JOURNAL OF QUATERNARY SCIENCE (2010) **25**(1) 79–87
This article is a U.S. Government work and is in the public domain in the U.S.A. Published in 2009 by John Wiley & Sons, Ltd.
Published online 30 September 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jqs.1314

Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change

GAVIN A. SCHMIDT*
NASA Goddard Institute for Space Studies and Center for Climate Systems Research, Columbia University, New York

Schmidt, G. A. 2010. Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change. *J. Quaternary Sci.*, **Vol. 25** pp. 79–87. ISSN 0267-8179.

Received 13 September 2008; Revised 20 May 2009; Accepted 22 May 2009

ABSTRACT: I discuss the role of model/data comparisons for past climate changes and use of such comparisons for enhancing credibility in future projections. I outline a framework in which data synthesis combined with suitable modelling targets should be able to reduce uncertainty in both. By focusing on areas that the latest Intergovernmental Panel on Climate Change assessment report (IPCC AR4) highlighted as being particularly uncertain in future projections, or where current models produce a very wide range of responses, the relevance of palaeoclimate data could be greatly enhanced. Specific targets include: the long-term behaviour of El Niño events and the potential response to volcanic and solar forcing; the variability of subtropical rainfall and the extent of the Hadley Circulation and their response to orbital and high-latitude forcing; ice sheet responses on sub-millennial timescales; multidecadal changes in the North Atlantic ocean circulation and, certainly, overall climate sensitivity. In each case, I highlight data synthesis steps and modelling approaches necessary for reducing the uncertainty. In particular, I stress the need for coordinated model simulation archives that are conformal to those used in simulations of the 20th century and beyond and the consistency of models used for past and future climate simulations. Published in 2009 by John Wiley & Sons, Ltd.

KEYWORDS: palaeoclimate; modelling; PMIP; IPCC; climate sensitivity.

Introduction

There are no true palaeoclimate analogues for the global changes projected for the 21st century in the Intergovernmental Panel for Climate Change (IPCC) Fourth Assessment Report (AR4) (Solomon *et al.*, 2007). However, many of the uncertainties highlighted in that report do involve aspects of the climate that have certainly changed in the past. Those changes, recorded in multiple archives, were sometimes forced externally or were manifestations of internal variability. Given this climate history, the role of modelling is to assess whether these changes are consistent with current theories of the climate system (and if not, why not), to test how realistic the sensitivity of the models is to various climate forcings and, hopefully, to give more credibility to the projections of these same models in response to future increases in anthropogenic forcing.

It is sometimes assumed that the goal of modelling is to produce a 'true' simulation of what happened in the past, but this is too simplistic. Even the best models, using as much data as can be brought to bear from satellites and other observations, cannot do this for the last few decades. Being able to do so deeper in the past is therefore very unlikely. Instead, the goal of palaeo-modelling is to sharpen the interpretations of past data and to help produce the most consistent explanations for what is seen in the data. This does not (thankfully) require a flawless re-creation of the past, but rather the robust simulation of how different aspects of climate fit together in the palaeoclimate contexts. In this context, 'robust' should be taken to mean that results are reproducible in multiple models and are not sensitive to *ad hoc* or arbitrary choices made in building any specific model. In the examples given below, I will try to elucidate how this works in practice.

Climate models span a spectrum of complexity from simple box models to full-blown high-resolution coupled ocean–atmosphere general circulation models (GCMs) with associated chemistry, aerosols and carbon cycle modules. While much of the discussion here applies across that spectrum, I will focus specifically on the kinds of models that are used for future projections – that is atmospheric GCMs (AGCMs), coupled atmosphere–ocean GCMs (AOGCMs) and Earth system models (ESMs) which additionally include more bio-geo-physicochemical modules.

These climate models are based on fundamental physics (conservation of energy, mass, radiative transfer, equations of

---

* Correspondence to: G. A. Schmidt, NASA Goddard Institute for Space Studies and Center for Climate Systems Research, Columbia University, 2880 Broadway, New York, NY 10025.
E-mail: gschmidt@giss.nasa.gov

motion etc.) combined with physically based empirical parameterisations that attempt to capture the phenomenology of unresolved processes (such as cloud formation or evaporation). They differ from statistical models in that we anticipate that they should still be useful for climates different from their 'calibration' period (usually the average climate of the last few decades for which we have substantial instrumental and remotely sensed information). The extent to which they are able to simulate out-of-sample climate changes in the past is directly relevant to how much confidence to place in projections of climate change in the future.

There are principally two different kinds of observations that are useful for constraining models: first, there are observations of small-scale, high-frequency relationships between climate variables (say humidity, updrafts and cloudiness) that are used for informing and evaluating specific parameterisations. Secondly, there are observations of large-scale phenomena that are emergent properties in this class of models; that is, they do not occur because of a single parameterisation, but through the interactions of many different aspects of the model. This includes changes in global temperature or in major modes of variability. Most observation–model comparisons involving palaeoclimate are with the second kind of observation and are thus an evaluation of how the models' emergent behaviour compares to reality.

These comparisons have been done for a long time. Early palaeoclimate modelling exercises (e.g. Kutzbach, 1981) identified key periods such as the mid Holocene where movement of the tropical rainbands, presumably as a function of orbital forcing, was a suitable target for testing the models. Coordinated palaeo-model/data comparison projects (such as COHMAP: the Cooperative Holocene Mapping Project (Wright and Bartlein, 1993) or the Paleoclimate Model Intercomparison Project (PMIP)) created communities and structures that could take these comparisons further. Other efforts have tested the overall climate sensitivity to forcings (Crucifix, 2006) or the temperature response to a change in the North Atlantic overturning circulation (Stouffer *et al.*, 2006).

Results from such comparisons have often shown where there is a mismatch (and where indeed the models do well), but on their own cannot be used to say exactly where the reason for any particular problem lies. These kinds of studies do not, in general, lead to changes to specific parameterisations, though they can indicate a possible need for models to include more processes (such as the dynamic vegetation effects on climate).

There is one kind of exception to this, exemplified by studies such as Schneider von Deimling *et al.* (2006) or the Palaeo-QUMP project (Edwards *et al.*, 2007). In those examples, multiple versions of a model are used to simulate the present and Last Glacial Maximum (LGM), and the versions that give the best estimate of the difference are more highly weighted. Given that each version differed in the values of specific parameters, an optimum value of the parameters could be chosen to maximise the fit with the palaeo-data. Thus this could be a demonstration of how past climate comparisons can impact parameterisations of processes. This approach will likely become more prevalent in the future, but will probably remain focused on specific periods (such as the LGM) where significant modelling work has already been done.

