) in the mid-Holocene (MH, ca 6000 yr B.P.) and (b) at the Last Glacial Maximum (LGM, ca 21000 yr B.P.) compared to present day. Data from the Global Lake Status Database (Kohfeld and Harrison, 2000).

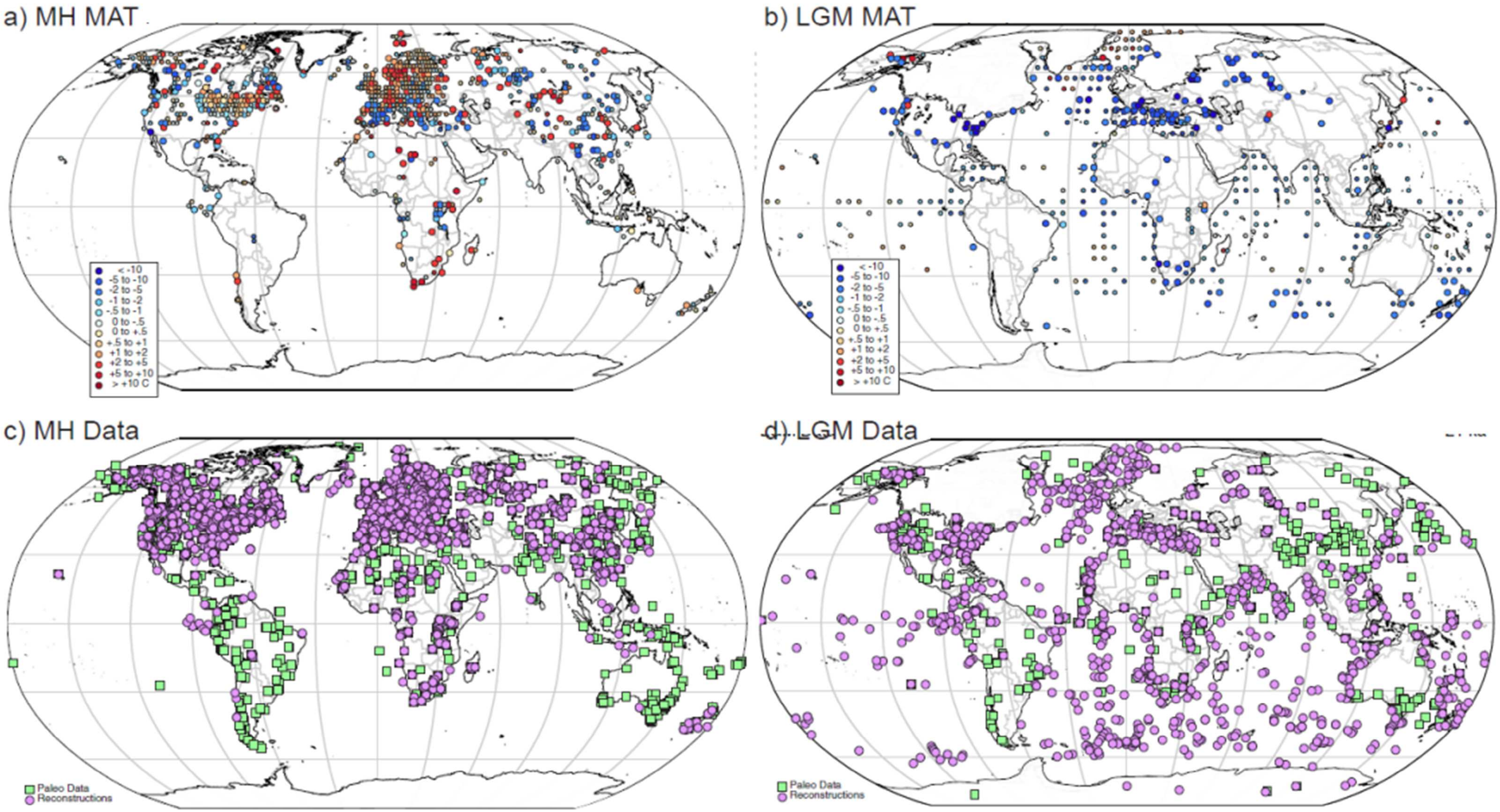


Figure 2: Reconstructed changes in mean annual temperature (MAT) (a) in the mid-Holocene (MH, ca 6000 yr B.P.) and (b) at the Last Glacial Maximum (LGM, ca 21000 yr B.P.) compared to present day. The reconstructions of ocean temperature are from the MARGO database (MARGO Project Members

