

## Using palaeo-climate comparisons to constrain future projections in CMIP5

G. A. Schmidt, J. D. Annan, P. J. Bartlein, B. I. Cook, Éric Guilyardi, J. C. Hargreaves, S. P. Harrison, Masa Kageyama, A. N. Legrande, B. Konecky, et al.

► **To cite this version:**

G. A. Schmidt, J. D. Annan, P. J. Bartlein, B. I. Cook, Éric Guilyardi, et al.. Using palaeo-climate comparisons to constrain future projections in CMIP5. *Climate of the Past*, European Geosciences Union (EGU), 2014, 10 (1), pp.221-250. <10.5194/cp-10-221-2014>. <hal-01089496>

**HAL Id: hal-01089496**

**<https://hal.archives-ouvertes.fr/hal-01089496>**

Submitted on 2 Dec 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Using palaeo-climate comparisons to constrain future projections in CMIP5

G. A. Schmidt<sup>1</sup>, J. D. Annan<sup>2</sup>, P. J. Bartlein<sup>3</sup>, B. I. Cook<sup>1</sup>, E. Guilyardi<sup>4,5</sup>, J. C. Hargreaves<sup>2</sup>, S. P. Harrison<sup>6,7</sup>, M. Kageyama<sup>8</sup>, A. N. LeGrande<sup>1</sup>, B. Konecny<sup>9</sup>, S. Lovejoy<sup>10</sup>, M. E. Mann<sup>11</sup>, V. Masson-Delmotte<sup>8</sup>, C. Risi<sup>12</sup>, D. Thompson<sup>13</sup>, A. Timmermann<sup>14</sup>, L.-B. Tremblay<sup>10</sup>, and P. Yiou<sup>8</sup>

<sup>1</sup>NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10025, USA

<sup>2</sup>Research Institute for Global Change, JAMSTEC, Yokohama Institute for Earth Sciences, Yokohama, Japan

<sup>3</sup>University of Oregon, Eugene, OR 97403, USA

<sup>4</sup>NCAS-Climate, University of Reading, Whiteknights, P.O. Box 217, Reading, Berkshire, RG6 6AH, UK

<sup>5</sup>Laboratoire d'Océanographie et du Climat: Expérimentation et Approches Numériques/Institut Pierre Simon Laplace, CNRS-IRD-UPMC – UMR7617, 4 place Jussieu, 75252 Paris Cedex 05, France

<sup>6</sup>Centre for Past Climate Change and School of Archaeology, Geography and Environmental Sciences (SAGES), University of Reading, Whiteknights, P.O. Box 217, Reading, Berkshire, RG6 6AH, UK

<sup>7</sup>Macquarie University, Sydney, NSW 2109, Australia

<sup>8</sup>Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre Simon Laplace, CEA-CNRS-UVSQ – UMR8212, CE Saclay l'Orme des Merisiers, 91191 Gif-sur-Yvette, France

<sup>9</sup>Brown University, 324 Brook St., P.O. Box 1846, Providence, RI 02912, USA

<sup>10</sup>McGill University, 805 Sherbrooke Street West Montreal, Quebec, H3A 0B9, Canada

<sup>11</sup>Pennsylvania State University, University Park, PA 16802, USA

<sup>12</sup>Laboratoire de Météorologie Dynamique/Institut Pierre Simon Laplace, 4, place Jussieu, 75252 Paris Cedex 05, France

<sup>13</sup>University of Arizona, Department of Geosciences, Gould-Simpson Building #77, 1040 E 4th St., Tucson, AZ 85721, USA

<sup>14</sup>University of Hawaii, 2525 Correa Road, Honolulu, HI 96822, USA

*Correspondence to:* G. A. Schmidt (gavin.a.schmidt@nasa.gov)

Received: 11 January 2013 – Published in Clim. Past Discuss.: 11 February 2013

Revised: 19 December 2013 – Accepted: 19 December 2013 – Published: 5 February 2014

**Abstract.** We present a selection of methodologies for using the palaeo-climate model component of the Coupled Model Intercomparison Project (Phase 5) (CMIP5) to attempt to constrain future climate projections using the same models. The constraints arise from measures of skill in hindcasting palaeo-climate changes from the present over three periods: the Last Glacial Maximum (LGM) (21 000 yr before present, ka), the mid-Holocene (MH) (6 ka) and the Last Millennium (LM) (850–1850 CE). The skill measures may be used to validate robust patterns of climate change across scenarios or to distinguish between models that have differing outcomes in future scenarios. We find that the multi-model ensemble of palaeo-simulations is adequate for addressing at least some of these issues. For example, selected benchmarks for the LGM and MH are correlated to the rank of future projections

of precipitation/temperature or sea ice extent to indicate that models that produce the best agreement with palaeo-climate information give demonstrably different future results than the rest of the models. We also explore cases where comparisons are strongly dependent on uncertain forcing time series or show important non-stationarity, making direct inferences for the future problematic. Overall, we demonstrate that there is a strong potential for the palaeo-climate simulations to help inform the future projections and urge all the modelling groups to complete this subset of the CMIP5 runs.

## 1 Introduction

The Coupled Model Intercomparison Project (Phase 5) (CMIP5) is an ongoing coordinated project instigated by the Working Group on Coupled Modelling (WGCM) at the World Climate Research Programme (WCRP) and consisting of contributions from over 25 climate modelling groups (and over 30 climate models) from around the world (Taylor et al., 2012). Multiple experiments are being coordinated, including historical simulations (1850–2005), future simulations following multiple representative concentration pathways (RCPs) and crucially, for the first time in CMIP, three sets of palaeo-climate simulations for the Last Glacial Maximum (LGM) (21 ka BP – Before Present), the mid-Holocene (MH) (6 ka BP) and the Last Millennium (850–1850 CE). The palaeo-climate simulations are also part of the Paleoclimate Model Intercomparison Project (Phase 3) (PMIP3) initiative.

The CMIP5/PMIP3 palaeo-simulations are true “out-of-sample” tests in that none of the models have been “tuned” to produce better palaeo-climates. Such tuning is not necessarily unwise (see Schneider von Deimling et al., 2006 for an example), but would complicate some of the potential analyses. Because the same models are being used for both past and future simulations, this archive of model output is a unique resource for research into the connections between model skill and model predictions, and has the potential to greatly improve assessments of future climate change.

There were many uncertainties in regional aspects of future climate projections highlighted in the Intergovernmental Panel on Climate Change (IPCC) 4th Assessment Report (AR4) (Meehl et al., 2007). These affected, for example, the future of sub-tropical rainfall, El Niño–Southern Oscillation (ENSO) changes, potential declines in the North Atlantic meridional circulation, and the fate of Arctic sea ice. Reducing the uncertainties in the projections could therefore have significant real world consequences for both adaptation and mitigation strategies.

There are three main classes of prediction uncertainty which relate to (a) the choice of scenario, (b) internal variability (sometimes described as initial condition uncertainty), and (c) the imperfections in the model (or structural uncertainty) (Hawkins and Sutton, 2009). Scenario uncertainties inevitably grow in importance with time, particularly after about 30 yr due to the timescales associated with economic change, CO<sub>2</sub> residence time and ocean thermal inertia. Initial condition uncertainty is globally important on scales of a few years (and longer at smaller spatial scales) but predictability is fundamentally limited by the chaotic dynamics of the atmosphere and upper ocean. Thus at the multi-decadal time horizon, reducing and/or better characterising structural uncertainty is the only way to potentially reduce overall uncertainty. These structural uncertainties (given a specific scenario of future emissions and other drivers) arise from a combination of model divergence – i.e. a large spread in model

predictions given the same future scenario, and model inadequacy – i.e. models that are collectively either incomplete, inaccurate or are missing processes or feedbacks. The first effect is explicit (though not completely explored) in a multi-model ensemble, while the second is implicit and needs to be assessed independently.

Observations provide the means to test the models and reduce these uncertainties but instrumental records of useful data targets are few (essentially limited to in situ networks of temperature and rainfall prior to the satellite era). Additionally, and perhaps more crucially, changes in the recent past are relatively small compared to projections for the future. Furthermore, the majority of skill metrics in historical (20th century) simulations do not provide much guidance for future projections: models that are either good or bad at simulating some aspect of modern climate – the climatology, seasonal cycle, or interannual variability – often give essentially the same spread for the future (Santer et al., 2009; Knutti et al., 2010b). The reasons for this can range from the tuning procedure in the models, disconnects between the important physics at different timescales or in response to different drivers, or the very different magnitudes of change. Palaeo-climate changes offer a substantially larger signal that is commensurate with projected future changes and although palaeo-climate records are often affected by substantial noise and difficulties in interpretation (Schmidt, 2010), the most robust reconstructions can provide a crucial test of model performance over a wider range than is possible with the 20th century climate alone.

There have been many previous evaluations of palaeo-climate simulations via earlier incarnations of PMIP, as well as in many individual studies (see the review by Braconnot et al., 2012). However, there has been a lack of analyses that quantitatively link future simulations or forecasts with skill or sensitivity in the palaeo-climate simulations (though see Hargreaves et al., 2013 for an example). This is partly because (prior to CMIP5) palaeo-simulations were not done with exactly the same versions of the models being used for future projections and partly due to a lack of suitable reconstructions for model evaluation. This paper is therefore specifically focused on making the connections between palaeo-climate changes and the future rather than on understanding palaeo-climate change for its own sake.

We break this task into three main areas: (1) examples of metrics that are robust across palaeo- and future simulations, where skill in palaeo-climate evaluations builds credibility for the projections going forward; (2) examples of metrics that discriminate between different models in the past and in the future, and thus may be used to weight model projections; and (3) examples where important caveats come into play that prevent constraints from being useful. We specifically include examples in the third section where important caveats currently limit the palaeo-climate constraints to provide guidance to others on pitfalls that can occur.

The scope of the paper is as follows: Sect. 2 discusses some background on dealing with the multi-model ensemble, issues arising from the use of palaeo-climate proxy data and the use of data-synthesis products; Sect. 3 discusses specific examples of skill metrics that may have predictive power in future simulations by showing robust behaviour across palaeo and future experiments; Sect. 4 gives examples that discriminate between future projections; Sect. 5 presents some exploratory analysis of additional potentially useful metrics that are problematic for various reasons; Sect. 6 concludes and discusses the potential for further work in this area.

## 2 Methodologies

### 2.1 Palaeo-climate reconstructions

Many of the problems in dealing with reconstructing climate from palaeo-data are specific to the type of record, the time period and resolution concerned – for instance, annually resolved tree rings have issues distinct from lower resolution ocean sediment or pollen records (e.g. Kohfeld and Harrison, 2000; Ramstein et al., 2007; Jones et al., 2009; Harrison and Bartlein, 2012). There are however a number of general issues that affect the use of such data for model evaluation, including the potential for multiple climate controls on a given record, the scale over which they are representative, the need to quantify (and take into account) reconstruction uncertainties, and the sparse and uneven site coverage.

Records used for palaeo-climate reconstructions are in general influenced by several different aspects of climate as well as, potentially, non-climatic factors. For instance, oxygen or hydrogen isotopes from ice cores, carbonates or organic matter, are climatically meaningful variables, but do not necessarily have a one-to-one, stationary relationship with temperature or precipitation (e.g. Werner et al., 2000; Schmidt et al., 2007; Masson-Delmotte et al., 2011). Vegetation, in addition to being influenced by several aspects of seasonal climate, is directly influenced by the atmospheric CO<sub>2</sub> concentration (Prentice and Harrison, 2009). There are several approaches that have been adopted to overcome this type of problem: the use of multi-proxy reconstruction techniques, forward modelling of the system within a climate model or using climate model output (see an example related to coral carbonate isotopes in Sect. 5.1) or other climate prior, and model inversion or data assimilation. Multi-proxy reconstructions rely on the idea that different types of record will be sensitive to different aspects of climate, and that pooling the information from each of these records therefore provides a more robust reconstruction of any specific climate variable. In the sense that forward modelling (and by extension model inversion techniques) are based on physical and or physiological knowledge of the given system, the use of these approaches may be a more robust way of dealing with

the non-stationarity issue – however, as with climate models, the results are constrained by the quality of the models and the degree to which the system is well-understood (see for example the discussion of CO<sub>2</sub> fertilisation in Denman et al., 2007).

The scale over which a record is representative can be a major issue in comparing palaeo-data and model output. All types of records are responding to basically local conditions, though the scale over which the record is representative will depend greatly on the variable and the resolved timescale. Many records, such as tropical ice core  $\delta^{18}\text{O}$ , may have strong correlations with climate further afield (e.g. Schmidt et al., 2007). Comparisons at local or regional scales often require some form of dynamical or statistical downscaling of model output, though there are many associated issues with this (Wilby and Wigley, 1997). Alternatively, upscaling reconstructions (for instance, through the use of gridding) can often reveal large-scale patterns that models could be expected to resolve, although this requires a sufficiently dense network of sites (see Sect. 3 for examples). Other approaches include the use of cluster analysis to classify types of model behaviour and to determine cohesive regions for comparison with the large-scale patterns in the observations (e.g. Bonfils et al., 2004; Brewer et al., 2007).

Palaeo-climate reconstructions are usually accompanied by estimates of measurement or structural uncertainty. However, in practice these uncertainties have rarely been propagated into large-scale synthetic products (except in terms of non-quantitative quality control measures, see e.g. COHMAP Members, 1988) and even more rarely taken into account when the reconstructions were used for model evaluation. However, quantitative measures of uncertainty have been included in more recent palaeo-climate syntheses (e.g. MARGO Project Members, 2009; Bartlein et al., 2011) and the use of fuzzy-distance measures (Guiot et al., 1999; Harrison et al., 2013) provides an explicit way to take account of data uncertainties if these cannot be expressed with Euclidean distance. It is worth noting that model-data differences cannot be expected to be smaller than the data uncertainties themselves.