Any new proposal for palaeo-model/data comparisons needs to be aware of these limitations and such studies need to be designed to maximise their utility, not for 'improving model parameterisations' but for enhancing credibility of model projections and improving quantitative understanding of the causes of past climate events.

A framework for model/palaeo-data comparisons

With 4.5 billion years of Earth history, tens of thousands of records of palaeoclimate, hundreds of different proxies, an uncountable number of questions to pose but a very limited set of palaeo-modellers with limited resources, how and why do certain periods and regions get selected as targets? The answer lies in the kind of question that models can be most usefully used for.

Models are mostly used to test hypotheses and one of the more common forms is whether a particular cause can produce an observed (or inferred) effect. More complicated questions can also be asked, such as what controls the relationship between two observed variables, or whether a set of observations is consistent with the intrinsic variability of the climate or not. In each case there is a set of underlying assumptions (the components and physics in the model itself and any specific forcing) and a target (an observed change, or derived relationship).

For suspected causes that are external to the model (a change in solar irradiance, or in the orbital parameters), the test is straightforward. A boundary condition is changed and the model response noted. The definition of 'external', though, depends on the scope of the model. A change in sea surface temperatures is external for an atmosphere-only model, but not for a coupled ocean–atmosphere model. A change in methane concentrations is external to a model without atmospheric chemistry, but internal to one that includes it. Testing hypotheses related to internal variables is also possible, but usually requires a little more effort as shown, for example, in examining the stability of the isotope thermometer in Greenland (Werner *et al.*, 2000).

It is unsurprising that most palaeoclimate modelling has focused on periods and issues where there are clear hypotheses to test (usually drawn from the palaeoclimate literature), and (equally importantly) where there are clear targets for model/palaeo-data comparisons. There are additional constraints of relevance (more people are interested in processes/causes that are applicable to the present day) and practicality (a simulation of 100 years is a lot easier than one of 100 000 years). Favourable targets can be further subdivided into timeslices (equilibrium simulations of the long-term conditions over a particular period of time), transient simulations and recurrent events with each requiring a different kind of simulation, input and data for comparison.

One fundamental aspect of the hypothesis and targets is how quantifiable they are. Vague invocations of freshwater forcings as a cause of some change without a magnitude or potential source are not particularly useful. Neither are targets that are basically undefined such as 'change' in the monsoon or a 'difference' in the wind. Much better are specific changes in physical variables: isotopic ratios in ice cores, multi-proxy estimates of summer sea surface temperature (SST) change or reconstructions of the Palmer Drought Severity Index (PDSI).

For timeslices such as the LGM or the Pliocene that are stable enough for equilibrium simulations to be appropriate, modellers require as many of the changed boundary conditions as possible – the land–ocean mask, topography/bathymetry and orbital configuration – in order to maximise the relevance of the experiment. Depending on the scope of the model, greenhouse gas levels, vegetation/land cover, ice sheet extent, aerosol distributions, SSTs and sea ice distribution may also be required. The targets for these runs will be observations from the period in question (averaged over a suitably long period so that a quasi-equilibrium assumption holds, but not so long that the boundary conditions change significantly). For the model/data comparison to be a useful constraint (or evaluation) for the model or the hypothesis, it is important that the full range of forcings are used and that their uncertainties are built into the comparison. If this is is not done, comparisons to the data (with their own inherent uncertainties) will be unable to distinguish between model problems and boundary forcing issues (this is further discussed below).

For relatively short-term climate variability (decades to hundreds of years), transient simulations are sometimes more appropriate. A good example is the last millennium, where there are some constraints on important short-term forcings (principally solar and volcanic, but also land use, orbital and other changes prior to the modern industrial period), and well-dated reconstructions of spatial temperature changes (Crowley, 2000; Jones and Mann, 2004; Ammann *et al.*, 2007). The 8.2 ka event also has a relatively well-constrained potential forcing (the final drainage of Lake Agassiz; Clarke *et al.*, 2004) and is seen in enough proxy records to provide a useful test of model responses (Renssen *et al.*, 2001; LeGrande *et al.*, 2006; LeGrande and Schmidt, 2008).

In both of the previous cases, the models are trying to match specific occurrences in the record that are assumed to have relatively simple causes. This is always compromised by the possibility that something else may have been happening then that was not included in the experiments. One way around that is to look at the impact of recurrent events, such as solar cycles, orbital cycles or volcanic eruptions, that lend themselves to a superposed-epoch analysis (SEA). An SEA takes multiple examples of the same phenomena and averages them together, to give a picture of what a typical ‘event’ would look like. Even though any one individual event may be happening in a different context, with other forcings and internal variability playing varying roles in the response, by averaging together many such instances the generic response can sometimes be discerned. This has been the approach used for volcanic and solar forcings in the last millennium (Shindell *et al.*, 2001, 2004) or El Niño–Southern Oscillations (ENSO) in the present day, but could well be extended to studies of Quaternary variability such as Dansgaard–Oeschger or Heinrich events or the response to periodic changes in orbital forcing.

Much can also be learned from putting all these kinds of simulations together and assessing how different model components and diagnostics co-vary. For instance, what is the relationship between tropical precipitation and insolation forcing (Prell and Kutzbach, 1992) or the latitude of westerly winds or poleward extent of the Hadley circulation as global temperatures change? Do any relationships exist at all? These insights can often be used much more generally to interpret palaeo-records even in the absence of specific model simulations for the exact period in question.

However, in every case, there is a clear need for data synthesis to proceed in tandem with model experiments. This is required for creating suitable boundary conditions, defining the forcings and, crucially, providing coherent targets for the model/data comparisons. Synthesis is required because comparisons need to be with records on common age scales, across different proxies, and are often spatially dispersed. Putting these kinds of databases together coherently has traditionally been extremely labour intensive. Examples range from BIOME6000 for vegetation changes in the mid Holocene (Prentice and Webb, 1998), CLIMAP and MARGO for glacial SSTs (CLIMAP, 1981; Kucera *et al.*, 2005) to composite records of climatically important volcanic eruptions (Gao *et al.*, 2008).