2009) and the reconstructions of land temperature are from Bartlein et al. (2011). The original site-based reconstructions are gridded to a 2° by 2° grid for the land and a 5° by 5° grid for the ocean (Harrison et al., 2013). The significance of the temperature changes are indicated by the dot sizes: large dots show where the confidence intervals of the reconstructions do not include 0.0 The lower panels show (c) MH and (d) LGM sites where quantitative reconstructions exist (dark magenta) and were it would be possible to make quantitative reconstructions, although these have not been made to date (green).

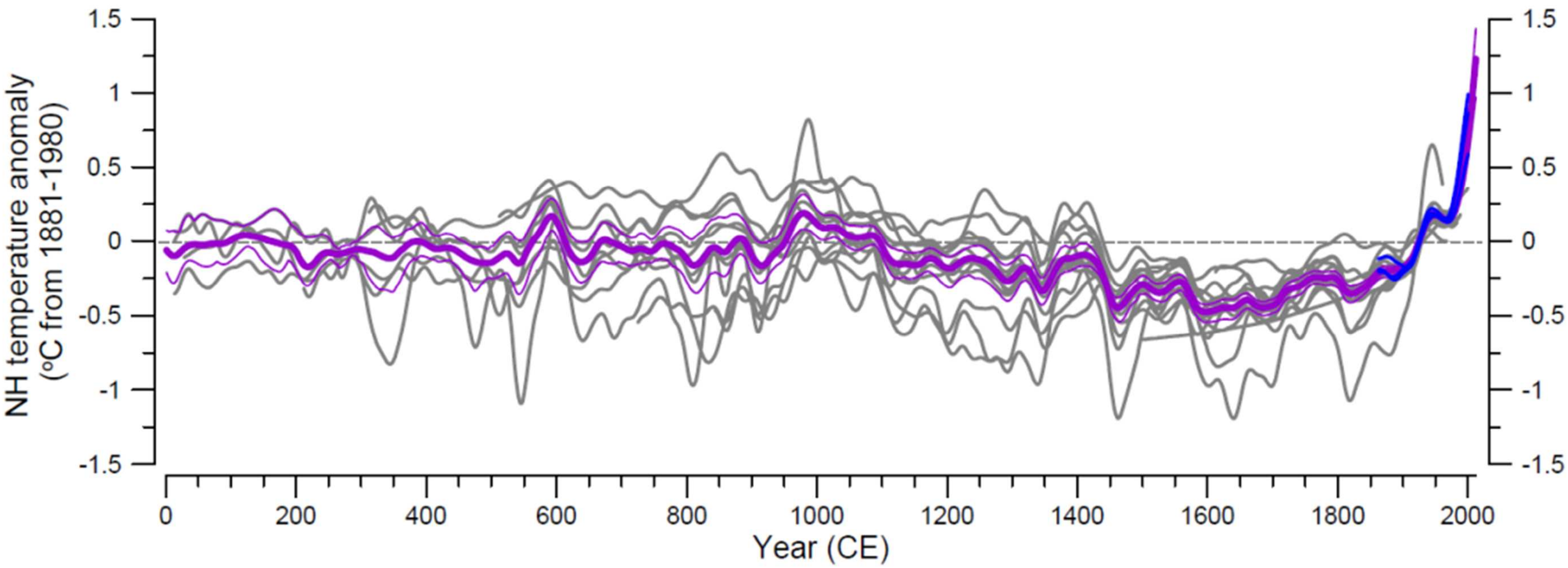


Figure 3: Reconstructed Northern Hemisphere global annual temperatures during the last 2000 years, redrawn from Masson-Delmotte et al. (2013). All series are anomalies from the 1881–1980 mean (horizontal dashed line) and have been smoothed with a filter that reduces variations on time scales less than about 50 years. Curves from instrumental records are plotted in blue, and the purple lines show a locally-weighted regression curve with a 25-yr window half-width fit to the original unsmoothed series, and the 95-percent bootstrap confidence intervals for that curve that show the impact of the individual series to the overall curve.