### 2.2 Modelling issues

There are two particular issues that are more problematic in palaeo-climate simulations than, for instance, simulations of the 20th century: model drift and forcing uncertainty. The issue of coupled climate model drift arises because of the long ( $\sim$  thousands of years) time required to bring the deep ocean into equilibrium in coupled ocean-atmosphere models. In some cases, insufficient spin-up time may have been allowed before specific experiments are started. While drift also affects transient historical simulations, the magnitude of the forcings in the 20th century means that residual drift is usually a small component of the transient response. For simulations of the last millennium though, the forcings are much

smaller, and drift in the early centuries of the simulation will be a larger fraction of the modelled change (Osborn et al., 2006; Fernández-Donado et al., 2013). One proposal to deal with this is via a correction using the drift in the control simulation (i.e. calculating a smooth trend and removing it from the perturbed simulation prior to analysis). While this works well for temperature, it is not very good for variables that exhibit threshold behaviour such as sea ice extent or precipitation. In practice, this issue needs to be assessed for each proposed comparison.

Second, there are important uncertainties in the forcings used for the palaeo-climate experiments. This is also true for aerosols in the historical simulations but such issues are more prevalent in palaeo-simulations. For example, the magnitudes of solar and volcanic forcing over the last millennium, and the size and height of ice sheets at the LGM are sources of major uncertainty. In the last millennium experiments, multiple forcing choices were proposed (Schmidt et al., 2011, 2012), but few groups have attempted (as yet) to comprehensively explore all the options, and this is also true for uncertainties associated with other time periods. If an insufficient range of different forcings is tested, it is plausible that mismatches between observations and simulations may be wrongly attributed to the model (or observations), when in fact they were related to a misspecified forcing (e.g. Kageyama et al., 2001).

Third, there are many aspects of past climate changes that are (currently) outside the scope of the available modelling within CMIP5 (and more widely). Variability in the last glacial period that involves complex ocean/ice sheet dynamics (such as Dansgaard–Oeschger events) are beyond what can be analysed directly since the CMIP5-class of models does not have sufficiently interactive dynamic ice sheets. There are also common biases across different models that have more to do with the state of computational technology than physics (for instance, poor or non-existent resolution of ocean eddies). Other examples can easily be found.

For clarity in the rest of the text, we define the term “ensemble” to denote the full multi-model database of results across all CMIP5 scenarios (which encompasses all palaeo-climate, historical, idealised and future projection simulations). The future projections used here consist of the four RCP scenarios (rcp26, rcp45, rcp6, rcp85) (future possibilities that correspond roughly to greenhouse gas radiative forcing at the year 2100, relative to the pre-industrial, of 2.6, 4.5, 6.0, and  $8.5 \text{ W m}^{-2}$ , respectively) along with idealised simulations that have been included to provide clean comparisons across models. The idealised simulations include a 1 % increasing  $\text{CO}_2$  simulation, the response to an abrupt increase to  $4x\text{CO}_2$ , atmosphere-only simulations such as amip, amip4xCO2 and amipFuture (where all models are forced by the same pattern of ocean temperatures from the historical period, with  $4x\text{CO}_2$ , and with a warm anomaly imposed respectively), or sstClim and sstClim4xCO2 simulations (where ocean temperatures are held constant under

pre-industrial or  $4x\text{CO}_2$  conditions). We use CMIP5 to refer to the entire database, including the PMIP3 simulations. Specific model simulations are referred to by their name in the CMIP5 database (i.e. rcp85, past1000, piControl etc.), while the scenarios or periods are referred to more generally using a standard abbreviation or name (e.g. the LGM, MH, RCP 4.5). We list the models that we have used in analyses in this paper, along with the specific experiments and simulation IDs, in Table 1. While the multi-model ensemble is a useful source for addressing structural uncertainty, it should be noted that the ensemble is not a controlled sample from a well-defined distribution of plausible simulations.

### 2.3 Approaches to comparing reconstructions and simulations

There has been a gradual evolution in the approaches for comparing reconstructed changes and climate model simulations from essentially qualitative graphical comparisons of output and reconstructions of the corresponding climatic variables (e.g. Braconnot et al., 2007) to more quantitative approaches that measure model-data mismatch via some “metric” or distance function (e.g. Sundberg et al., 2012; Izumi et al., 2013). Metrics based on correlations or rms differences between fields of data and model output have been commonly used in model evaluation for current climate (e.g. Taylor, 2001; Schmidt et al., 2006; Gleckler et al., 2008). These methods provide opportunities for both inter- and intra-generational model comparisons (Reichler and Kim, 2008; Harrison et al., 2013). The concept of “skill” as adopted in the numerical weather prediction community is useful as a quantitative test of model performance: that is, does a model produce a more accurate prediction (match to the palaeo-climate record), than that which would be achieved by a simple null hypothesis (Hargreaves et al., 2013)? Most studies and metrics have focused on time slice or time series comparisons, though it is worth pointing out that nothing precludes comparing the simulations and palaeo-record in the frequency domain (e.g. Lovejoy and Schertzer, 2012b).

While most standard comparisons focus on evaluating individual model simulations against the reconstructions, a different approach is to focus on the collective performance of the ensemble as a whole. For instance, Hargreaves et al. (2011) tested the ability of the PMIP2 ensemble to represent the Last Glacial Maximum in terms of its “reliability”, defined as the adequacy of the ensemble, considered in probabilistic terms, in predicting the changes documented in the palaeo-climate archives during that interval. Multi-model ensemble means can be informative and will generally outperform individual models (Annan and Hargreaves, 2011), but care must be taken to assess the suitability of each included model and (any) weighting of individual models needs to be well justified (Knutti et al., 2010a).

**Table 1.** List of models, institutions and experiments used in the analyses in this paper. Experiment names use the CMIP5 database shorthand, and run numbers are the “rip” coding for each experiment.

Model name	Model institution	Experiments	Run numbers
ACCESS-1.0	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	historical	r1i1p1
BCC-CSM1	Beijing Climate Center, China Meteorological Administration, China	piControl midHolocene rcp85	r1i1p1 r1i1p1 r1i1p1
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	piControl amip sstClim sstClim4xCO2 rcp85	r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1
CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada	historical rcp45	r[1-5]i1p1 r[1-5]i1p1
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France	piControl historical midHolocene lgm IpctCO2 abrupt4xCO2 rcp45 rcp85	r1i1p1 r[1-10]i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organisation in collaboration with the Queensland Climate Change Centre of Excellence, Australia	piControl historical midHolocene rcp45 rcp85	r1i1p1 r[1-10]i1p1 r1i1p1 r[1-10]i1p1 r1i1p1
EC-EARTH	EC-Earth consortium	piControl historical midHolocene rcp45 rcp85	r1i1p1 r7i1p1 r1i1p1 r[1, 2, 6-9, 11, 12, 14]i1p1 r1i1p1
FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University, China	piControl midHolocene rcp85	r1i1p1 r1i1p1 r1i1p1
FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences; and CESS, Tsinghua University, China	amip midHolocene piControl rcp85 sstClim4xCO2 sstClim	r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1
CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici	piControl rcp85	r1i1p1 r1i1p1
GFDL-CM2.1	NOAA Geophysical Fluid Dynamics Laboratory, US	historical	r1i1p1
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory, US	piControl amip amip4xCO2 sstClim rcp85	r1i1p1 r1i1p1 r1i1p1 r1i1p1 r1i1p1

Table 1. Continued.

Model name	Model institution	Experiments	Run numbers
GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory, US	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp85	r1i1p1
GFDL-ESM2M	NOAA Geophysical Fluid Dynamics Laboratory, US	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp85	r1i1p1
GISS-E2-H	NASA Goddard Institute for Space Studies, US	piControl historical	r1i1p1 r[1–5]i1p[12]
GISS-E2-R	NASA Goddard Institute for Space Studies, US	piControl	r1i1p1, r1i1p141
		historical	r1i1p[12], r[45]i1p3
		past1000	r1i1p12[1–8]
		midHolocene	r1i1p1
		lgm	r1i1p15[01]
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp45	r[1–5]i1p1
		rcp85	r1i1p1
HadCM3	Hadley Center, UK Met. Office, UK	historical	r[1–10]i1p1
HadGEM2-CC	Hadley Center, UK Met. Office, UK	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp45	r1i1p1
		rcp85	r1i1p1
HadGEM2-ES	Hadley Center, UK Met. Office, UK	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp45	r[1–3]i1p1
		rcp85	r1i1p1
INM-CM4	Institute for Numerical Mathematics, Russia	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp45	r1i1p1
		rcp85	r1i1p1
IPSL-CM5A-LR	Institut Pierre-Simon Laplace, France	piControl	r1i1p1
		historical	r[1–4]i1p1
		midHolocene	r1i1p1
		lgm	r1i1p1
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp45	r[1–4]i1p1
		rcp85	r1i1p1
IPSL-CM5A-MR	Institut Pierre-Simon Laplace, France	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp45	r1i1p1
		rcp85	r1i1p1

Table 1. Continued.

Model name	Model institution	Experiments	Run numbers
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	piControl	r1i1p1
		midHolocene	r1i1p1
		lgm	r1i1p1
		past1000	r1i1p1
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp85	r1i1p1
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	piControl	r1i1p1
		historical	r1i1p1
		midHolocene	r1i1p1
		rcp45	r[1–3]i1p1
		rcp85	r1i1p1
MPI-ESM-P	Max Planck Institute for Meteorology, Hamburg, Germany	piControl	r1i1p1
		historical	r1i1p1
		past1000	r1i1p1
		lgm	r1i1p1
		midHolocene	r1i1p1
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp85	r1i1p1
		MPI-ESM-LR	Max Planck Institute for Meteorology, Hamburg, Germany
historical	r[1–3]i1p1		
rcp85	r1i1p1		
MPI-ESM-MR	Max Planck Institute for Meteorology, Hamburg, Germany	piControl	r1i1p1
		amip	r1i1p1
		amip4xCO2	r1i1p1
		amipFuture	r1i1p1
		sstClim	r1i1p1
		sstClim4xCO2	r1i1p1
		historical	r[1–3]i1p1
		rcp85	r1i1p1
MRI-CGCM3	Meteorological Research Institute, Tsukuba, Japan	piControl	r1i1p1
		midHolocene	r1i1p1
		lgm	r1i1p1
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp85	r1i1p1
NCAR-CCSM4	National Center for Atmospheric Research, US/Dept. of Energy/NSF	piControl	r1i1p1
		amip	r1i1p1
		amip4xCO2	r1i1p1
		amipFuture	r1i1p1
		sstClim	r1i1p1
		sstClim4xCO2	r1i1p1
		midHolocene	r1i1p1
		lgm	r1i1p1
		1pctCO2	r1i1p1
		abrupt4xCO2	r1i1p1
		rcp45	r1i1p1
		rcp85	r1i1p1
NCAR-CESM1	National Center for Atmospheric Research, US/Dept. of Energy/NSF	historical	r[5–7]i1p1

Table 1. Continued.

Model name	Model institution	Experiments	Run numbers
NCAR-CESM1-CAM5	National Center for Atmospheric Research, US/Dept. of Energy/NSF	piControl	r1i1p1
		amip	r1i1p1
		amipFuture	r1i1p1
		rcp85	r1i1p1
NCAR-CESM1-BGC	National Center for Atmospheric Research, US/Dept. of Energy/NSF	piControl	r1i1p1
		rcp85	r1i1p1
NorESM1-M	Norwegian Climate Centre, Norway	piControl	r1i1p1
		historical	r[1–3]i1p1
		midHolocene	r1i1p1
		rcp45	r1i1p1
		rcp85	r1i1p1
NorESM1-ME	Norwegian Climate Centre, Norway	piControl rcp85	r1i1p1 r1i1p1

## 2.4 Linking past and future

The key task of this paper is to provide guidance and examples for deciding on whether the palaeo-climate simulations have a connection to the future projections, and if so, what the comparison to palaeo-reconstructions can imply for the future. We stress that robust links between past and future simulations can only be derived if the model configurations used are the same in the different experiments. A previously common practice of using a lower resolution or differently tuned or scoped model for past simulations than for future projections, while perhaps convenient for efficiency, is not appropriate because such variations often have led to substantial differences in sensitivity. Thus, all the examples discussed below link models that were identical (excepting boundary conditions and forcings) in the past and future CMIP5 simulations.

We distinguish two ways in which palaeo-data-model comparisons can be used as a guide to the future: (1) as a validation of a robust relationship between diagnostics across models and scenarios, or (2) as a method to discriminate between differently skillful models. In the first case, one would search for properties or correlations that we expect to be features of all climates within the ensemble, determine whether that is the case, and use the palaeo-data to provide some independent support for that relationship. In the second case, there is a prerequisite that for the diagnostic chosen, the “skill” metric when it is compared to a reconstruction actually correlates to future outcomes within the ensemble. If this is not the case, then the skill in that diagnostic is orthogonal to the spread in the projections and cannot be used to constrain them. Even when such a relationship is found, we need to consider whether it is physically meaningful to be confident that it has not arisen either through chance due to a small sample size or as an artifact of the model or the experimental design. To gain confidence in such palaeo-constraint, we

also need to understand the physical processes that explain the connections between past and future.

While connections may in principle be highly complex, it is natural as a first step to consider whether a correlation exists between past and future behaviour in the same diagnostic. The search for useful metrics (in this sense) using modern data has generally been disappointing (Knutti et al., 2010b), although there have been a small number of cases where apparently meaningful relationships have been found (Boé et al., 2009; Hall and Qu, 2006; Brient and Bony, 2012; Fasullo and Trenberth, 2012). It is notable that the first three examples relate future climate changes to externally forced changes in the modern climate (decadal or seasonal variations), rather than using metrics based on the climatological mean state alone. This lends support to our working hypothesis that past variations seen in palaeo-climate simulations will be informative about the future.

Where a credible relationship between past and future is found, there is a range of methods that can be applied to use observations to constrain future predictions (Collins et al., 2012). One method, applied by both Boé et al. (2009) and Hall and Qu (2006), is to take the observational estimate, and use the relationship (often linear) embodied in the correlation between past and future model output to project this value into the future. An attractive feature of this approach, beyond its simplicity, is that it readily allows extrapolation of the observed relationship in the case where the true value is suspected of lying outside the model range. An alternative approach, which has been widely applied to perturbed physics ensembles, is more explicitly Bayesian and considers the ensemble as a probabilistic sample. For the prior, equal weight is typically assigned to each ensemble member. Probabilistic weights are then calculated for each member of the ensemble, according to their performance in reproducing the observations. This weighted ensemble now represents the posterior estimate of future change. This method uses the model

spread as a prior constraint which, depending on one's viewpoint and the specific case in question, may be considered either a strength or weakness (Collins et al., 2012).