One aspect of modelling that is making data–model comparisons easier is the improvement in incorporating forward modelling of proxies within the GCMs themselves. This refers to the advances in including tracers such as water isotopes, and modules for atmospheric chemistry and aerosols, biogeochemical cycles in the ocean etc., such that the model output can be directly compared to palaeo-records of the same quantities (oxygen isotopes in ice cores, ocean sediment and speleothems; trace gases and aerosols in ice cores; dust in various archives etc.). Palaeoclimate modelling is in fact benefitting greatly from the work that has gone into improving model simulations for the 20th century. The additions of chemistry, aerosols (including dust), carbon cycles and dynamic vegetation modules have been mostly driven (and funded) through concerns arising from the increasing anthropogenic impact on the Earth system (from CO_2 emissions, air pollution, deforestation etc.) – though it is certainly the case that palaeoclimate applications have also driven progress in these directions (e.g. Joussaume, 1990). In many cases, the diagnostics from these new tracers and modules are exactly what has been recorded in palaeo-archives (e.g. LeGrande *et al.*, 2006). These methods allow much more of our understanding of those tracers to be built into the models than was ever brought to bear in interpreting them as climate proxies. Subtleties and non-stationarity in climate/proxy connections become much easier to assess.

From MIPs to GRIPs

Early attempts at palaeoclimate model/data comparisons (e.g. PMIP) succeeded in fostering communities of scientists who were focused on making the necessary data syntheses and designing appropriate model experiments. However, the modelling that was possible at the time (mid 1990s) was not ideal for answering the questions being posed. For instance, the first phase of the PMIP included a test of the purely atmospheric response to orbital forcing at 6 ka, with fixed (modern) SSTs and atmospheric composition. This was a useful exercise for comparing model outputs with each other, but when comparing the models to the collated observations for the 6 ka period little could be said other than the models and data differed. This problem arose because the hypothesis that was being tested (implicitly, can the mid Holocene climate be explained solely by changes in orbital forcing but with no ocean, vegetation or greenhouse gas feedbacks?) was not close enough to the question that we would ideally like to ask (can we understand the changes in climate during the mid Holocene?). The difference in these questions – in the assumptions about the ocean or vegetation principally – precluded any strong conclusions about the adequacy of the models, or the consistency of the boundary conditions. Nonetheless, this kind of experiment was useful at least in indicating that the unchanged boundary conditions may well have been important. Subsequent experiments have borne out this contention, at least in part (Braconnot *et al.*, 1999). Similar AGCM experiments for the LGM using CLIMAP SSTs were sufficient to demonstrate that the ocean boundary conditions (used as input) could not be reconciled with the target terrestrial proxy data (CLIMAP, 1981; Rind and Peteet, 1985).

A more recent example of a pure MIP is the palaeo-hosing experiment (Stouffer *et al.*, 2006), where coupled models were all given identical (and completely artificial) freshwater hosing boundary conditions in order to provide insight into the generic response of models, but without a direct connection to any particular real event.

These above examples demonstrate that when the boundary conditions and/or the model configurations are not sufficiently conformal to the reality of the period in question it is natural for the focus to be on model–model intercomparisons (MIPs), i.e. how different models react to similar changes in forcing. While carefully controlled MIPs are indeed useful, particularly in exploring differences due to model physics, they are often misconstrued; that is, a mismatch between models and reality is assigned to 'model problems', while it may well have been related to the basic experimental design. An example of this problem is provided by the results reported in Otto-Bliesner *et al.* (2009a), where comparisons between the PMIP2 LGM models are made with the MARGO database of SST changes. Since none of the models used appropriate dust or vegetation changes, they note that the experimental design might itself lead to an underestimate of the LGM cooling by perhaps 1°C – clearly a non-negligible amount.

A GCM–reality intercomparison project (GRIP) can be thought of as the next phase of an MIP but with a subtle change in emphasis. The goal is not to compare one model with another but to find robust explanations of features seen in the real-world data. Specifically, if there is uncertainty in a particular component of a palaeoclimate experiment – the height of the ice sheet at the LGM, for instance – a fuller range of plausible ice sheet topography should be used. By contrast, if, as in earlier experiments, a single solution is imposed on all models, then systematic impacts of that choice cannot be explored. Similarly, if it is known that greenhouse gases or dust levels changed, forcing them to remain constant might simplify the experiment but at the cost of relevance to the comparison.

For any palaeo-GRIP, different modelling groups will have different levels of functionality and accordingly they will each make different sets of assumptions in make their 'best effort' to create an appropriate simulation. These choices should of course be clearly documented and quantified. This was (mostly) the case for the 20th-century runs in CMIP3 (the Coupled Model Intercomparison Project archive set up in support of IPCC AR4), where there were large differences in the kind and nature of the aerosol forcing, whether ozone depletion was included or whether and how volcanic stratospheric forcings were imposed. The disadvantage of this approach is that model-to-model comparisons are not as clean, but the advantages are that the model simulations now span a wider range of the true uncertainty in forcings and more robust conclusions can be drawn about the necessity of various processes. For instance, by segregating the models with and without a representation of the Antarctic ozone hole, Miller *et al.* (2006) showed that this forcing was essential for correctly simulating the increase in westerly winds in the Southern Hemisphere in recent decades.

Computational resources and modelling expertise have advanced considerably since the first PMIP experiments, and many reasons previously advanced for not doing more appropriate experiments have fallen by the wayside (coupled models are more available, various estimates of dust and vegetation forcings exist, etc.). The consequence of a palaeo-GRIP approach will be that the model solutions for the particular comparison will be more dispersed but they will hopefully span more of the range of plausible solutions. This will likely lead to more robust conclusions, as well as to improved targets for future research. Criticisms of this approach have focused on the difficulty of clean model–model comparisons, but it is worth pointing out that even in seemingly highly controlled experiments (such as the 1% increasing CO_2 simulations) the actual forcing imposed on each model will vary by more than 10% due to differences in radiation transfer code and the base climate, a factor that is rarely considered in comparisons. However, as with the difference between the 20th-century all-forcing runs and the 1% increasing CO_2 simulations in CMIP3, the two approaches can and do exist in tandem.

The next phase in palaeoclimate modelling (whether it is described as a GRIP or not) must be made more useful in constraining future projections, and for that we need to turn away from idealised comparisons to true best efforts – including all valid and known sources of uncertainty.