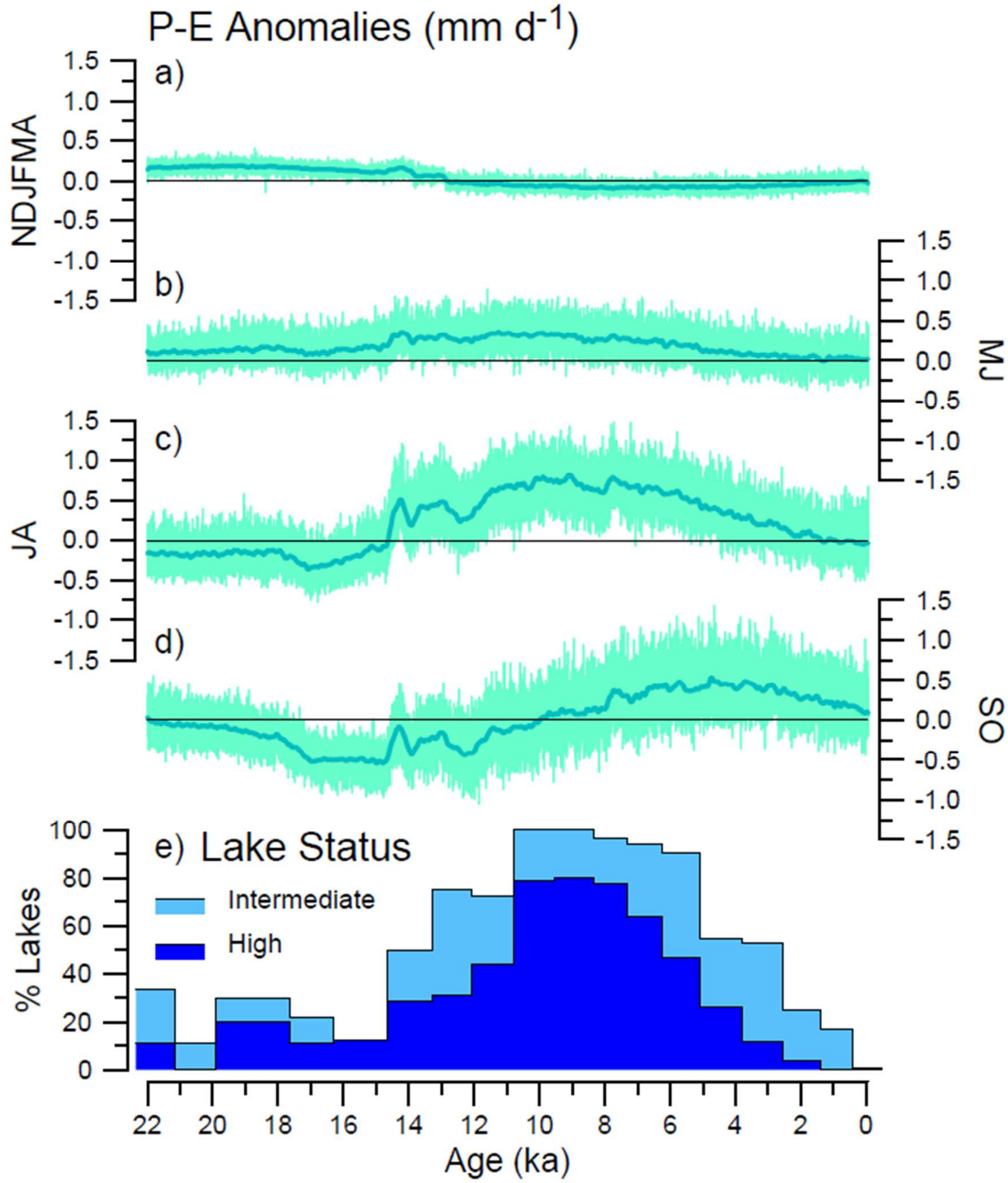


Figure 4: Simulated and observed evolution of the hydroclimate of northern Africa during the past 21,000 years. Simulated precipitation minus evaporation (P-E) is an area-average over all land cells with centre points between 9.28° and 24.12° N, for (a) winter and spring (November, December, January, February, March, April: NDJFMA), (b) for pre-monsoon (May, June: MJ), (c) monsoon (July, August: JA) and (d) late monsoon (September, October: SO) intervals from the TraCE-21k simulation (http://www.cgd.ucar.edu/ccr/TraCE/; Liu et al., 2009). Lake status (e) in the equivalent region (7.42° to 29.69° N) is derived from data in the Global Lake Status Database (Kohfeld and Harrison, 2000), with the original 1 kyr 14C reporting intervals converted to calendar ages using the intcal13.14c calibration curve (Reimer et al. 2013).