### 3 Robust relationships in past and future simulations

In this section we highlight examples of physically based correlations between key diagnostics that show similar relationships in the palaeo-climate simulations and in future projections (or the more idealised warming scenarios) and whose fidelity can be assessed using the palaeo-climate record. If these conditions are realised, the observations can be used to support the model results, and thus help provide contingent future predictions of one diagnostic given a potential change in the other.

An important issue for assessing future climate impacts is to what extent the large-scale mean temperature response can be used as an index for more regional changes. We consider the relationships between global mean temperature and temperature changes in the tropics and other regions in Sect. 3.1, and relationships between land and ocean temperatures in Sect. 3.2.

#### 3.1 Relationships between regional and global temperature change

A common feature in future and palaeo-simulations is that some parts of the world warm or cool at different rates. In future climate simulations, the high latitudes warm more than the low latitudes, as is also observed during the recent instrumental era. This “polar amplification” is also present in LGM simulations and data, with a stronger cooling in the high latitudes than in the tropics (Masson-Delmotte et al., 2006a, b). Izumi et al. (2013) investigate high vs. low latitude temperature changes in lgm, midHolocene, historical, 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> PMIP3/CMIP5 simulations and find broadly consistent relationships for lgm, historical and increased GHG forcings, between mean annual SST changes, w.r.t to piControl, over the northern extratropics and the northern tropics. However, the relationship is not consistent for mean annual air temperatures in the lgm simulations compared to the others because of the particular impact of the northern ice sheets. Here we examine the relationships between changes in global mean temperature and change over large-scale regions. The uneven distribution of the palaeoclimatic reconstructions indeed suggests a focus on specific regions, rather than the globe.

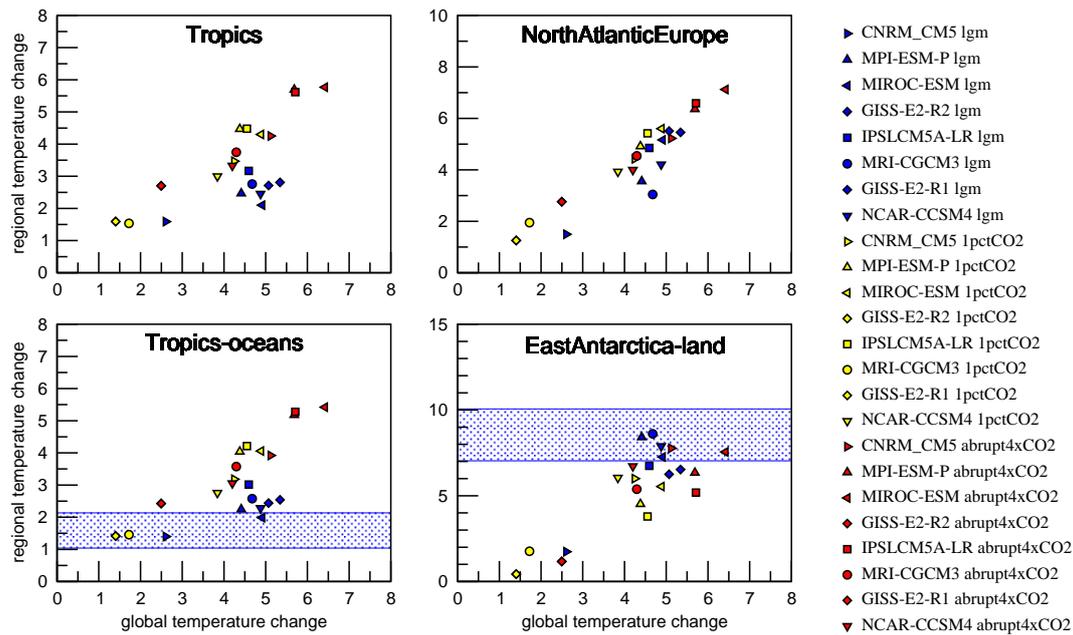
The main climate forcings for the LGM are the lower concentrations in atmospheric greenhouse gases and the presence of Laurentide and Fennoscandian ice sheets in the northern extratropics. The ice sheets have a strong local albedo effect (e.g. Braconnot et al., 2012) but also affect the mid-latitude large-scale atmospheric circulation due to the associated change in topography (e.g. Pausata et al., 2011;

Rivière et al., 2009; Laine et al., 2009). However, away from the direct ice sheet perturbations, we expect that the greenhouse gas forcing would be the main forcing for the LGM climate change and thus patterns of response may be similar to future warmer climates (Hewitt and Mitchell, 1997).

We analyse the comparison between the mean annual surface air temperature change over a region compared to the global mean change for the abrupt4xCO<sub>2</sub>, 1pctCO<sub>2</sub> and lgm CMIP5 simulations from the 8 models for which the results were available at the time of the analysis. We have considered the tropics (land + oceans) and the tropical oceans as targets, because they have been used previously in perturbed physics ensemble studies (Schneider von Deimling et al., 2006; Hargreaves et al., 2007), East Antarctica, for which the temperature change has been shown to scale with global temperature change for the LGM and the CMIP3 2xCO<sub>2</sub> and 4xCO<sub>2</sub> changes (Masson-Delmotte et al., 2006a, b) and the mid-latitude region of the North Atlantic and Europe.

Figure 1 shows a clear relationship between the tropical and global temperature change for the 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> anomalies, both for the combined land and ocean grid cells (top-left panel) and for ocean grid cells alone (bottom-left panel), and this relationship is consistent across these two experiments. The relationship for the LGM is ambiguous because the results for 7 out of the 8 models cluster around the same values. These appear to fall outside the relationship which can be derived from the 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> simulations, with a smaller LGM tropical temperature change for a given global temperature change. This may be because of an outside influence of the LGM northern hemisphere ice sheets on the global mean for this particular climate. Furthermore, the models which simulate the smallest (largest) warming for increased CO<sub>2</sub> are not those which simulate the smallest (largest) cooling for LGM. This implies that either the impact from the lower GHG concentrations are not symmetric compared to those for increased GHG concentrations, or that the ice sheet remote impact extends to the tropics (as inferred by Laine et al., 2009). The relationship appears more consistent across experiments for East Antarctica (Fig. 1, bottom-right panel) and, surprisingly given the proximity of the ice sheets, over the North Atlantic/Europe region (Fig. 1, top right panel).

In the second row of Fig. 1, we indicate the range of the reconstructed LGM regional response. In the case of the tropical oceans (bottom-left plot), this range is computed from the MARGO (2009) data. Uncertainties are derived using a bootstrap method, randomly drawing 1000 samples (of random size, and with replacement) from the initial MARGO data set. For each drawn site, we assume a Gaussian probability function centered on the mean reconstruction and with a standard deviation equal to the uncertainty given in the data set and we draw a possible value considering this probability distribution function. We obtain 1000 estimates of the mean value and compute its mean  $\pm 2$  standard deviations, which defines the shaded blue band on the Fig. 1.



**Fig. 1.** Model average regional vs. global temperature changes for the glacial (in blue, the pre-industrial – LGM difference is shown), years 120 to 140 of the 1pctCO<sub>2</sub> simulations (in yellow, in comparison to piControl) and years 100 to 150 of the abrupt4xCO<sub>2</sub> simulations (in red, in comparison to piControl). For the bottom plots, the regional model average is taken only over grid boxes that correspond to proxy data sites within the defined region (reconstruction range shown in blue shading). For the tropical oceans (bottom-left panel), the blue shaded band shows the average  $\pm 2 \times$  standard deviations of the present – LGM warming as evaluated by the bootstrap method from the MARGO (2009) reconstructions, taking into account the uncertainty on the reconstructions (see main text). For East Antarctica, the blue shaded band corresponds to the range of available reconstructions (5 sites)  $\pm 1^\circ\text{C}$  (Braconnot et al., 2012). Definition of the regions: Tropics:  $23^\circ\text{S}$ – $23^\circ\text{N}$ ; North Atlantic Europe:  $45^\circ\text{W}$ – $90^\circ\text{E}$ ;  $35$ – $45^\circ\text{N}$ ; East Antarctica:  $5^\circ\text{W}$ – $165^\circ\text{E}$ ;  $70$ – $80^\circ\text{S}$ . The results have been computed for all models in the database on 23 July 2012 for which there were results for the lgm, piControl, 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> simulations.

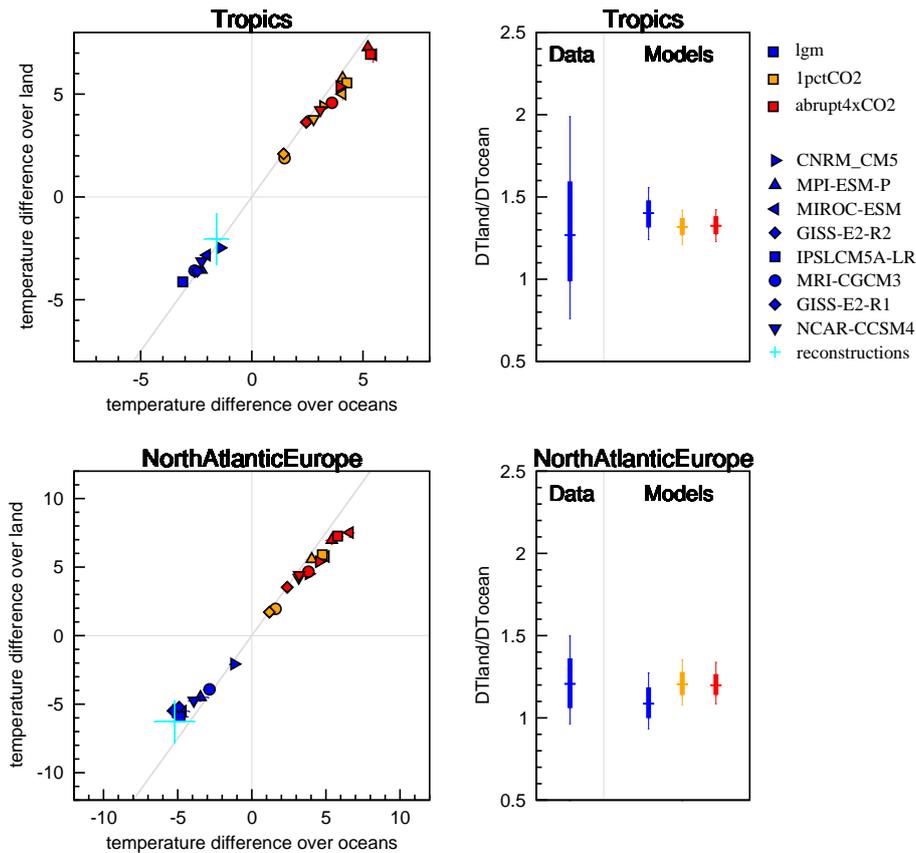
For East Antarctica, we “only” have 5 points, so we simply consider the uncertainty of  $\pm 1^\circ\text{C}$  on the reconstruction (Masson-Delmotte et al., 2006) and the range of available reconstructions. In both cases, the available data discriminate between the models, with 2 models out of 8 falling in the range of the reconstructions for the tropical oceans and of 4 models out of 8 in the case of East Antarctica.

In summary, the range of model results for increased CO<sub>2</sub> scenarios shows that there is a relationship between regional and global temperature changes for all regions considered here. The range of simulated LGM regional/global average temperature change is smaller than in the increased CO<sub>2</sub> runs. The results are consistent with the relationship derived in future scenarios for East Antarctica and the North Atlantic/Europe region. For the tropics, the LGM ratio is smaller than that seen in future scenarios, which could be due to the impact of ice sheets on the global mean temperature change. Both data and models suggest an amplification of changes from the tropics to Antarctica and the data can help constrain the global LGM temperature change to 4.2 to 5°C, but only weakly constrain the expected sensitivity to abrupt4xCO<sub>2</sub> forcing (from 4.2 to 6.5°C). Additional sensitivity experiments will be needed to test the individual impacts of CO<sub>2</sub> and ice sheets and better understand the full

LGM response and the inter-model differences. These results are based on only 8 models and will need to be revisited when a larger number of simulations are available.

### 3.2 Land–ocean contrasts

Even though models show biases in the LGM when directly compared to reconstructions (Fig. 1) there are large-scale relationships which appear to be consistent for different climates. For instance, model results have consistently shown that for the LGM, the continents cooled more than the ocean (e.g. Braconnot et al., 2007, 2012; Lañé et al., 2009), while, in a symmetric manner, predictions for future climate show a stronger warming over land than over the oceans (e.g. Sutton et al., 2007; Drost et al., 2012). The ratio between cooling over non-glaciated land and cooling over the ocean for the LGM tropics was  $\sim 1.3$  in the PMIP1 computed sea surface temperature (SST) simulations (Pinot et al., 1999), a result close to the ratio of  $\sim 1.5$  found in both the PMIP2 fully coupled LGM experiments (Braconnot et al., 2012) and CMIP3 future projections (Sutton et al., 2007). Izumi et al. (2013) evaluated this land-sea ratio from the CMIP5 lgm, piControl, historical, 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> simulations and found consistent land-sea ratios for global changes and for



**Fig. 2.** Left panels: average surface air temperature change, compared to piControl, over land compared to over the oceans for the tropics ( $23^{\circ}\text{S}$ – $23^{\circ}\text{N}$ ) and the North Atlantic and Europe region ( $45^{\circ}\text{W}$ – $90^{\circ}\text{E}$ ,  $35$ – $45^{\circ}\text{N}$ ). LGM – piControl in blue, 1pctCO<sub>2</sub> – piControl in orange, abrupt4xCO<sub>2</sub> – piControl in red. For the latter 2 periods, the averages have been computed over the same years as Fig. 1. The results have been computed for all models in the database on 23 July 2012 for which there were results for the lgm, piControl, 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub>. The grey lines indicate the 1 : 1.5 ratio in both plots. The results from the reconstructions are based on the MARGO (2009) data for the oceans and on the Bartlein et al. (2011) data for the continents. The error bars show twice the standard deviation of the distribution of the mean temperature changes over the region, as estimated by the same bootstrap method as described for the bottom-left plot of Fig. 1 and in Sect. 3.1. Right panels: plots show the land-sea ratio computed from the data, again with the same bootstrap method. The horizontal bar shows the median ratio, the thick vertical line shows the 25–75th percentiles and the thin lines the 10–90th percentiles. The model results are computed by a bootstrap method applied on all model results, as explained in the text. The definition of the thin/thick vertical lines and horizontal bar are the same as for the data.

the northern and southern extratropics as well as for the tropics, though the ratio varies with latitude and is smallest in the tropics.