Palaeo-modelling in IPCC AR4

For the first time in an IPCC report, AR4 contained a chapter specifically devoted to palaeoclimate (Jansen *et al.*, 2007). Other papers in this special issue will discuss the scope and use of palaeo-data within it, but here I draw the reader's attention specifically to its use of palaeo-GCMs. Specific figures and conclusions focused almost exclusively on the PMIP target periods (the LGM and mid Holocene) and the last millennium because these were the only periods for which multi-model comparisons were available. Individual papers on other topics (forward modelling for the 8.2 ka event (LeGrande *et al.*, 2006), the Eemian (Otto-Bliesner *et al.*, 2006), hosing experiments (Stouffer *et al.*, 2006) etc.) were referred to but not systematically assessed. Notably, no systematic comparisons of GCM results in palaeoclimate contexts were referred to in the discussions of future projections for ENSO, regional hydrology or ice sheet development or collapse.

Even in the cases where multi-model ensembles were examined, the palaeo-models used were not generally the ones used in the climate projections discussed in the rest of the report (with a couple of notable exceptions) (Braconnot *et al.*, 2007). This lack of 'traceability' from palaeo-modelling to future projections is a fundamental constraint on how useful any palaeo-modelling work will be. The reasons for this state of affairs are rather obvious: models used for palaeo work are often of an older vintage, or have been designed specifically for palaeo studies, and computational restrictions often mean that palaeo simulations are of a lower priority than more standard climate simulations. Ideally, the same models should be used in both contexts and preparations for AR5 and beyond will stress this. A second-best alternative is to ensure that any palaeo-model in the ensemble is also used to simulate one or two of the AR5 representative marker scenarios so that its responses can be compared to the ensemble of future projections. For instance, if the models that perform best in palaeo-scenarios have a preferred response in future scenarios, there might be reason to more highly weight other models with a similar projection, but which were not tested in the palaeo-simulation. It is worth pointing out that weighting strategies will be matters of debate for many years until enough experience has been accrued for how to optimally meld disparate sources of model credibility.

Increasing the relevance of palaeo-modelling in future assessments

While there are many interesting questions in palaeoclimate that are amenable to model studies, the ones that will be most utilised in future assessments are likely those that have a direct relation to the uncertainties identified in AR4. These are areas where either model responses were not robust, where known important processes were not included, where modern model/

data comparisons show systematic problems or where there are no relevant modern data to compare to. Examples that have received considerable attention are the sensitivity and intrinsic variability of regional rainfall, ice sheet dynamics, carbon cycle feedbacks, polar amplification, overall climate sensitivity, abrupt change, vegetation, ocean circulation, ENSO sensitivity and possibly the prevalence of extreme events (in the broadest sense). All of these are high-priority targets where palaeomodelling may be able to contribute to reducing uncertainties. For each of these topics, Table 1 highlights appropriate palaeomodelling exercises and datasets that could be compared to add insight (see also the PMIP Phase II White Paper available at http://pmip2.lsce.ipsl.fr/share/overview/PMIP2_White_Paper.pdf and the summary of the 2008 PMIP workshop; Otto-Bliesner *et al.*, 2009b).

The table is by no means complete, but there are perhaps some puzzling lacunae. In particular, I do not list the Younger Dryas under the Abrupt climate change topic. The reason for this is a consequence of the framework outlined above. Specifically, without a good quantifiable hypothesis to test, model/data comparisons are always ambiguous and will not be useful in constraining future changes. In that regard, the 8.2 ka event is a much better modelling target for quantifying freshwater impacts on the Meridional Overturning Circulation (MOC) (Schmidt and LeGrande, 2005). Similarly, the deglaciation from the LGM is a fascinating process, but the deglacial history of the ice sheets and runoff are not well enough known to usefully force a coupled model, and ice sheet models themselves are not far enough advanced to be used in a multi-thousand-year coupled experiment. Thus, for the time being at least, those experiments will be interesting, but not constraining. I have also neglected the onset of the last glaciation. This has received some modelling attention already (Khodri *et al.*, 2001), but for obvious reasons is not high on the agenda of the IPCC.

The different model simulation classes (equilibrium vs. transient) imply that different output will be required for any organised archive. Quasi-equilibrium simulations require roughly 100 a worth of monthly data (to allow for an assessment of potential changes in interannual/decadal variability), with single-year or decade snapshots of high resolution (daily or 3-hourly) data. Getting to the point where these experiments are possible nonetheless requires multi-centennial (to thousand-year) spin-ups. Transients need simulations that extend over the length of the event – 100–200 a for the 8.2 ka event, 1000+ a for the last millennium. They also require at least a few ensemble members in order to assess the importance of the internal variability and periodic snapshots of high temporal resolution data.

The data sources listed have had very varied levels of synthesis applied. Some products are relatively mature (lake levels and pollen maps at the mid Holocene (Jolly *et al.*, 1998; Prentice *et al.*, 2000), North American PDSI reconstructions (Cook *et al.*, 2004), ice sheet reconstructions (Licciardi *et al.*, 1998; Peltier, 2004), LGM sea surface conditions (Kucera *et al.*, 2005), while other efforts have barely started. Thus the data list should be considered more aspirational than operational.

Given that there are substantial ambiguities in some of the input boundary conditions, different modelling groups will choose to make slightly different simulations (or sets of simulations) for this period (similarly to how the 20th-century AR4 experiments were conducted). As long as the differences are documented this is no impediment to using these results to assess, for instance, the relationship of tropical warming to atmospheric forcings.

Priorities for IPCC-related palaeomodelling and data synthesis

Given the targets and questions highlighted above, what are the model simulations that would go furthest in helping address them? Different time periods and forcings obviously exercise different parts of the models and explore different regions of the climate system phase space. Nonetheless, there are some periods that are more practical than others and priorities need to be decided accordingly. Note that in the specifications for the CMIP5 archive being constructed in support of the AR5, the mid Holocene, LGM and last millennium were highlighted as being desired experiments, alongside the 20th-century and future simulations.

Early to mid Holocene

The early to mid Holocene has been the workhorse of palaeomodelling since the first experiments from Kutzbach (1981) and it still holds potential for improving understanding, in particular, of hydrological responses of the climate system.