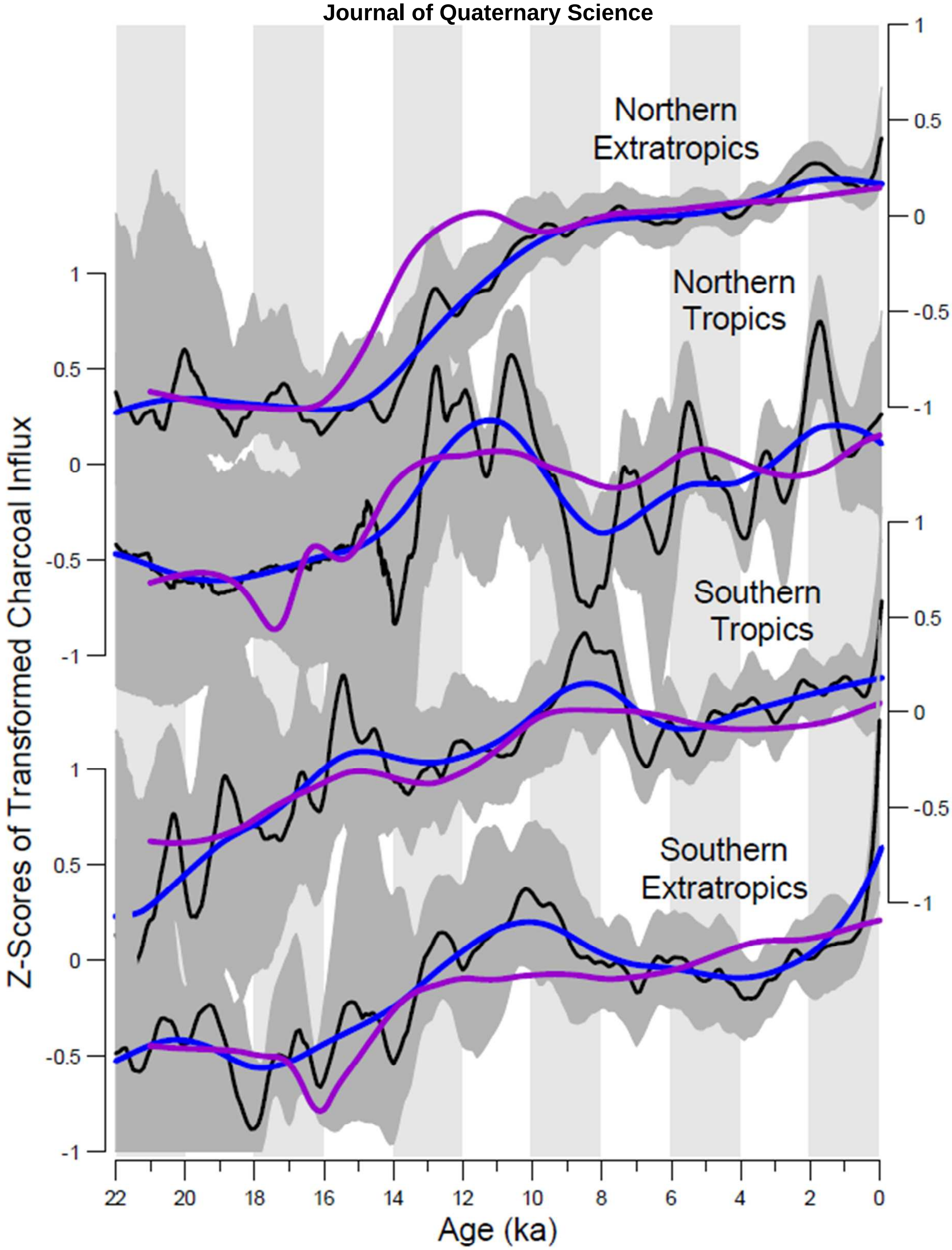