Figure 2 shows temperature changes over land vs. oceans for the tropics and North Atlantic/Europe. As also shown by Izumi et al. (2013), the relationship between temperature changes over land and over the oceans appears to be broadly consistent for the lgm, 1pctCO<sub>2</sub> and abrupt4xCO<sub>2</sub> results, though the degree of agreement in the land-sea contrast varies across regions. Furthermore, even though models appear to overestimate temperature changes both over land and oceans in the tropics, and to underestimate them in the North Atlantic/Europe region, the land-ocean ratio appears to be consistent with the data. The right-hand panels in Fig. 2 were built by estimating the distribution of the ratios between

mean temperature changes over land and over the oceans using a bootstrap method on the data from Bartlein et al. (2011) for continental data and MARGO (2009) for the ocean data and taking their uncertainties into account (as for Fig. 1). A similar approach is used for the models, selecting only from points where target reconstructions exists. Coefficients are taken from a linear regression constrained to pass through the origin from 1000 trials. The results show that for the tropics, the land-sea ratio is consistent for the different periods, and model-derived ratios are themselves consistent with the reconstructions. This is also true for the North Atlantic/Europe region, although in this case, the LGM results are more offset from the increased CO<sub>2</sub> results.

We conclude that these relationships are robust, although the reasons for this appear to be imperfectly understood (Lambert et al., 2011) and will require, as for the results from Sect. 3.1, additional sensitivity and process-based analyses. Incidentally, it is worthwhile to note that the land–ocean relationship was previously used to highlight the inconsistency between an earlier compilation of tropical LGM sea surface temperatures and adjacent continental reconstructions (Rind and Peteet, 1985).

#### 4 Palaeo-derived measures of skill that discriminate between models

In this section we highlight diagnostics for which we have commensurate palaeo-climate information and for which the skill metrics across the ensemble serve to discriminate between models that show different behaviours in future projections. This requires that we demonstrate that differences in future sensitivity are correlated to past sensitivities, and that palaeo-reconstructions exist that can effectively weight the projections from models with more realistic sensitivity in the past more highly in an ensemble projection. We illustrate this with three examples: in Sect. 4.1, we look at a simple binary grouping of model behaviour related to South American rainfall that can be evaluated using information from the mid-Holocene. Section 4.2 revisits attempts to constrain overall climate sensitivity using information from the LGM, and Sect. 4.3 looks at the potential to estimate sea ice sensitivity to Arctic warming through results from the mid-Holocene.

##### 4.1 Rainfall change in South America

Projections of precipitation change in South America have a large spread in the CMIP3 (Meehl et al., 2007) and CMIP5 (Knutti and Sedláček, 2012) archives. In future projections, most models simulate a dipole of precipitation change in northern South America. However, the sign and magnitude of this dipole depends on the model: some models simulate drier conditions in Guyana, Venezuela and Colombia and wetter conditions in Nordeste and eastern Brazil, while some model simulate the opposite changes (Fig. 3).

We define the precipitation dipole as the annual-mean precipitation averaged over 0–8° N; 50–60° W (hereafter “Guyana”) minus the annual-mean precipitation averaged over 5–15° S, 35–45° W (hereafter “Nordeste”). We divide 28 different models from the CMIP5 archive into two equal groups. Models where the dipole is weak or negative in the changes in precipitation between rcp85 and piControl are placed in group 1; models which have a strong positive dipole are in group 2. All of the models simulate similar patterns of present-day precipitation, although models in group 2 tend to have a more pronounced double ITCZ. Among the models,

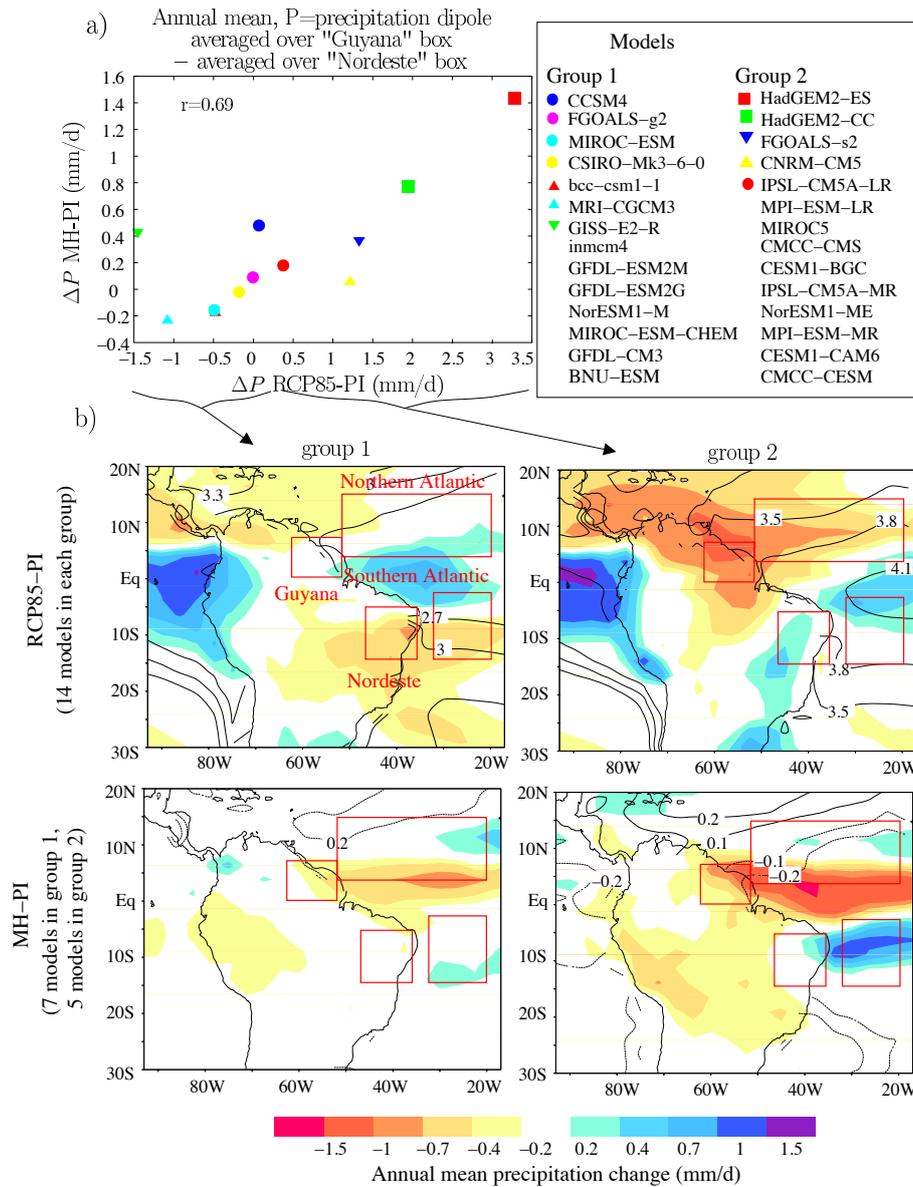
midHolocene output was available for 7 models in group 1 and for 5 models in group 2.

Figure 3 shows a link between precipitation change in the future and in the MH. Models in group 1 simulate wetter conditions in Guyana and drier conditions in Nordeste, associated with a northward shift of the ITCZ in the rcp85 and a broadening of the ITCZ in the MH simulations. Conversely, group 2 models simulate drier conditions in “Guyana” and wetter conditions in “Nordeste”, associated with a southward shift of the Intertropical Convergence Zone (ITCZ). They show a similar dipole in the MH, with a strong southward shift of the ITCZ. Thus the models from a particular group show essentially the same change in the dipole pattern and the same shift in the ITCZ in both future and MH simulations. These patterns are robust relative to the numbers of groups or the number of models included in any group. Palaeo-data from South America show drying everywhere except northeastern Brazil (Prado et al., 2013), a response which is more consistent with group 1 than group 2.

The processes underlying these patterns can be investigated using a variety of other CMIP5 simulations. Table 2 shows correlations between precipitation changes and other features of the simulations. Shifts in the ITCZ in the future projections are associated with shifts in the SST dipole in the Atlantic: models that shift the ITCZ the furthest southwards are those with the strongest warming south of the Equator relative to the rest of the Atlantic. However, while ITCZ shifts in response to SST dipoles are expected (e.g. Kang et al., 2008), this is not the dominant pattern for the MH to PI change. Some of the model behaviours are seen in the amipFuture and sstClim4xCO2 simulations, indicating that the intrinsic response of the atmosphere to a given SST change plays a key role in the formation of the dipole. This is consistent with the fast atmospheric response to CO<sub>2</sub> being an important component of the total precipitation response in global warming (e.g. Bala et al., 2010; Bony et al., 2013). Models that have reduced precipitation over northern South America in the MH simulations also have reduced precipitation in the projections and under 4xCO<sub>2</sub>. These models have the strongest land surface warming in response to both 4xCO<sub>2</sub> and MH forcing. Although the precipitation response of the different groups of models to a change in forcing differs, within each model group the response to different forcing (SST changes, orbital forcing, 4xCO<sub>2</sub>) is similar. This suggests that common mechanisms are involved in the precipitation response to all forcings, and that we can expect future changes to resemble those predicted by the group 1 models. A more quantitative assessment of these changes still remains to be finalised.

##### 4.2 LGM constraints on climate sensitivity

The LGM has been a prime target for assessments of climate sensitivity since it is a quasi-stable period with significant climate differences from today, with reasonably



**Fig. 3.** (a) Relationship between the precipitation dipole change from pre-industrial to future climate under RCP 8.5 for the 2080–2100 and the precipitation dipole change from pre-industrial to mid-Holocene. The precipitation dipole is defined as the difference of precipitation change in RCP 8.5 between the “Guyana” region and the “Nordeste” region. Only those models within each group that had both rcp85 and midHolocene data available at the time of the analysis are plotted. Other models that provided only rcp85 data are listed for completeness, but without any markers. (b) Maps of precipitation changes from piControl to rcp85 (top panels) and from piControl to midHolocene (bottom panels) in average over all available models in group 1 (left panels) and in group 2 (right panels). Contours show corresponding SST changes. The boxes over land and ocean show the areas used in the dipole definitions.

well-known boundary conditions and sufficient data to reconstruct large-scale climate shifts (e.g. Lorius et al., 1990; Edwards et al., 2007; Köhler et al., 2010; Schmittner et al., 2011; PALAEOSENS Project Members, 2012). This provides a good opportunity to apply the methods described in Sect. 2 as a proof-of-concept estimate of the equilibrium climate sensitivity based on the CMIP5 LGM simulations.

We use an ensemble of opportunity consisting of 7 models which participated in the PMIP2 experiment, together with 7 CMIP5 models for which sufficient data were available (at time of writing). Estimates of the climate sensitivities of these models were obtained from a variety of sources and were derived using a range of methods: For the PMIP2/CMIP3 models, sensitivity was generally calculated using a slab ocean coupled to the atmospheric component

**Table 2.** Correlation of different variables with future precipitation change in the RCP 8.5 scenario. Precipitation changes are defined as in Fig. 3: annual-mean precipitation averaged over 0–8° N; 60–50° W minus annual-mean precipitation averaged over 5–15° S; 45–35° W; SST dipole changes are defined as the annual-mean change in SST averaged over 3–15° N; 50–20° W minus the annual-mean change in SST averaged over 3–15° S; 20–30° W (see boxes on Fig. 3b). Land surface warming is the annual-mean warming averaged over 0–15° S; 50–70° W. The double ITCZ index is defined as the annual-mean precipitation averaged over the southern branch (3–7° S; 20–35° W) minus the annual-mean precipitation averaged over the northern branch (3–7° N; 20–35° W).

Variable	Correlation ( <i>r</i> )	No. of models	<i>p</i> value
midHolocene-piControl: $\Delta$ precip	0.69	12	0.013
rcp85-piControl: $\Delta$ SST dipole	0.47	28	0.012
midHolocene-piControl: $\Delta$ SST dipole	0.22	12	Not significant
amipFuture-amip: $\Delta$ precip	0.55	9	0.125
sstClim4xCO2-sstClim: $\Delta$ precip	0.67	12	0.017
sstClim4xCO2-sstClim: $\Delta$ SAT (land)	0.62	12	0.032
midHolocene-piControl: $\Delta$ SAT (land)	0.55	12	0.063
double ITCZ index in piControl	–0.48	28	0.010

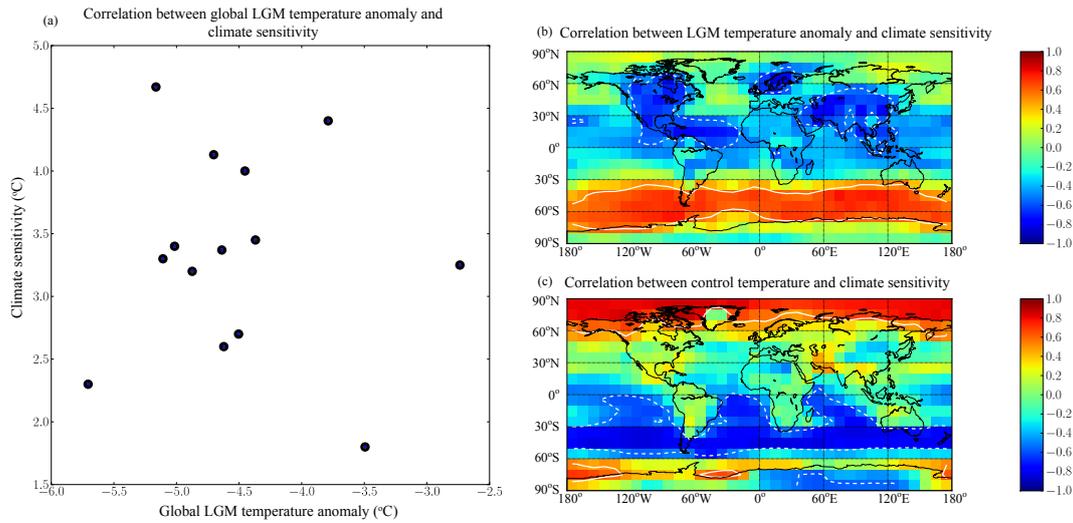
(Meehl et al., 2007), whereas in CMIP5, the most readily available estimates use a regression based on a transient simulation (Andrews et al., 2012). These estimates are not perfectly commensurate, with some models reporting a 10% difference in the two methods (e.g. Schmidt et al., 2014). Unfortunately, some of the PMIP2 models used for the LGM simulations differ from the CMIP3 versions for which the sensitivity estimates were made (for example, MIROC3.2). Thus, while the values used here may be somewhat inconsistent and imprecise, we expect the uncertainty arising from these sources (around 0.5 °C) to be modest in comparison to the range of values represented across the ensemble (roughly 2–5 °C). The boundary conditions for the LGM simulations are essentially unchanged between PMIP2 and CMIP5 (save for changes in the shape of the imposed ice sheets), allowing us to consider these experiments as broadly equivalent though there are some systematic biases due to the total ice volume and resulting changes in land/sea mask (Kageyama et al., 2013). Limitations in the boundary conditions (such as the exclusion of dust and vegetation effects) which we do not attempt to account for here, could introduce additional bias and uncertainty into our result. For these and other reasons discussed below, these results should be considered as a proof of concept rather than conclusive.