Table 1 Potential palaeoclimate targets for reducing uncertainties in future projections

Target	Periods/experiments	Forward models	Datasets
Subtropical rainfall	Mid Holocene	O isotopes, veg.	Speleothems, pollen, lakes
	High medieval	Tree rings	Gridded PDSI
Ice sheet dynamics	LGM, deglaciation	ice sheets	Isostasy, moraines
	Eemian, Pliocene	(as above)	(as above)
Carbon cycle	Glacial-to-interglacial	Carbon cycle	Ice cores
	Last millennium, Holocene	(as above)	(as above)
Polar amplification	LGM	Water isotopes	Ice cores
	Eocene, Pliocene	Veg., isotopes	Ocean sediment, palaeosols
Climate sensitivity	LGM	n/a	Multi-proxy
	Pliocene	n/a	(as above)
Abrupt change	8.2 ka, ‘Green Sahara’	O isotopes, veg.	Multi-proxy, lake-levels
Vegetation	Mid Holocene, LGM	Dynamic veg.	Pollen, palaeosols
	Pliocene, Eocene	(as above)	(as above)
N Atlantic MOC	LGM, 8.2 ka	C & O isotopes, Pr/Th	Ocean/ice cores
	Generic D/O events	(as above)	(as above)
ENSO sensitivity	Medieval, mid Holocene	Isotopes	Corals
	Generic solar or volcanic	(as above)	(as above)
Extremes(?)	Last millennium	?	Multi-proxy

Published in 2009 by John Wiley & Sons, Ltd.

The relevant boundary condition changes are dominated by the change in precession and obliquity, while greenhouse gas changes are small. Land cover changes in the high northern latitudes and in the 'Green Sahara' are useful as targets for dynamic vegetation models or as input to simpler configurations. The relevant questions for IPCC reports revolve around changes in hydrology in the northern subtropics and mid-continental regions – from the Sahel to China, explanations for the apparent reduction of ENSO variability and, potentially, the skill of models in replicating Arctic ocean conditions that may have had little to no summer ice (Funder and Kjaer, 2007). The wealth of relevant questions, calibrated targets, existing databases (lake levels, dust changes, temperature changes, cave isotope changes) and ease of implementation must make 6 ka the primary target for any future palaeo-GRIP.

Enhancements to the existing datasets could include more quantitative syntheses of ocean-based proxies (Mg/Ca ratios, δ^{18}O) (notwithstanding the difficulty in dating sediments of this age), a more comprehensive calibration of Arctic proxies, greater availability of global land cover datasets that can be used as a forcing in some configurations, and a wider use of forward models for water isotopes and vegetation.

Related experiments for 10–8 ka have similar advantages, and the existence of a remnant Laurentide ice sheet until perhaps 7 ka (or even later) adds the potential to test ice sheet models as well (Carlson *et al.*, 2008). However, the synthesis of observational data at these periods is not very far advanced. Thus, while useful (perhaps in combination with providing initial conditions for the 8.2 ka simulations discussed below), they are not yet a high priority.

LGM

As the most recent period with a large quasi-equilibrium global mean temperature difference from today, the LGM remains key in assessing questions of overall climate sensitivity and truly global perturbations to the carbon cycle, dust and other aerosols, ice sheets, vegetation and ocean circulation. The wealth of changes are a modelling challenge, but for all the periods discussed this is the one where as wide a range as possible of models and boundary conditions must be used. Orbital forcing and changes in greenhouse gases (GHGs) are well known, while sea-level and land-mask changes can be reasonably approximated. However, ice sheet extent, topography and changes in freshwater routing are more uncertain than has been recognised in PMIP experiments to date. Multiple reconstructions exist for each of the ice sheets (e.g. Licciardi *et al.*, 1998; Peltier, 2004; Siegert and Dowdeswell, 2004) and, while there is broad consistency in ice sheet outlines, the ice topography in different reconstructions can have important impacts on freshwater runoff and ocean circulation (A. N. LeGrande, pers. comm.). Vegetation types have been estimated (Ray and Adams, 2001) (somewhat subjectively) or estimated using online dynamic vegetation models (Crucifix and Hewitt, 2005; Ramstein *et al.*, 2007). Several additional Earth system components (atmospheric aerosols, including dust, and chemistry) have only been loosely incorporated so far (e.g. Kohfeld and Harrison, 2001, for dust).

It is of particular importance that the radiative forcings and freshwater transports for this period are more accurately characterised, or at minimum that the difference that any of these components might make (within plausible limits) is explored. Protocols for these experiments will need to allow for more divergence in the model configurations in future, including allowing different ice sheet topographies, river routing changes and the optional addition of dust, chemistry, aerosols and vegetation changes.

The last millennium

The last 1000 a or so (say 850–1850 CE) can be considered an extension of the instrumental record of climate change and a period when forced and internal variability of the climate system can be examined in the absence of large anthropogenic forcing. The changes over that period are driven by solar, volcanic and some land use change. Orbital and greenhouse gas changes are small (but are easily incorporated). Increasing numbers of hemispheric and regional reconstructions on annual and decadal resolution are becoming available (Moberg *et al.*, 2005; Mann *et al.*, 2008), and the targets for model/data comparison are legion. These simulations will also provide an invaluable test-bed for pseudo-proxy assessments of reconstruction methods (such as the PR Challenge, http://www.pages.unibe.ch/science/prchallenge/index.html).

These are transient experiments, but an appropriate long-term equilibrium run is needed for the initialisation if climate drifts that can be as large as the expected climate signal are to be avoided (Osborn *et al.*, 2006). The period that has been highlighted in the next phase of PMIP and which is a secondary target in CMIP5, from 850 CE onwards, allows for a clear view of any medieval climate anomaly to develop.

The 8.2 ka event

The 8.2 ka event is the last abrupt climate change visible in the Greenland ice core records and is clearly recognised in multiple proxies across the North Atlantic region (and perhaps more widely) (Morrill and Jacobsen, 2005). It is a short event (perhaps 160 a according to layer counting in the ice cores (Thomas *et al.*, 2007), and there is a very strong quantifiable candidate for its primary cause: the almost contemporaneous draining of Lake Agassiz (Clarke *et al.*, 2004), which added perhaps 2.5–5 Sv of freshwater into Hudson Strait over a 6-month to 1-year period. For modellers it is a test of (a) the sensitivity of the North Atlantic MOC to forcing and (b) how changes in the MOC affect key subsystems (isotopes, methane, dust, regional rainfall) (LeGrande *et al.*, 2006). The base conditions are similar enough to today that using pre-industrial control climate will likely give a reasonable response, but there are likely to be dependencies on the existence of the remnant ice sheet, the state of Labrador deep-water production and perhaps the differences in insolation (LeGrande and Schmidt, 2008).