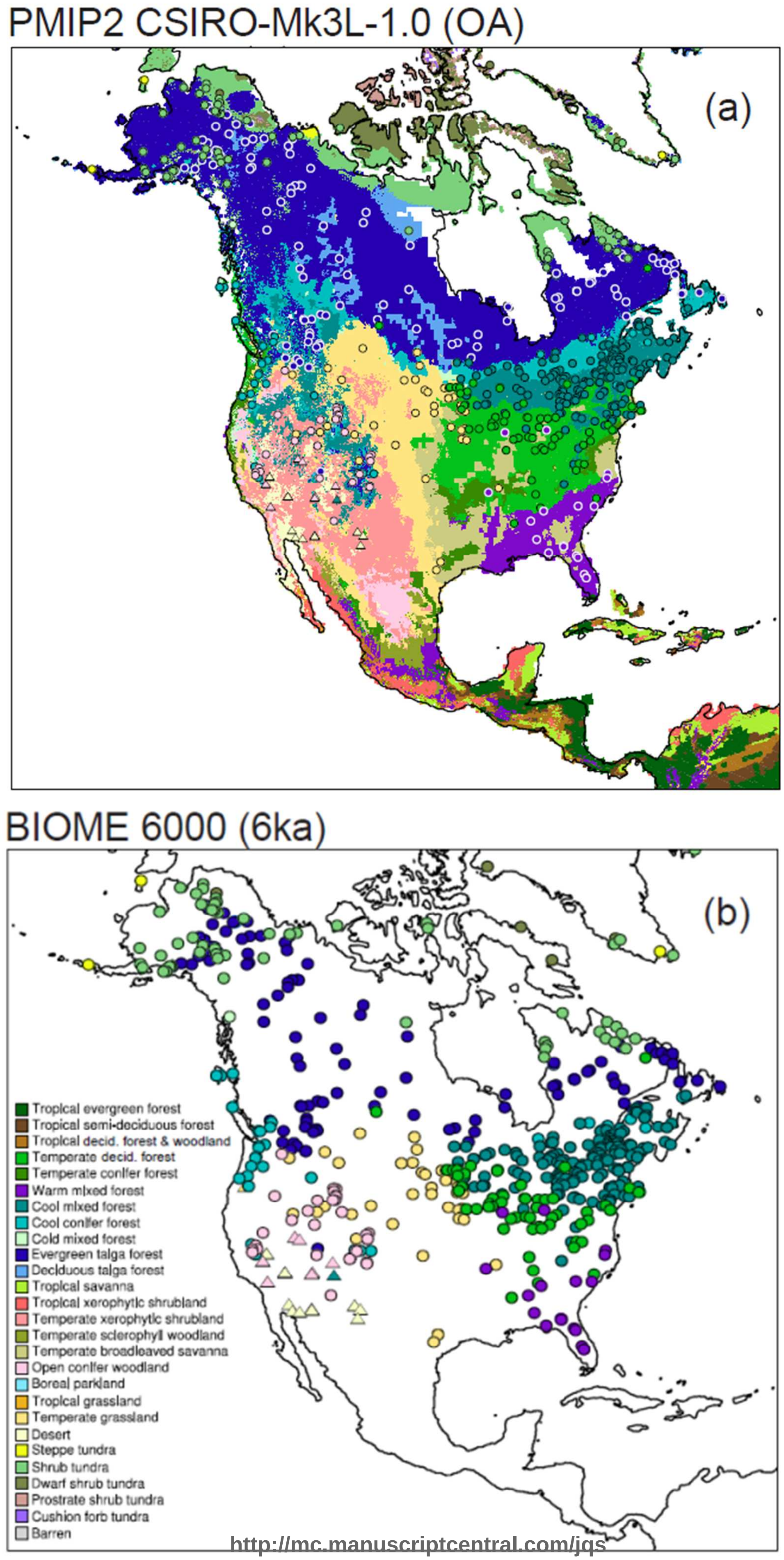