The LGM was associated with a large negative radiative forcing with respect to the pre-industrial including substantially lower concentrations of greenhouse gases (e.g. Köhler et al., 2010). However, the ensemble does not show the expected negative correlation between climate sensitivities and their globally averaged LGM temperature anomalies (over the full 100 yr of simulation output) (Fig. 4a, see also Crucifix, 2006). Hargreaves et al. (2012) analysed the PMIP2 ensemble on a regional basis and found their LGM temperature changes in the tropics to exhibit a negative correlation with climate sensitivity, most strongly in the latitude band 20° S–30° N. Results from the PMIP3 models are consistent

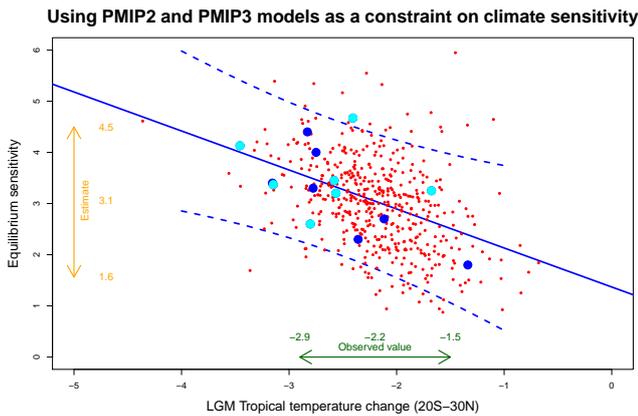
with this relationship, but do not strengthen it. When we combine all models into one ensemble, the correlation over this region weakens to –0.54 but it is significant at the 95% level.

The correlation is generally insignificant at higher latitudes where the feedbacks in response to large cryospheric changes may be very different to those exhibited in a future warmer climate. There is also a strong positive correlation in the southern ocean (i.e. colder LGM anomalies are linked with *lower* sensitivity), possibly due to the large range of biases in the control climate (Fig. 4c). The correlation of piControl temperatures to sensitivity points to the Arctic and the southern oceans as regions where base climatology strongly impacts sensitivity, probably via cloud effects (see Trenberth and Fasullo, 2010 for a discussion). The significant negative correlation between the LGM temperature anomalies in the latitude band 20° S–30° N, and the climate sensitivities of the models (Fig. 5), is physically plausible, since this region is far from the cryospheric and sea ice changes of the LGM, and the forcing here is dominated by the reduction in greenhouse gas concentrations. Assuming that the correlation with tropical temperatures provides a valid constraint on the real climate system, we can use this correlation to project a reconstruction of past change onto the future, as in Boé et al. (2009).

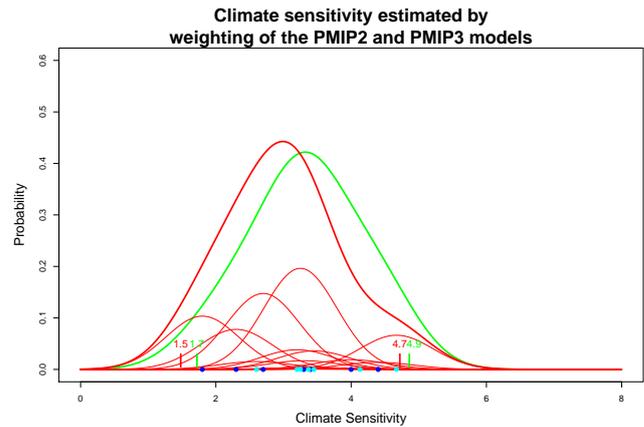
Annan and Hargreaves (2013) generated a new estimate of LGM temperature changes, based on a combination of several multiproxy data sets, and the ensemble of PMIP2 models. The method does not depend on the magnitude of changes estimated by the models, but only their spatial patterns. Due to the suspicion that the tropical temperatures at the LGM from the MARGO synthesis are too warm (e.g. Telford et al., 2013), we focus here on the results from the sensitivity test in Annan and Hargreaves (2013), where reconstructed tropical temperatures were decreased uniformly by 1 °C. Using the resulting estimate of LGM



**Fig. 4.** (a) Global mean LGM temperature change versus overall climate sensitivity to  $2\times\text{CO}_2$ . (b) correlation between local LGM air temperature anomaly and climate sensitivity across the model ensemble. (c) correlation across the model ensemble between control run temperatures and climate sensitivity.



**Fig. 5.** Using LGM tropical temperature as a constraint on climate sensitivity. Cyan and blue dots represent PMIP2 and CMIP5 simulations respectively. Linear regression and predictive uncertainty range are plotted as solid and dashed blue lines respectively. Small red dots represent a Monte Carlo sample from the estimated proxy-derived reconstruction, mapped onto the climate sensitivity.



**Fig. 6.** Climate sensitivity estimated through weighting of the PMIP2 and PMIP3 models. Blue and cyan dots represent PMIP2 and CMIP5 simulations respectively. Green curve shows prior distribution of climate sensitivity (based on equal weighting of the models). Thick red curve shows posterior distribution, after weighting according to match to the LGM tropical temperature. Thin red curves show the individual models’ contributions to the posterior after weighting. Vertical bars indicate 5, 50 and 95 percentiles.

temperature change in the  $20^\circ\text{S}$ – $30^\circ\text{N}$  latitude band of  $-2.2 \pm 0.7^\circ\text{C}$  (at 90 % confidence), the predicted value for climate sensitivity arising from the correlation is  $3.1^\circ\text{C}$ , with a 90 % interval of  $1.6$ – $4.5^\circ\text{C}$  calculated by Monte Carlo sampling, but this range is sensitive (by up to  $0.4^\circ\text{C}$ ) to the reconstruction uncertainties (Annan and Hargreaves, 2013).

In a more explicitly Bayesian approach, we can initially assign equal probability to each model in the ensemble. This choice can be questioned, given both the range of model complexities, and also the possible inter- or intra-generational similarities between models of related origins (Masson and Knutti, 2011). However, quantifying these

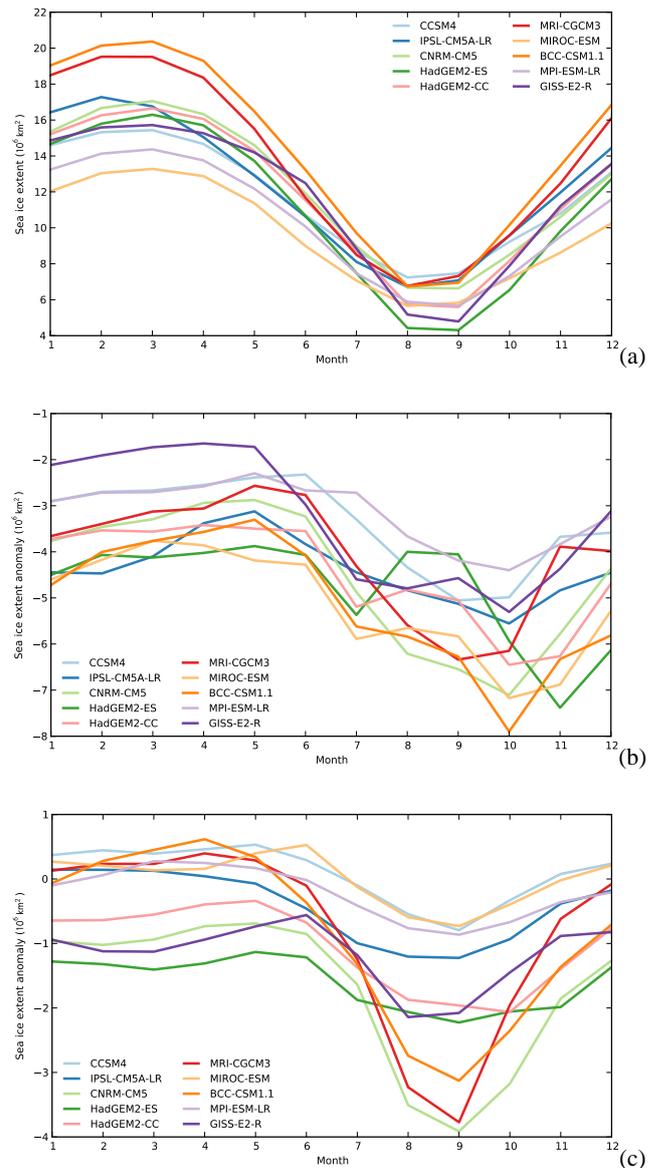
issues is far from straightforward, so we make our choice for reasons of practicality and in order to demonstrate the utility of the overall method. A standard kernel density estimation based on the ensemble leads to the prior distribution presented as the green curve in Fig. 6, which has a 90 % range of  $1.7$ – $4.9^\circ\text{C}$  and a median of  $3.3^\circ\text{C}$ . The observationally derived estimate of tropical temperature gives rise to the natural Gaussian likelihood function  $L(M|\bar{O}) = P(O|\bar{M})$  from which the weights are calculated (where  $O$  represents

the reconstructed observations,  $M$  the model simulation,  $L(M|\bar{O})$  the likelihood of the model result given the reconstructed data, and  $P(O|\bar{M})$  the probability of the reconstructions assuming the models are correct). The posterior distribution is shown in red, the bulk of which has been shifted to lower values with the median reducing to  $3.0^\circ\text{C}$ . Its 90% probability range has moved slightly less to  $1.5\text{--}4.7^\circ\text{C}$ . The reason for the upper limit here remaining high is that the highest sensitivity model in the ensemble has been assigned a fairly large weight since it matches the reconstruction well. The small size of the ensemble means that this approach is rather sensitive to the presence or absence of particular models in the ensemble.

These two approaches differ considerably in their use of the model ensemble. In the latter case, the ensemble is directly used as a prior estimate, which therefore imposes quite a strong constraint on climate sensitivity even before the observational constraints are used. The former method may be considered as roughly equivalent to using a prior that is uniform in the observed variable (here tropical temperature), although this approach is rarely presented in explicitly Bayesian terms. Despite the different assumptions and approaches, these methods both generate similar estimates for the climate sensitivity – both assigning higher probability towards the lower end of the model range. The ranges are comparable with other palaeo-climate-derived estimates of  $2.3\text{--}4.8^\circ\text{C}$  (68% confidence interval, PALAEOSENS Project Members, 2012) but, given the small ensemble size and possible naïveté of the assumptions made here, these estimates may not be robust and need to be tested using a larger ensemble.

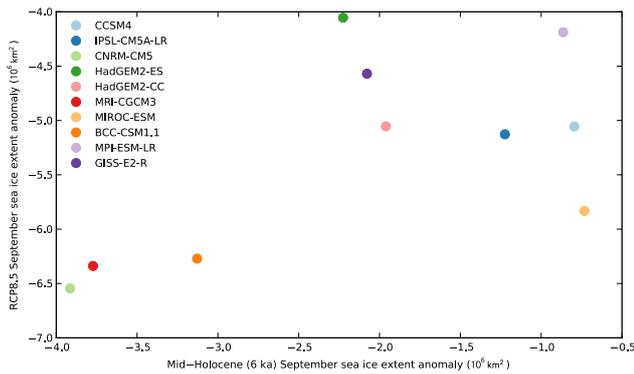
### 4.3 Arctic Sea ice sensitivity constraints from the mid-Holocene

The rate and pattern of Arctic sea ice change in future decades is of interest due both to the surprisingly rapid changes currently occurring and the large spread in model estimates in, for instance, the onset of summertime “ice-free” conditions (Stroeve et al., 2012; Massonet et al., 2012). Recent studies (Mahlstein and Knutti, 2012; Abe et al., 2011) have demonstrated that biases in sea ice volume have a strong impact on the simulated responses to radiative perturbations, and that there may be a possibility to discriminate among models based on interannual modes of sea ice variability. The mid-Holocene simulations (driven mainly by changes in orbital forcing) provide an orthogonal test of Arctic sea ice sensitivity since MH insolation changes imply that NH summers were warmer than summers today (see Kutzbach, 1981 and many subsequent papers). Palaeo-data from the circum-Arctic region indicates that this warmth was accompanied by reductions in sea ice extent at least during some months of the year (Dyke and Saville, 2001; de Vernal et al., 2005; McKay et al., 2008; Funder et al., 2011; Polyak et al., 2010; Moros et al., 2006).



**Fig. 7.** Sea-ice extent in CMIP5 models in  $10^6 \text{ km}^2$ . **(a)** 30-yr mean seasonal cycle for the period 1870–1900, **(b)** the anomaly in sea ice extent for the period 2036–2065 in RCP 8.5, and **(c)** the anomaly at the mid-Holocene.

The CMIP5 MH simulations (Fig. 7c) consistently show decreases in sea ice extent from July/August through to November relative to the pre-industrial. Changes in winter months (December–February) do not agree in sign across the models, though these changes are not well characterised in the palaeo-data either. There is a relationship (Fig. 8) between the size of the anomaly at the MH and in future projections (using 2036–2065 in rcp85), presumably reflecting the underlying sensitivity of the sea ice model and Arctic climate in general (see also O’ishi and Abe-Ouchi, 2011). We focus on the 30 yr period centered on 2050 since that is when there



**Fig. 8.** Relationship between the September RCP 8.8 anomaly (derived from Fig. 7b) and the September MH anomaly (from Fig. 7c) across the CMIP5 models.

is a very large spread in individual model projections in the RCP 8.5 simulations (Fig. 7).