No other palaeo-event potentially involving the North Atlantic MOC is as well documented, or has as quantifiable a forcing, nor a base climate close to present, nor as tractable a duration. This implies that this event is an essential target to develop for future comparisons of MOC sensitivity.

The Pliocene

As the last period which was consistently warmer than the Holocene for hundreds of thousands of years, the Pliocene holds much interest for examining what the long-term consequences of enhanced anthropogenic warming might be. Identified forcings for this period (usually taken to be the mid Pliocene, roughly 3.2 Ma ago) are mildly elevated greenhouse gas levels, small changes in orography (in particular in the Rockies and African highlands), vegetation

Published in 2009 by John Wiley & Sons, Ltd.

changes (poleward extensions of the treeline) and substantially less glacial ice – consistent with a global sea-level rise of around 20 m (Dowsett and Cronin, 1990; Haywood *et al.*, 2009).

Two kinds of tests are relevant to the IPCC report: the first is to test whether these boundary conditions are sufficient to produce temperature changes comparable to those inferred (a global change of about 2–3°C above pre-industrial) (Dowsett *et al.*, 1996); and secondly, whether dynamic vegetation and ice sheet models can produce solutions consistent with the Pliocene Research, Interpretation and Synoptic Mapping (PRISM) reconstructions (Chandler *et al.*, 2008). These questions tie in directly to the issue of Earth system sensitivity – the long-term change in the climate due to a change in forcings including all of the feedbacks – not just the selected fast feedbacks normally considered part of the standard (Charney) climate sensitivity (Hansen *et al.*, 2008; Lunt *et al.*, 2008). As a warm period, it is likely to be more useful in assessing this than the LGM which, though better observed, is perhaps not as relevant to the future.

The Eemian/last interglacial period

One additional warm period that is of interest is the last interglacial (roughly 125 ka) or Eemian. Orbital changes are the major forcing and, as with the Pliocene, the issue is whether models can match the elevated temperatures – particularly in the Arctic region (perhaps 3–5°C above pre-industrial) (Otto-Bliesner *et al.*, 2006; Overpeck *et al.*, 2006). As with the Pliocene, the magnitude and extent of Greenland and West Antarctic ice sheets are a crucial parameter since they provided some 4–6 m of sea-level rise at that time.

Control simulations

The control simulation is obviously not a key period for palaeo-modelling, but some thought is required to ensure that modellers choose the most relevant experiments with which to compare the palaeo-model results. This is mainly an issue for the equilibrium experiments (Pliocene, LGM, 6 ka, Eemian). For maximum overlap with the more standard IPCC simulations the controls should be the same – and that mandates using a generic pre-industrial base case with no (or very small) anthropogenic GHGs, near-modern insolation and modern geography. The switch between the equilibrium control runs and the transient simulations for the 20th century should be as smooth as possible and so the control needs to have forcings as close to the beginning of the transient run as possible. Different groups have made slightly different decisions on when to start their transients and so differ on the details of that forcing (whether they use 1750 or 1850 greenhouse gas levels, 1950 vs. 2000 orbital configuration) but as long as this clearly noted, it is not likely to make any substantial difference. More subtle effects might be associated with the levels of solar irradiance or background volcanic forcing – average mid-19th-century values are probably most useful.

Facilitating synthesis

One of the major lessons to be drawn from the history of palaeoclimate modelling and the IPCC AR4 archives is that these endeavours require significant facilitation and widespread contributions to get off the ground. The overall goal must be to open up both analysis of model output and collation of observational data to allow for as many cross-model and cross-experiment comparisons as possible.

For model simulations, the CMIP3 archive is a good archetype. The amount of data stored was prodigious, but so were the analyses done and the number of external participants involved. Over 500 publications have used these data to date. With the specifications of the AR5 archive (CMIP5) having been decided (at time of writing), the next phase of PMIP will be using exactly the same prototype (for diagnostics, metadata etc.). Storage of model output will need to be distributed based on available resources, and many initiatives to make dealing with the copious amounts of data easier are under development. The conformability of both models and diagnostics across the palaeo-simulations and the future simulations will allow the palaeo-comparisons to directly feed into the assessment of future projections.

When it comes to observational and proxy data, the database requirements are more complicated. Current practice is to archive proxy data with a specific age model, the raw or processed measurements and perhaps an interpretation in terms of temperature or precipitation. Unfortunately, the archives are not complete, age models are not static (through adjustments of the radiocarbon calibration, or reassessments of absolutely dated chronologies, such as the layer counting in Greenland ice cores), new data points are rarely added to the archive, and conventions for both dating (calendar years vs. BP (before present with respect to 1950) or BP (2000)) and age model construction vary widely. This makes it next to impossible for researchers to easily pull together snapshots of climate change for any particular interval or event.

It is vitally important to realise that data archiving is not the same as data synthesis, and it is the latter step that is crucial in allowing for more interesting model/data comparisons. It goes almost without saying that making the synthesis of data easier is the *sine qua non* of making true palaeo-GRIPs successful. There are many good ideas being discussed, but the principles of what is required are clear: (i) datasets need to be citable in ways that include updates and corrections; (ii) age–model calculations (and the metadata associated with them) need to be archived with the raw data; and (iii) databases need to actively allow for time-slice calculations, using consistent variations to the age models – for instance, as improvements are made to the ^{14}C calibrations or to the age models of common tuning targets. Ideally, the main existing archives (National Climate Data Center (NCDC) Paleo and Pangaea) will be able to migrate towards more full-service databases, but upgrading existing datasets to meet new requirements is likely to be hard. For instance, Pangaea does not keep track of how what the calendar year convention is in the data they archive (Hannes Grobe, pers. comm.).

Funding agencies can also play a significant role in fostering synthesis. Since synthesis often requires multidisciplinary teams, and needs to address issues with data that already exist, these kinds of proposals often fare poorly when included in competitive calls for proposals when stacked against new data-gathering ideas, or in very specific disciplinary calls. Progress can be made by setting aside money specifically for synthesis activities that are required to be cross-disciplinary or to deal only with existing data. It also needs to be recognised that the skills in putting data together are often not the same as the skills of a geochemist, for instance, in generating it. New kinds of science (and scientists) are likely to be required (Schmidt and Moyer, 2008).