Figure 5: Observed and predicted zonal changes in biomass burning over the past 21 kyr. Composite charcoal influx curves for the northern extratropics (30° N– 90° N), northern tropics (0– 30° N), southern tropics (0– 30° S) and southern extratropics (30° S– 90° S) with confidence intervals based on bootstrap resampling by site. The black curves and gray envelopes show locally weighted regression fitted values and confidence intervals using a window (half) width of 500 yrs, while the blue curves are fitted values for a window (half) width of 2000 yrs. The purple lines show values of charcoal predicted using a generalized additive model developed using zonally averaged charcoal values and zonally averaged temperature and precipitation minus evaporation (P-E) over land from a transient simulation of the ECBILT-CLIO model (Timm and Timmermann, 2007).


PMIP2 CSIRO-Mk3L-1.0 (OA)
(a)
BIOME 6000 (6ka)
(b)
Tropical evergreen forest
Tropical semi-deciduous forest
Tropical decid. forest & woodland
Temperate decid. forest
Temperate conifer forest
Warm mixed forest
Cool mixed forest
Cool conifer forest
Cold mixed forest
Evergreen taiga forest
Deciduous taiga forest
Tropical savanna
Tropical xerophytic shrubland
Temperate xerophytic shrubland
Temperate sclerophyll woodland
Temperate broadleaved savanna
Open conifer woodland
Boreal parkland
Tropical grassland
Temperate grassland
Desert
Steppe tundra
Shrub tundra
Dwarf shrub tundra
Prostrate shrub tundra
Cushion forb tundra
Barren