This correlation exists despite the variations in the cause of the ice loss (summer insolation versus greenhouse-gas-related forcing). Although the small size of the ensemble raises questions about the robustness of the relationships, the MH ice extent anomaly can be used to estimate the likely loss in future projections. Given the qualitative nature of the palaeo-data, we are not yet able to make a quantitative projection, but there is some support given to the models with a larger projected change. A more detailed approach using more specific and local diagnostics in comparison to a wider proxy network will likely give a more quantitative result (Tremblay et al., 2014).

## 5 Exploratory metrics: potential and limitations

While the examples given above show direct connections between past and future in ways that can be used relatively straightforwardly, there are a number of reasons why other diagnostics may not be as useful. In this section we provide examples of where the palaeo-climate information has yet to be explored, is ambiguous, or where connections seen in palaeo-climate changes do not translate easily into the future for some reason. This may be related to forcing ambiguities, climate-change-related non-stationarity in climate/proxy relationships, or potentially, a poor understanding or representation of the dominant processes. While these examples are not directly informative about the future, they illustrate how the palaeo-simulations can be explored in ways that illuminate key uncertainties and, potentially, provide more opportunities in the future.

Section 5.1 focuses on the potential for shorter-term extremes in temperature and precipitation at the regional scale to be predicted by large-scale seasonal anomalies. Section 5.2 deals with the issue of the evaluation of models over the historical period in the tropical Pacific using forward

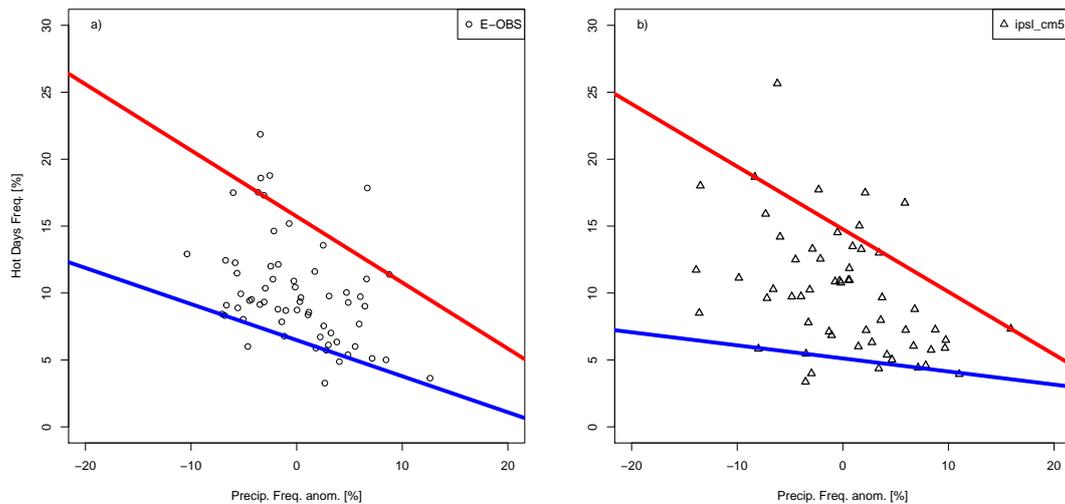
models of coral-based proxies. Section 5.3 addresses diagnostics in the frequency domain that are strongly affected by uncertainties in the forcing fields, rather than intrinsic properties of the models. Finally, Sect. 5.4 provides an example of how connections between past and future hydroclimate diagnostics may be non-stationary.

### 5.1 Regional extremes

Extreme climate events such as heat waves and cold spells can have long-lasting impacts on society or ecosystems (IPCC SREX, 2012) and there have been analyses of the impact of heat waves during recent centuries in Europe (Le Roy Ladurie, 2004, 2006; Barriendos and Rodrigo, 2006; Camuffo et al., 2010). The development of such events spans days to a few weeks, so that they are largely intra-seasonal by nature (Seneviratne et al., 2012). In such a context, the generally linear relationship between palaeo-climate reconstructions and actual climate can be strongly distorted. Since extreme events are by definition rare, large numbers of examples are required to get good statistics. Simulations of the past millennium offer a promising tool to investigate modelled extremes since they sample a longer time series and bigger range of possible cases than in most other simulations. The strongest limitation for an application of this method to palaeoclimatic data has been the necessity of dealing with daily data in order to resolve extreme value distributions (which may be non-Gaussian) and the need for palaeo-archives that record extreme variables (e.g. Donnelly and Woodruff, 2007). However, if we can demonstrate the robustness of the relationships between short and longer-term statistics over long periods of time, and/or their dependence on external forcings, we can potentially constrain the behaviour of temperature extremes in the future.

The statistical analyses of (daily) temperature hot extremes of the 20th century have shown that temperature is generally a bounded variable, for which the upper bound can be computed from the statistical parameters of extremes (Parey et al., 2010a, b). Diagnostic studies focusing on the probability distribution of temperature and precipitation extremes are often based on the application of Extreme Value Theory (EVT), though simpler metrics have also been used (e.g. Hansen et al., 2010). EVT describes the behaviour of the probability distribution near the tails, and allows one to estimate return periods for extremes that are longer than the period of observation (Coles, 2001). It has been applied to meteorological observations (Parey et al., 2010a), reanalysis data (Nogaj et al., 2006) and model simulations (Kharin et al., 2005, 2007) in order to estimate trends in extremes.

Extremes of hot and cold temperatures are correlated with mean temperatures over the northern extra-tropics (Yiou et al., 2009). Since very few models have archived daily outputs of temperature or precipitation on multi-century timescales, there has been no assessment of whether this is true over the longer term (Jansen et al., 2007). However, daily resolution



**Fig. 9.** Illustration of quantile regressions between the percentage of summer hot days (i.e. exceeding the 90th quantile of daily mean temperature in June–July–August) in western Europe ( $10^{\circ}$  W– $30^{\circ}$  E;  $36$ – $61^{\circ}$  N) and the precipitation frequency anomaly with respect to the 1948–2011 mean in winter–spring (January to May). The precipitation frequency is computed over southern Europe ( $10^{\circ}$  W– $30^{\circ}$  E;  $36$ – $46^{\circ}$  N) and is defined as the percentage of days with precipitation exceeding 0.5 mm. The quantile regressions are computed for the 10th and 90th quantiles of the hot day frequency, following Quesada et al. (2012). The red lines are for the 90th quantile regressions and the blue lines are for the 10th quantile regression. (a) shows the quantile regression for western Europe from the updated EOBS data set (Haylock et al., 2008) between 1950 and 2012 where each point represents a year. (b) is for the “historical” simulation (1960–2008) of the IPSL-CM5A-MR model (Dufresne et al., 2013). Both panels show a widening of the quantile regression for low values of precipitation frequency, indicating a consistency of the model simulation with observations.

data was requested for simulations in the CMIP5 archive (Yiou et al., 2012).

Summer heat waves are generally preceded by droughts in the winter and spring in extratropical regions (Fischer et al., 2007; Vautard et al., 2007). The mechanism involves a positive feedback between sensible heat fluxes, evapotranspiration and temperature (Schär et al., 1999), and this has also been found in global and regional simulations of the future (Seneviratne et al., 2006, 2010; Quesada et al., 2012). Quantile regression provides a useful statistic metric to investigate the linkage between precipitation in winter and spring and summer temperature. Ordinary least-squares regression focuses on the mean values of related variables, but by setting a threshold based on the upper/lower quantiles of the variable to be predicted, regression coefficients related to the high (or low) values of this variable are obtained (Koenker, 2005). The purpose of quantile regression is to investigate the conditional dependence between variables: for instance, the dependence structure could be different for small and large predictors. Hence, differences of slopes for small and high quantiles show that the relation between the predictand and predictor depends on the value of the predictor. An interesting feature is that quantile regression is not very sensitive to outliers, because the regression is performed on the ranks rather than the values themselves. We illustrate this diagnostic in Fig. 9, by computing the quantile regression for 90th and 10th deciles of the summer hot day frequency and winter–spring precipitation frequency anomaly in the IPSL-CM5A-MR historical

simulation and the E-OBS gridded data set (Haylock et al., 2008).

As in Quesada et al. (2012), the frequency of hot days is defined by the percentage of days in western Europe ( $10^{\circ}$  W– $30^{\circ}$  E;  $36$ – $61^{\circ}$  N) between June and August whose temperature anomaly exceeds the 90th quantile over a reference period (1948–2011). The frequency of rainy days is the percentage of days in southwestern Europe ( $10^{\circ}$  W– $30^{\circ}$  E;  $36$ – $46^{\circ}$  N) between January and May whose precipitation exceeds 0.5 mm. This is a simplistic index for soil moisture (or drought) but it does have a significant predictive skill to European summer temperature variations (Vautard et al., 2007). More sophisticated indices of drought or soil moisture marginally improve the predictive skill (Seneviratne and Koster, 2012). In Fig. 9 the quantile regression slopes illustrate the asymmetry of the precipitation or temperature dependence for hot or cool summers in western Europe (Hirschi et al., 2011; Mueller and Seneviratne, 2012; Quesada et al., 2012; Seneviratne and Koster, 2012).

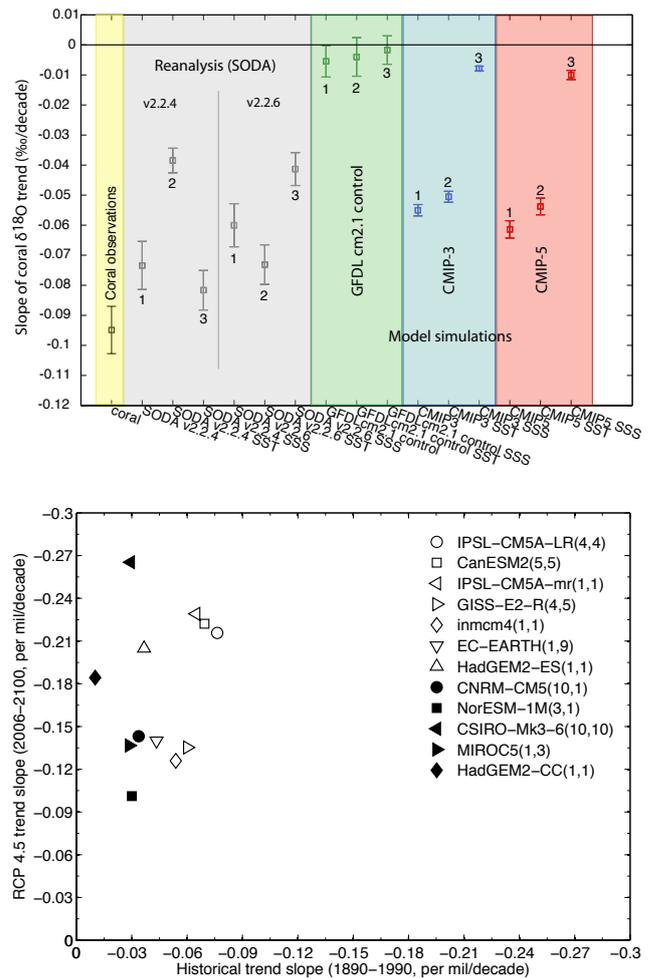
The 90th and 10th quantile regression lines are not parallel. Thus while the general picture is that a dry winter/spring tends to favor a hot summer and wet winter–spring conditions are generally followed by cool summers, dry winter–spring conditions can be followed by cool summers as well as heat waves (large spread between low and high quantiles). This is due to the fact that the genesis of heat waves can be broken in just a few days, due to fast variations of the synoptic atmospheric circulation (Hirschi et al., 2011; Quesada

et al., 2012). This feature has been tested on CMIP3 and some CMIP5 simulations for the present and future scenarios and shows that the seasonal predictability of large European hot summers decreases under drier conditions in southern Europe, although their frequency increases (Quesada et al., 2012). By looking at the last millennium simulations we will be able to examine the stability in time of these patterns, and hence potentially constrain future changes.

### 5.2 20th-century changes in tropical Pacific climate

The response of the tropical Pacific Ocean to anthropogenic climate change is uncertain, partly because we do not have a good understanding of how the region has responded to drivers in the past. Instrumentally based estimates of SST over the 20th century are not internally consistent (Deser et al., 2010), and model simulations have a wide spread of 20th-century trends (Thompson et al., 2011). Trends in the tropical Pacific are particularly challenging because the instrumental record is sparse even for the early 20th century and long-term in situ measurements of SST are uncommon. High-resolution palaeo-climate records, particularly the large network of tropical Pacific coral  $\delta^{18}\text{O}_{\text{calcite}}$  records, can be used to extend the observational record and assess long-term trends. These  $\delta^{18}\text{O}_{\text{calcite}}$  records respond to the combined effects of SST and the isotopic composition of seawater ( $\delta^{18}\text{O}_{\text{sw}}$ ) (which is strongly correlated to sea surface salinity, SSS) and can reveal changes on longer timescales.

To address the limitations of historical observations, model simulations and coral proxy records in the tropical Pacific, Thompson et al. (2011) used a forward-modelling approach to generate synthetic coral records (i.e. pseudo-corals) from observational and climate model output and test whether these pseudo-corals are in agreement with the network of coral  $\delta^{18}\text{O}_{\text{c}}$  observations. The forward model for  $\delta^{18}\text{O}_{\text{c}}$  calculates isotopic variations as a function of SST and SSS anomalies, with an SST- $\delta^{18}\text{O}_{\text{c}}$  slope of  $-0.22\text{‰}\text{°C}^{-1}$  and the SSS- $\delta^{18}\text{O}_{\text{sw}}$  slope varying by region (LeGrande and Schmidt, 2006). When driven with historical SST and SSS data, the simple model of  $\delta^{18}\text{O}_{\text{c}}$  is able to capture the spatial and temporal pattern of ENSO and the linear trend observed in 23 Indo-Pacific coral records between 1958 and 1990 (Thompson et al., 2011). The observed trends were driven primarily by warming at the coral sites, though SSS was responsible for approximately 40% of the shared  $\delta^{18}\text{O}_{\text{c}}$  trend. However, pseudo-coral records calculated from CMIP3 and CMIP5 historical simulations could not reproduce the magnitude of the secular trend (Fig. 10, upper panel), the change in mean state, or the change in ENSO-related variance observed in the coral network from 1890 to 1990. While the observational coral network suggests a reduction in ENSO-related variance and an El Niño-like trend over the 20th century, CMIP3 and CMIP5 simulations vary greatly on both points.