Data gathering and synthesis are, however, both necessary: the first for insight into climate change itself, but the second to ensure that those insights are made maximum use of and reach the maximum audience. In many ways, it is the success of the few syntheses products (CLIMAP, multi-century multi-proxy reconstructions, Mann *et al.* (1998) and others) that have brought palaeoclimate research the increased attention that allows funding for the more specialised work.

Conclusion

While many aspects of climate change in the future are relatively robust, large areas with dramatic and significant potential impacts still remain uncertain. Palaeoclimate information may not be able to reduce all of those uncertainties but it is the only test-bed for the models that project large changes in the future to be evaluated using true out-of-sample tests of comparable magnitude. As a community, we have not come close to using the information already gathered as efficiently as we could – a task that is surprisingly undervalued in comparison with the (still necessary) gathering of new records. Some of the suggestions made here for more active databases and better archiving have been made elsewhere, but perhaps the link between lowering the barriers of entry to palaeo-data synthesis and the increased relevance of palaeoclimate itself have not been made clear as often. A higher public profile, increased funding and an ability to shape the science in AR5 and beyond will rely upon a community effort to provide more IPCC-relevant science – and, to a large extent, that means model/data comparisons. This does not imply that the data-gathering community needs to mothball their mass spectrometers, but it does mean that researchers, funders and archivists need to give more thought to how data, once gathered, will be used. At many of the modelling centres a greater appreciation of the role of palaeoclimate is required and a mainstreaming of palaeo-relevant modules (such as for water isotopes) and sufficient computer time for palaeo-model runs should be made higher priorities.

Acknowledgements I would like to thank the editors for inviting me to submit this review and the many people who I have discussed this topic with over the last year or so. In particular, conversations at the Leverhulme Climate Symposium in Cambridge (2008), the PAGES/CLIVAR workshop in Trieste (2008) and at the International Conference on Paleoceanography in Shanghai (2007) were extremely helpful. Special thanks to Mike Mann, Peter Huybers, Eric Steig, Kim Cobb, Mark Sidall, Reto Knutti, Jacqueline Flückiger, Sandy Harrison, Bette Otto-Bliesner, Caspar Ammann and Eric Wolff who have contributed (willingly or not) to making this a much more rounded article. Comments from Sylvie Joussaume and Patrick Bartlein on an early draft were very helpful. Climate modelling at GISS is supported by the NASA Modeling and Analysis Program.

References

Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA. 2007. Solar influence on climate during the past millennium: results from transient simulations with the NCAR Climate System Model. *Proceedings of the National Academy of Sciences* **104**: 3713–3718.

Braconnot P, Joussaume S, Marti O, de Noblet N. 1999. Synergistic feedbacks from ocean and vegetation on the African monsoon response to mid-Holocene insolation. *Geophysical Research Letters* **26**: 2481–2484.

Braconnot P, Otto-Bliesner B, Harrison S, Joussaume S, Peterchmitt J-Y, Abe-Ouchi A, *et al.* 2007. 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.

Carlson A, LeGrande A, Oppo D, Came R, Schmidt G, Anslow F, Licciardi J, Obbink E. 2008. Rapid early Holocene deglaciation of the Laurentide ice sheet. *Nature Geoscience* **1**: 620–624.

Chandler M, Dowsett H, Haywood A. 2008. The PRISM Model/Data cooperative: Mid-Pliocene data–model comparisons. *PAGES News* **16**: 24–25.

Clarke G, Leverington DW, Teller JT, Dyke AS. 2004. Paleohydraulics of the last outburst flood from glacial Lake Agassiz and the 8200 BP cold event. *Quaternary Science Reviews* **23**: 389–407.

CLIMAP. 1981. Seasonal reconstructions of the Earth's surface at the glacial maximum. *Map Chart Series MC–36*, Geological Society of America, Boulder, CO.

Cook E, Woodhouse CA, Eakin CM, Meko DM, Stahle D. 2004. Long-term aridity changes in the western United States. *Science* **306**: 1015–1018.

Crowley TJ. 2000. Causes of climate change over the past 1000 years. *Science* **289**: 270–277.

Crucifix M. 2006. Does the Last Glacial Maximum constrain climate sensitivity? *Geophysical Research Letters* **33**: L18701.

Crucifix M, Hewitt C. 2005. Impact of vegetation changes on the dynamics of the atmosphere at the Last Glacial Maximum. *Climate Dynamics* **25**: 447–459.

Dowsett H, Cronin T. 1990. High eustatic sea level during the middle Pliocene: evidence from southeastern US Atlantic coastal plain. *Geology* **18**: 435–438.

Dowsett H, Barron J, Poore R. 1996. Middle Pliocene sea surface temperatures: a global reconstruction. *Marine Micropaleontology* **27**: 13–25.

Edwards T, Crucifix M, Harrison S. 2007. Using the past to constrain the future: how the palaeorecord can improve estimates of global warming. *Progress in Physical Geography* **31**: 481–500.

Funder S, Kjaer KH. 2007. Ice free arctic ocean, an early Holocene analogue. *Eos, Transactions of the AGU*, Fall Meeting Supplement **88**: Abstract.

Gao C, Robock A, Ammann C. 2008. Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *Journal of Geophysical Research* **113**: D23111.

Hansen J, Sato M, Kharecha P, Beerling D, Berner R, Masson-Delmotte V, Pagani M, Raymo M, Royer DL, Zachos JC. 2008. Target atmospheric CO_2: where should humanity aim? *Open Atmospheric Science Journal* **2**: 217–231.

Haywood A, Dowsett H, Valdes P, Lunt D, Francis J, Sellwood B. 2009. Introduction: Pliocene climate, processes and problems. *Philosophical Transactions A* **367**: 3–17.

Jansen E, Overpeck J, Briffa KR, Duplessy J-C, Joos F, Masson-Delmotte V, *et al.* 2007. Palaeoclimate,. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL Cambridge University Press: Cambridge, UK.

Jolly D, Harrison S, Damnati B, Bonnefille R. 1998. Simulated climate and biomes of Africa during the Late Quaternary: comparison with pollen and lake status data. *Quaternary Science Reviews* **17**: 629–657.

Jones P, Mann M. 2004. Climate over past millennia. *Reviews of Geophysics* **42**: RG2002.

Joussaume S. 1990. Three-dimensional simulations of the atmospheric cycle of desert dust particles using a general circulation model. *Journal of Geophysical Research* **95**: 1909–1941.

Khodri M, Leclainche Y, Ramstein G, Braconnot P, Marti O, Cortijo E. 2001. Simulating the amplification of orbital forcing by ocean feedbacks in the last glaciation. *Nature* **410**: 570–574.