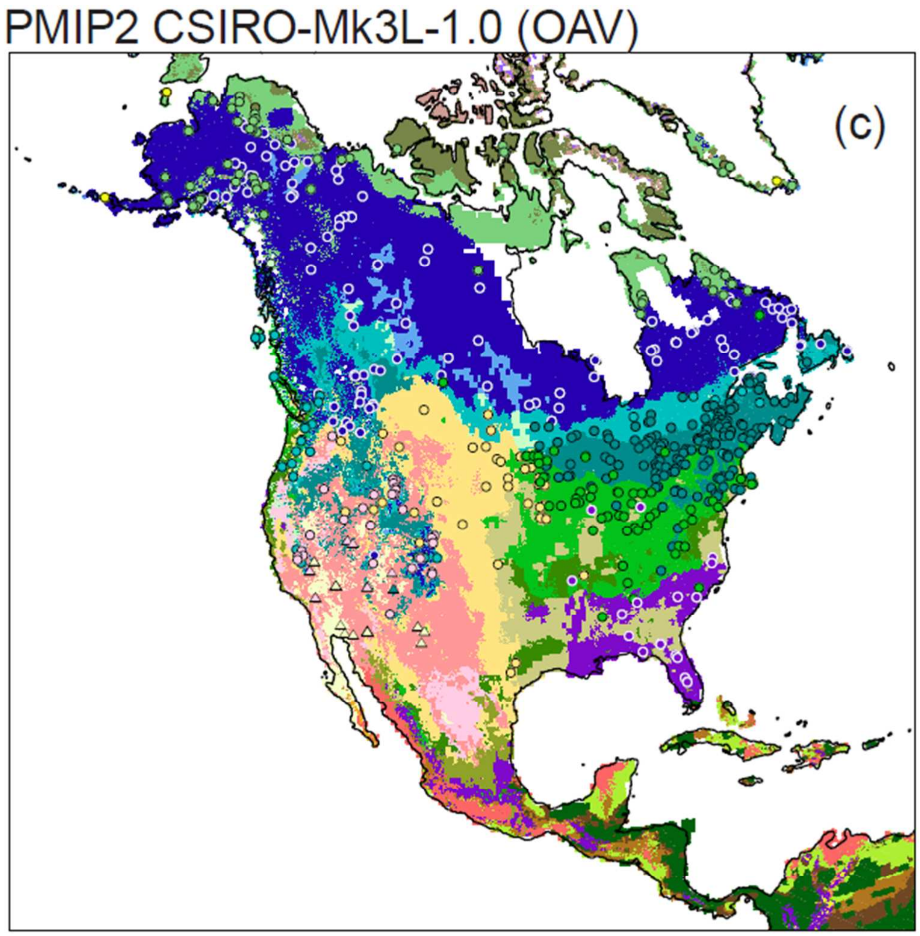


Figure 6: Simulated and observed vegetation changes across North America during the mid-Holocene (MH, ca 6000 yr B.P.). The simulations were made using the BIOME4 biogeography model (Kaplan et al., 2003) driven by long-term averages of monthly mean temperature, sunshine and precipitation derived from Palaeoclimate Modelling Intercomparison Project (PMIP2) simulations made with the (a) CSIRO-Mk3L-1.0 coupled ocean-atmosphere (OA) and (c) ocean-atmosphere-vegetation (OAV) models. The observed vegetation during the MH (b) is derived from the BIOME6000 dataset (Prentice et al., 2000; Bigelow et al., 2003). The OAV model does not shows appreciably greater agreement with the observed vegetation then the less complicated OA model.