**Fig. 10.** Upper panel: magnitude of the trend in  $\delta^{18}\text{O}_{\text{c}}$  (‰/decade, computed from a simple linear regression through the trend PC) in corals (far left), Simple Ocean Data Assimilation (SODA) 20th-century reanalysis (Carton and Giese, 2008; Giese and Ray, 2011; Compo et al., 2011), a 500 yr control run from GFDL CM2.1 (Wittenberg, 2009), and the CMIP3 and CMIP5 multi-model ensembles. In each case,  $\delta^{18}\text{O}_{\text{c}}$  was modelled from SST and SSS (1), SST only (2), and SSS (3). Lower Panel: magnitude of the  $\delta^{18}\text{O}_{\text{c}}$  trend (‰/decade, computed from a simple linear regression through the trend PC) over 1890–1990 in pseudocorals modelled from CMIP5 historical simulations and over 2006–2100 in the RCP 4.5 projections where numbers in parenthesis indicate the number of runs in the historical and RCP 4.5 ensemble, respectively.

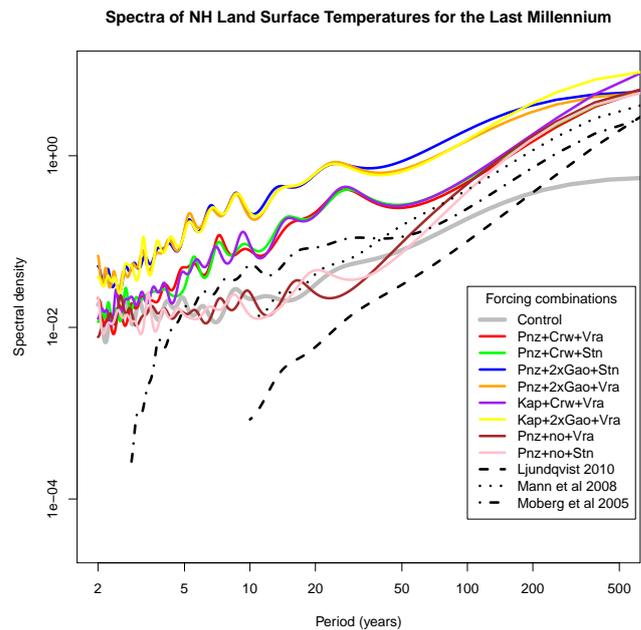
The differences between observed and GCM-derived  $\delta^{18}\text{O}_{\text{c}}$  trends may stem from the simplicity of the forward model for  $\delta^{18}\text{O}_{\text{c}}$ , bias in the coral records, and/or errors in the GCM SST and SSS responses, or indicate an important role for unforced variability. Isotope-enabled coupled control simulations highlight uncertainties in the SSS- $\delta^{18}\text{O}_{\text{sw}}$  relationship and suggest that short-term isotope variability may play a minor role (Russon et al., 2013; Thompson et al., 2013). Previous work has also highlighted potential

biases in simulated salinity fields as a source of the discrepancy (Thompson et al., 2011; 2013). For example, CMIP3 and CMIP5 simulations display weak and spatially heterogeneous SSS trends, such that the magnitude of the  $\delta^{18}\text{O}_c$  trend in pseudo-corals simulated from CMIP3 and CMIP5 SSS is indistinguishable from the trends observed in individual centuries of an unforced control run (Fig. 10, upper panel). Further, trends in mean state and ENSO-related variance within the basin are highly variable among the CMIP5 models, and even between ensemble members of the same model, and much of this model spread may be attributed to differences in the simulated SSS fields. On the other hand, while pseudo-corals, modelled from the new SODA 20th-century reanalysis of SST and SSS, display greater agreement with the observed coral trends, two recent versions of this product disagree regarding the relative contribution of SST and SSS. These results suggest that more work is needed to constrain the magnitude of the observed 20th-century salinity trend throughout the tropical Pacific Ocean. This work provides an example of the utility of forward models in investigating potential biases in both the models and proxy data, which may be used for further model development and exploration and improvement of model metrics.

Despite the disagreement among models and runs regarding the change over the 20th century, the CMIP5 projections converge upon a more El Niño-like (e.g. warmer eastern equatorial Pacific) mean state change by 2100 under RCP 4.5 (with only one model suggesting the opposite), consistent with the CMIP3 projections (Meehl et al., 2007). However, the models still disagree about the change in ENSO-related variance. Further, there is no clear relationship between the magnitude of the simulated 20th-century  $\delta^{18}\text{O}_c$  trend and the projected future  $\delta^{18}\text{O}_c$  trend in the CMIP5 ensemble (Fig. 10, lower panel). This suggests that an agreement of the simulated 20th-century change in the tropical Pacific with that of the observational coral network would not be a reliable indicator of future trends. Nonetheless, this work highlights key uncertainties in the observed and simulated salinity trends within the basin and thus provides a basis for further development of the models and this potential metric. More generally, it shows the utility of a forward modelling approach in palaeo-model/data comparisons to highlight key functional dependencies in specific proxies and investigate potential biases in both models and reconstructions.

### 5.3 Decadal to multi-decadal variability

In contrast to the spatial domain used in other examples here, this section highlights two analyses in the frequency domain that illustrate the important role of relatively uncertain forcings in assessing skill in model simulations of decadal to multi-decadal variability. Given the short instrumental period, it might be hoped that longer time series from proxy reconstructions for the last millennium could be used to



**Fig. 11.** Spectra from an ensemble of LM simulations using the same model but driven with different sets of forcings compared with Ljungqvist (2010), Mann et al. (2008) and Moberg et al. (2005) reconstructions. The clustering of simulations is driven entirely by changes in the volcanic forcing data set used, with the simulations with the most decadal and multi-decadal variability using the Gao et al. (2008) reconstruction. Only in the examples where no volcanic forcing is used at all is the impact of different solar forcing reconstructions detectable. Spectra derived using MEM with 30 poles, from 850 to 2005, after correction for control run drift using a loess low-frequency estimate derived from the control run. Key abbreviations: Land use: Pnz (Pongratz et al., 2008), Kap (Kaplan et al., 2011); Solar: Vra (Vieira et al., 2011), Stn (Steinilber et al., 2009); Volcanic: 2xGao (twice the forcing from Gao et al., 2008), Crw (Crowley and Unterman, 2013).

constrain internal variability, and hence the unforced spread in projections over the next few decades.

In Fig. 11, we show the maximum-entropy method (MEM) spectra (using 30 poles) for the NH mean land surface temperature over 8 last millennium simulations (850–2005) with the GISS-E2-R model that were run with different combinations of plausible solar, volcanic and land use forcings (Schmidt et al., 2011, 2012). The spectra are similar for models that have the same volcanic forcing, and significantly different when the volcanic forcing is derived from a different data set or where no volcanic forcing was imposed at all. Specifically, interannual to multi-decadal variability is much larger when volcanoes are imposed, and the larger the volcanic forcing, the greater the variability, with the largest response in simulations using the Gao et al. (2008) reconstruction (Gao), compared to the Crowley and Unterman (2013) reconstruction (Crw). Note that the implementation of the Gao et al. volcanic forcing in these

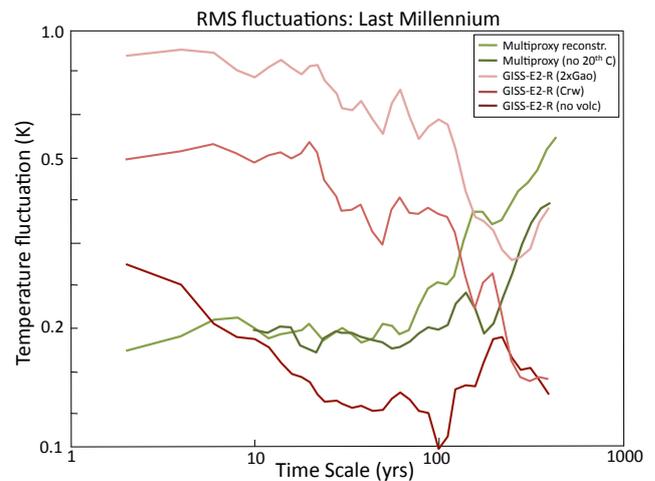
simulations was misspecified and gave roughly twice the expected radiative forcing. However, given the uncertainties in specifying volcanic forcing (for instance, associated with the effective radius of the particles), the exercise is nonetheless useful in highlighting the role of forcings in determining variance. In contrast, the difference between two different solar forcings (Vieira et al., 2011; Steinhilber et al., 2009) is not detectable in this metric.

The no-volcano simulations underestimate the decadal/multi-decadal variance seen in two of the three reconstructions, while the with-volcano simulations overestimate it. The lowest-frequency bands in the models (primarily driven by orbital forcing, and the 20th century anthropogenic trend) have slightly larger variance than in the reconstructions.

Another analysis of variability as a function of timescale is one focused on power law scaling (Lovejoy and Schertzer, 1986). Several scaling studies of GCMs demonstrate that they generally simulate the statistics (including spectral scaling exponents) reasonably well up to  $\approx 10$  yr scales (e.g. Fraedrich and Blender, 2003; Zhu et al., 2006; Rybski et al., 2008; Lovejoy and Schertzer, 2013; Vyushin et al., 2012). However, tests at lower frequencies are strongly affected by the solar and volcanic forcings as well as the possible impacts of slow processes such as deep ocean or land-ice dynamics which are perhaps poorly represented or missing.

Following Lovejoy and Schertzer (2012a), we calculate the root mean square (rms) fluctuation as a function of timescale, from months to centuries, for the NH land temperatures using the same eight runs of the GISS-E2-R model used above, for the period 1500–1900 CE (Fig. 12). Since simulations are strongly clustered according to changes in the volcanic forcing used (Fig. 11), for simplicity we averaged over the three Gao and three Crw volcanic and the two no-volcanic runs. For comparison, we show the mean of the same metric from three multiproxy reconstructions (Huang et al., 2000; Moberg et al., 2005; Ljungqvist, 2010). The multiproxy average is processed with and without the 20th century to indicate the importance of that period for the scaling behaviour – in all cases the variance in the multi-decadal to century scale is greatly enhanced by the recent anthropogenic trend. These curves show fluctuations stable with scale over the low frequency weather regime (years to decades) but increasing in the climate regime (decades to centuries) (Fig. 12).

The comparison with the GISS-E2-R simulations is illuminating. First, we note that the slopes for the simulations show decreasing variance from annual to centennial scale, in contrast to the reconstructions. Only the volcano-free runs (bottom) clear have increasing variance with scale in the centennial and longer periods, though with a magnitude of variance at all scales that is too low. Volcanic forcings add variance at all scales, but producing larger magnitudes than inferred from the reconstructions.



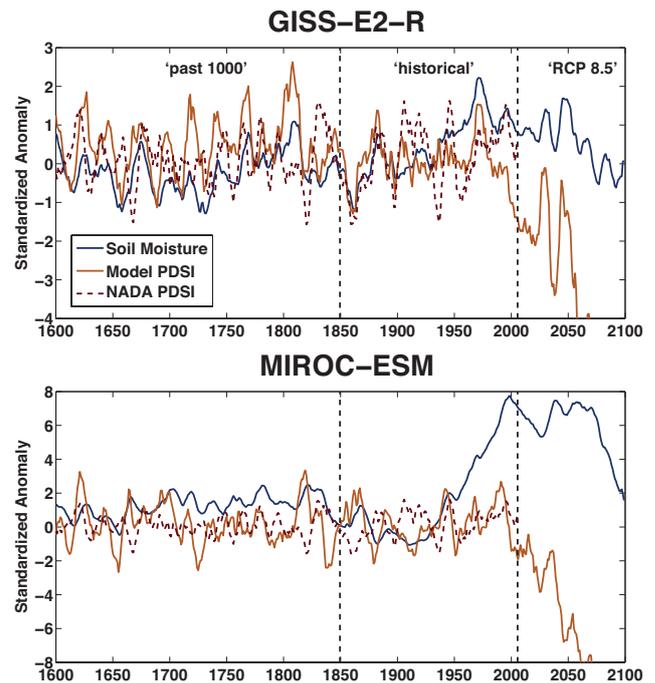
**Fig. 12.** rms fluctuations of instrumental and palaeo-climate reconstructions compared to drift-corrected simulations of the Northern Hemisphere land temperature for the period 1500–1900. 2xGao and Crw refer to GISS-E2-R simulations using the (2x) Gao et al. (2008) and Crowley and Unterman (2013) reconstructions of volcanic forcing. The multiproxy reconstruction used is an average of three NH estimates, and the rms fluctuations are separately shown for the periods 1000–1900 and 1000–1980.

Both these results demonstrate clear mismatches in behaviour between the models' simulated variance at different scales and the inferred variability from multi-proxy reconstructions. However, there are strong sensitivities to the (uncertain) external forcing functions, precluding a straightforward attribution of the mismatch to potentially misspecified forcings, missing mechanisms, insufficient "slow" variability or problems in the reconstructions. Specifically, reconstructions may have frequency-dependent biases that vary depending on the methodology and source data. For instance, boreholes used in Huang et al. (2000) do not have high frequency variability, while low frequency variability in tree-ring-based reconstructions is hard to capture. In models, the importance of decadal and multi-decadal variance in the Pacific or Atlantic sectors vary widely and are poorly constrained from observations, and there may be significant issues with the forcing functions themselves. Other analyses (i.e. Schurer et al., 2013) have examined the coherence of last millennium simulations and the proxy reconstructions and found that while signatures of multiple forcings can be determined, there is a mismatch in the magnitude of the response to volcanoes coherent with the conclusions drawn above. It therefore remains unclear what implications these tests have for future projections, but improvements in the forcing data sets or a focus on more specific comparisons may prove fruitful in future analyses.