Kohfeld K, Harrison S. 2001. DIRTMAP: the geological record of dust. *Earth Science Reviews* **54**: 81–114.

Kucera M, Rosell-Melé A, Schneider R, Waelbroeck C, Weinelt M. 2005. Multiproxy approach for the reconstruction of the glacial ocean surface (MARGO). *Quaternary Science Reviews* **24**: 813–819.

Kutzbach JE. 1981. Monsoon climate of the early Holocene: climate experiment with Earth's orbital parameters for 9000 years ago. *Science* **214**: 59–61.

LeGrande AN, Schmidt GA. 2008. Ensemble, water-isotope enabled, coupled general circulation modeling insights into the 8.2-kyr event. *Paleoceanography* **23**: PA3207.

LeGrande AN, Schmidt GA, Shindell DT, Field C, Miller RL, Koch D, Faluvegi G, Hoffmann G. 2006. Consistent simulations of multiple proxy responses to an abrupt climate change event. *Proceedings of the National Academy of Sciences* **103**: 837–842.

Licciardi JM, Clark PU, Jenson JW, MacAyeal DR. 1998. Deglaciation of a soft-bedded Laurentide ice sheet. *Quaternary Science Reviews* **17**: 427–448.

Lunt DJ, Foster GL, Haywood AM, Stone EJ. 2008. Late Pliocene Greenland glaciation controlled by a decline in atmospheric CO_2 levels. *Nature* **454**: 1102–1105.

Mann ME, Bradley RS, Hughes MK. 1998. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**: 779–787.

Mann M, 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.

Miller RL, Schmidt GA, Shindell DT. 2006. Forced annular variations in the 20th century Intergovernmental Panel on Climate Change Fourth Assessment Report models. *Journal of Geophysical Research* **111**: D18101.

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

Morrill C, Jacobsen R. 2005. How widespread were climate anomalies 8200 years ago? *Geophysical Research Letters* **32**: L19701.

Osborn T, Raper S, Briffa K. 2006. Simulated climate change during the last 1,000 years: comparing the ECHO-G general circulation model with the MAGICC simple climate model. *Climate Dynamics* **27**: 185–197.

Otto-Bliesner B, Marshall S, Overpeck J, Miller G. 2006. Simulating Arctic climate warmth and icefield retreat in the Last Interglaciation. *Science* **311**: 1751–1753.

Otto-Bliesner B, Schneider R, Brady EC, Kucera M, Abe-Ouchi A, Bard E, Braconnot P, Crucifix M, Hewitt CD, Kageyama M, Marti O, Paul A, Rosell-Melé A, Waelbroeck C, Weber SL, Weinelt M, Yu Y. 2009a. 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, Joussaume S, Braconnot P, Harrison SP, Abe-Ouchi A. 2009b. Modeling and data syntheses of past climates. *Eos, Transactions of the AGU* **90**: 93.

Overpeck J, Otto-Bliesner B, Miller G, Muhs D, Alley R, Kiehl J. 2006. Paleoclimatic evidence for future ice sheet instability and rapid sea level rise. *Science* **311**: 1747–1750.

Peltier W. 2004. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annual Review of Earth and Planetary Sciences* **32**: 111–149.

Prell WL, Kutzbach JE. 1992. Sensitivity of the Indian monsoon to forcing parameters and implications for its evolution. *Nature* **360**: 647–652.

Prentice I, 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, Webb T III. 1998. BIOM. **6000**: reconstructing global mid-Holocene vegetation patterns fro palaeoecological records. *Journal of Biogeography* **25**: 997–1005.

Ramstein G, Kageyama M, Guiot J, Wu H, Hély C, Krinner G, Brewer S. 2007. How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison. *Climate of the Past* **3**: 331–339.

Ray N, Adams JM. 2001. A GIS-based vegetation map of the world at the Last Glacial Maximum (25,000–15,000 BP). *Internet Archaeology* **11**.

Renssen H, Goosse H, Fichefet T, Campin J-M. 2001. The 8.2 kyr BP event simulated by a global atmosphere–sea-ice–ocean model. *Geophysical Research Letters* **28**: 1567–1570.

Rind DH, Peteet D. 1985. Terrestrial conditions at the last glacial maximum and CLIMAP sea surface temperature estimates: are they consistent? *Quaternary Research* **24**: 1–22.

Schmidt GA, LeGrande AN. 2005. The goldilocks abrupt climate change event. *Quaternary Science Reviews* **24**: 1109–1110.

Schmidt GA, Moyer L. 2008. A new kind of scientist. *Nature Reports Climate Change* 102–103.

Schneider von Deimling T, Ganopolski A, Held H, Rahmstorf S. 2006. How cold was the Last Glacial Maximum? *Geophysical Research Letters* **33**: L14709.

Shindell DT, Schmidt GA, Miller RL, Rind D. 2001. Northern hemisphere winter climate response to greenhouse gas, ozone, solar and volcanic forcing. *Journal of Geophysical Research* **106**: 7193–7210.

Shindell DT, Schmidt GA, Mann ME, Faluvegi G. 2004. Dynamic winter climate response to large tropical volcanic eruptions since 1600. *Journal of Geophysical Research* **109**: D05104.

Siegert M, Dowdeswell J. 2004. Numerical reconstructions of the Eurasian ice sheet and climate during the Late Weichselian. *Quaternary Science Reviews* **23**: 1273–1283.

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds). 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, UK.

Stouffer RJ, Yin J, Gregory JM, Dixon KW, Spelman MJ, Hurlin W, *et al*. 2006. Investigating the causes of the response of the thermohaline circulation to past and future climate changes. *Journal of Climate* **19**: 1365–1387.

Thomas E, Wolff E, Mulvaney R, Steffensen J, Johnsen SJ, Arrowsmith C, White J, Vaughn B, Popp T. 2007. The 8.2 ka event from Greenland ice cores. *Quaternary Science Reviews* **26**: 70–81.

Werner M, Mikolajewicz U, Heimann M, Hoffmann G. 2000. Borehole versus isotope temperatures on Greenland: seasonality does matter. *Geophysical Research Letters* **27**: 723–726.

Wright HE, Bartlein PJ. 1993. Reflections on COHMAP. *The Holocene* **3**: 89–92.

Published in 2009 by John Wiley & Sons, Ltd.