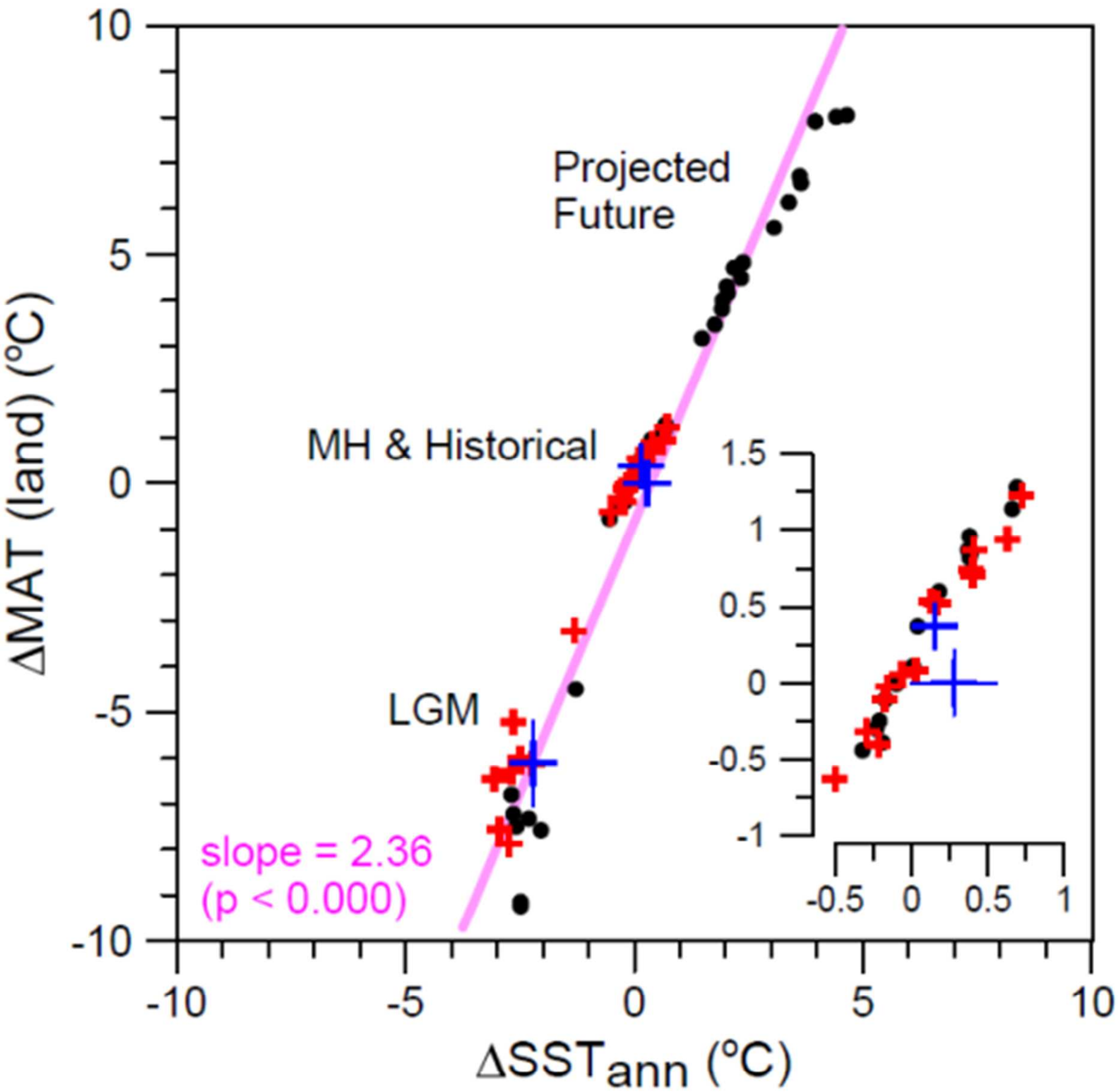
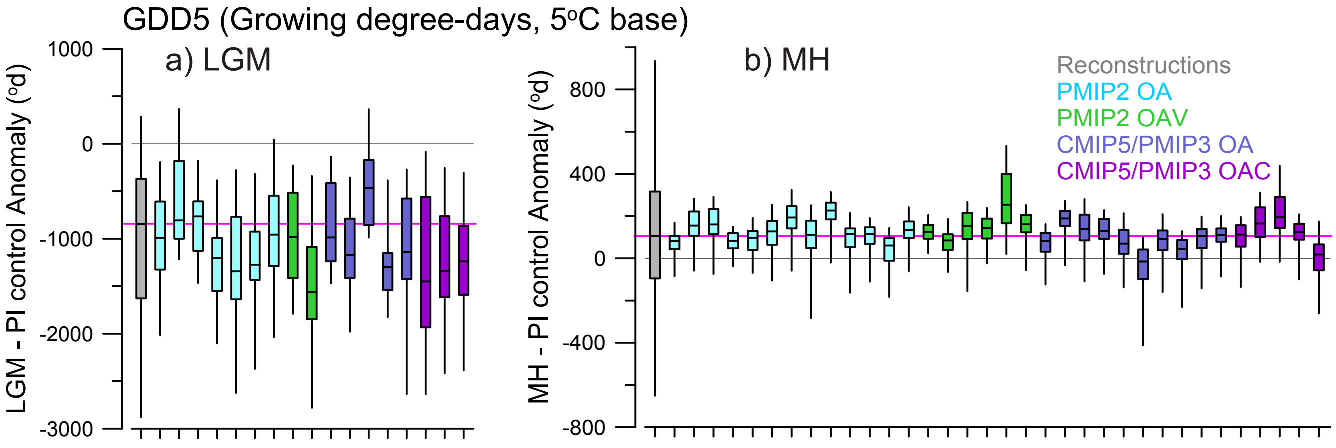


Figure 7: Scatter plots showing changes in land-ocean contrast in past, present, and projected climates. The black dots are the simulated long-term mean differences (experiment minus pre-industrial Control) in the relative warming/ cooling over global land and global ocean. The red crosses show simulated changes where the model output has been sampled only at the locations for which there are temperature reconstructions for the Last Glacial Maximum (LGM, ca 21000 yr B.P.) or mid-Holocene (MH), or observations for the historical (post 1850 CE) interval. Area-weighted averages of the palaeoclimate data are shown by a bold blue cross, with reconstruction uncertainties (standard deviation) shown by the finer lines. The inset shows data points for the MH and historical intervals.


GDD5 (Growing degree-days, 5°C base)
a) LGM
LGM - PI control Anomaly (°d)
1000
0
-1000
-2000
-3000
b) MH
MH - PI control Anomaly (°d)
800
400
0
-400
-800
Reconstructions
PMIP2 OA
PMIP2 OAV
CMIP5/PMIP3 OA
CMIP5/PMIP3 OAC