#### 5.4 Non-stationarity in hydroclimate diagnostics

Hydroclimate variability can be quantified using a range of variables, including precipitation, soil moisture, lake levels, or other synthetic indices (e.g. Nigam and Ruiz-Barradas, 2006). Most models provide output for these diagnostics, but often these variables are not directly derivable from palaeo-climate archives, creating a challenge when conducting model-data comparisons. However, calibrations of networks of precipitation-sensitive tree ring widths have been used to reconstruct the Palmer Drought Severity Index (PDSI) in North America and Asia over the Common Era (Cook et al., 2004, 2010). PDSI is calculated using temperature-derived estimates of the evapo-transpiration and precipitation, and nominally represents a normalised index of soil moisture, with negative values indicating drought and positive values indicate wetter than normal conditions. There are many outstanding issues with using variations of the index globally to assess drought, in definition and availability and quality of inputs and sensitivity (e.g. contrast Sheffield et al., 2012 and Dai, 2013). However, we focus here on the question of how well this index, if derived from GCM output, reflects simulated soil moisture and whether this relationship changes over time.

From two GCMs (GISS-E2-R and MIROC-ESM), we calculated PDSI using simulated temperature and precipitation from the GISS-E2-R and MIROC-ESM models using the Thornthwaite method. We compared this index against the standardised (zero mean, unit standard deviation, over the 1850–1950 period, 10 yr smoothing) total column soil moisture model output for the Central Plains of North America (105–90° W; 32–48° W) (Fig. 13). Prior to the start of the industrial period in 1850, PDSI and soil moisture track each other closely in both models (GISS-E2-R:  $r = 0.82$ ; MIROC-ESM:  $r = 0.50$ ). Beginning near the middle of the twentieth century, however, the two indices diverge dramatically. In one model (GISS-E2-R) the correlation weakens considerably ( $r = 0.33$ ), while in the other model (MIROC-ESM) the sign of the correlation reverses ( $r = -0.29$ ). The PDSI changes over the 21st century would suggest severe and unprecedented drought. In contrast, the simulated soil moisture trends indicate a more modest shift towards drying (GISS-E2-R) or even wetter conditions over the coming decades (MIROC-ESM). The divergence in projections is related to the treatment of evapo-transpiration (ET) in the model versus in the PDSI (Thornthwaite) calculation. In this PDSI calculation, temperature is used as a proxy for the energy available while in the GCMs the soil energy and moisture budgets are calculated directly using explicit physical models. In reality, ET becomes increasingly decoupled from temperature as the temperature increases, a factor reflected in the model soil moisture but not in the PDSI index. For time periods with strong transient changes in temperature (e.g. the late 20th century and into the future), our analysis suggests that the



**Fig. 13.** Standardised anomalies for PDSI and soil moisture in two models (GISS-E2-R and MIROC-ESM) using a past1000 simulation, and a historical+rcp85 continuation. For reference, the tree-ring-based reconstruction is plotted (dashed line) (Cook et al., 2010), though this would not be expected to line up exactly with the model simulations. All data smoothed with a 10 yr running mean.

usefulness of PDSI for projecting drought and hydroclimate trends is limited.

While this example of a diagnostic divergence is specific to the PDSI and soil moisture, there are wider implications that might need to be explored in other metrics. It is not unusual for a proxy to have a non-stationary response to a climate variable of interest (see Sect. 5.2 for another example), and it is incumbent on the investigator to ensure that any consequences of this are fully explored.

## 6 Conclusions and recommendations

In this paper, we have focused on the opportunities provided by “out-of-sample” palaeo-climate experiments within the CMIP5 framework, and specifically how measures of skill in modelling palaeo-climate change might inform future projections of climate change.

We have given examples that show that some relationships are robust across the multi-model ensemble, over multiple simulations and in the palaeo-data (Sect. 3) and examples of skill measures that are well correlated with the simulated magnitude of future change, thus allowing the likely magnitude of future changes to be constrained (Sect. 4). We also give examples of cases (Sect. 5) where there is a need for caution because of the limitations with models,

the experimental setup used in CMIP5, or with the palaeo-climate data itself.

Our examples illustrate the general requirements for attempts to use the palaeo-climate simulations to quantitatively constrain future projections. Each example makes use of a specific target (or targets) from a palaeo-climate reconstruction of change that is within the scope of the modelled system, defines a metric of skill that quantifies the accuracy of the modelled changes and assesses the connection to a future prediction. The successes and problems discussed above lead naturally to a set of guidelines that could profitably guide future research for both modellers and the palaeo-data community:

1. Palaeo-simulations need to be performed with models that are also being used for future projections and produce model diagnostics that are commensurate in all experiments (as in CMIP5).
2. The more extensive the structural uncertainty examined (across models, boundary conditions etc.) the more robust any resulting constraints will be. Some of our analyses (i.e. Sects. 4.1 and 4.2) are limited by the small number of palaeo-simulations currently available in the CMIP5 database, and we hope that the demonstration of their potential to address questions relevant to the future will encourage other modelling groups to complete and archive these simulations.
3. Palaeo-data targets should be spatially representative synthesis products with well-characterised uncertainties. Our analyses rely heavily on the use of synthesis data products, for instance the MARGO data set for the LGM (MARGO, 2009), pollen-based reconstructions for the mid-Holocene (Bartlein et al., 2011), multi-proxy reconstructions of hemispheric temperature (e.g. Moberg et al., 2005), or gridded tree-ring-based reconstructions of PDSI for the last millennium (Cook et al., 2010). Such products are invaluable, but there is a need for increased transparency of included uncertainties and continued expansion of targets (e.g. see Müller et al., 2011 for sea ice extent). Increasing model complexity and scope, for instance by including a carbon cycle, fire models or online tracers such as water isotopes, necessitates the creation of new synthesis products (e.g. charcoal records: Daniou et al., 2012; or sea surface carbonate isotopes: Oppo et al., 2007) if useful comparisons are to be made. Examples in Sects. 4.3 and 5.1 illustrate the need for more efforts in this direction.
4. Skill metrics may be impacted by uncertainties in external forcing (and thus not solely characterise the realism of modelled processes; as in the spectra generated in Sect. 5.3), or may have non-stationary relationships with impacts of interest (Sect. 5.4). Improved forward

modelling of palaeo-data (as in Sect. 5.2) will be increasingly important.

5. Relationships between targets in the past and the future predictions should be examined and not assumed. Not all mismatches of palaeo-models and reconstructions are related to factors important for future sensitivities, and not all divergences in future projections are correlated to differences in palaeo-climate skill.

The periods and hypotheses tested using palaeo-climate simulations are far more limited than the number of interesting features in the palaeo-climate record. The three periods selected for CMIP5 were chosen on the basis of their relative maturity (the existence of prior sets of experiments, already tested issues, existing data syntheses), but additional periods are also potentially useful – the mid-Pliocene (2.5 million years ago), the transient 8.2 ka event, the last interglacial, the peak Eocene, etc. (see Schmidt, 2010 for justifications). Some of these periods are already being examined in a coordinated fashion (e.g. Haywood et al., 2013 and Dolan et al., 2012 for the Pliocene), and we hope that more coordinated experiments will be started. Further expansions of the model experiments will lead to increases in the production of higher frequency diagnostics (daily and sub-daily variations), and include perturbed physics ensembles to better characterise the model structural uncertainty. On the data side, much greater efforts to create palaeo-data synthesis products with robust uncertainty estimates are possible. All of these expansions will create possibilities for more and better tests of model performance and hence potentially lead to better constraints on future projections. In the meantime, there is still a huge untapped scope for more informative palaeo-model comparisons that can be made using the existing databases.

*Acknowledgements.* This paper arose from a workshop at the Bishop Museum, Honolulu in March 2012 organised by the PAGES-CLIVAR Intersection Panel. Funding from the NASA Modeling and Analysis Program, NOAA, NSF, AORI, PAGES and CLIVAR is gratefully acknowledged. We acknowledge the WCRP WGCM, which is responsible for CMIP, and we thank the climate modelling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Individual researchers were funded from the French ANR ISOTROPIC and CHEDAR projects, NSF grant ATM-0902133, the Japanese Ministry of Environment S-5 fund, NOAA grant NA10OAR4310115, and the Canadian Sea Ice and Snow Evolution (CanSISE) Network. We thank Rumi Ohgaito and Tetsuo Sueyoshi (JAMSTEC) for help in some of the analysis, Richard Healy for model support, and Françoise Vimeux and Sandrine Bony for discussions. Comments from Tamsin Edwards and an anonymous reviewer greatly improved the clarity of the paper.

All authors contributed to the conception and scope of the paper. Sections 1, 2 and 6 were jointly written by all authors. Principle contributions for each technical section are as follows: Sect. 3.1: M. Kageyama; Sect. 3.2: M. Kageyama, S. P. Harrison; Sect. 4.1: C. Risi; Sect. 4.2: J. D. Annan, J. C. Hargreaves; Sect. 4.3: L.-B. Tremblay; Sect. 5.1: P. Yiou; Sect. 5.2: D. Thompson; Sect. 5.3: A. N. LeGrande, G. A. Schmidt, S. Lovejoy; Sect. 5.4: B. I. Cook.

Edited by: M. Crucifix

## References

- Abe, M., Shiogama, H., Nozawa, T., and Emori, S.: Estimation of future surface temperature changes constrained using the future-present correlated modes in inter-model variability of CMIP3 multimodel simulations, *J. Geophys. Res.*, 116, D18104, doi:10.1029/2010JD015111, 2011.
- Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models, *Geophys. Res. Lett.*, 39, L09712, doi:10.1029/2012GL051607, 2012.
- Annan, J. D. and Hargreaves, J. C.: Understanding the CMIP3 multimodel ensemble, *J. Climate*, 24, 4529–4538, doi:10.1175/2011JCLI3873.1, 2011.
- Annan, J. D. and Hargreaves, J. C.: A new global reconstruction of temperature changes at the Last Glacial Maximum, *Clim. Past*, 9, 367–376, doi:10.5194/cp-9-367-2013, 2013.
- Bala, G., Caldeira, K., and Nemani, R.: Fast versus slow response in climate change: implications for the global hydrological cycle, *Clim. Dynam.*, 35, 423–434, doi:10.1007/s00382-009-0583-y, 2010.
- Barriendos, M. and Rodrigo, F.: Study of historical flood events on Spanish rivers using documentary data, *Hydrolog. Sci. J.*, 51, 765–783, 2006.
- Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis, B. A. S., Gajewski, K., Guiot, J., Harrison-Prentice, T. I., Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppä, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A. E., Williams, J., and Wu, H. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, *Clim. Dynam.*, 37, 775–802, 2011.
- Boé, J. L., Hall, A., and Qu, X.: September sea-ice cover in the Arctic Ocean projected to vanish by 2100, *Nat. Geosci.*, 2, 1–3, 2009.
- Bonfils, C., de Noblet-Ducoudré, N., Guiot, J., and Bartlein, P.: Some mechanisms of mid-Holocene climate change in Europe, inferred from comparing PMIP models to data, *Clim. Dynam.*, 23, 79–98, 2004.
- Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., and Denvil, S.: Robust direct effect of carbon dioxide on tropical circulation and regional precipitation, *Nat. Geosci.*, 6, 447–451, 2013.
- Braconnot, P., Otto-Bliesner, B., Harrison, S., Joussaume, S., Peterchmitt, J.-Y., Abe-Ouchi, A., Crucifix, M., Driesschaert, E., Fichefet, Th., Hewitt, C. D., Kageyama, M., Kitoh, A., Laîné, A., Loutre, M.-F., Marti, O., Merkel, U., Ramstein, G., Valdes, P., Weber, S. L., Yu, Y., and Zhao, Y.: Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features, *Clim. Past*, 3, 261–277, doi:10.5194/cp-3-261-2007, 2007.
- Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, *Nat. Clim. Change*, 2, 417–424, doi:10.1038/nclimate1456, 2012.
- Brewer, S., Guiot, J., and Torre, F.: Mid-Holocene climate change in Europe: a data-model comparison, *Clim. Past*, 3, 499–512, doi:10.5194/cp-3-499-2007, 2007.
- Brient, F. and Bony, S.: How may low-cloud radiative properties simulated in the current climate influence low-cloud feedbacks under global warming?, *Geophys. Res. Lett.*, 39, L20807, doi:10.1029/2012GL053265, 2012.
- Camuffo, D., Bertolin, C., Diodato, N., Barriendos, M., Dominguez-Castro, F., Cocheo, C., della Valle, A., Garnier, E., and Alcoforado, M. J.: The western Mediterranean climate: how will it respond to global warming?, *Climatic Change*, 100, 137–142, 2010.
- Carton, J. A. and Giese, B. S.: A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA), *Mon. Weather Rev.*, 136, 2999–3017, 2008.
- COHMAP Members: Climatic changes of the last 18,000 years: observations and model simulations, *Science*, 241, 1043–1052, 1988.
- Coles, S.: An introduction to statistical modeling of extreme values, Springer, London, New York, 2001.
- Collins, M., Chandler, R. E., Cox, P. M., Huthnance, J. M., Rougier, J., and Stephenson, D. B.: Quantifying future climate change, *Nat. Clim. Change*, 2, 403–409, 2012.
- Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, *Q. J. Roy. Meteorol. Soc.*, 137, 1–28, 2011.
- Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M., and Stahle, D. W.: Long-Term Aridity Changes in the Western United States, *Science*, 306, 1015–1018, 2004.
- Cook, E. R., Seager, R., Heim Jr., R. R., Vose, R. S., Herweijer, C., and Woodhouse, C.: Megadroughts in North America: placing IPCC projections of hydroclimatic change in a long-term palaeo-climate context, *J. Quaternary Sci.*, 25, 48–61, 2010.
- Crowley, T. J. and Unterman, M. B.: Technical details concerning development of a 1200 yr proxy index for global volcanism, *Earth Syst. Sci. Data*, 5, 187–197, doi:10.5194/essd-5-187-2013, 2013.
- Crucifix, M.: Does the Last Glacial Maximum constrain climate sensitivity?, *Geophys. Res. Lett.*, 33, L18701, doi:10.1029/2006GL027137, 2006.
- Dai, A.: Increasing drought under global warming in observations and models, *Nat. Clim. Change*, 3, 52–58, doi:10.1038/nclimate1633, 2013.
- Daniau, A.-L., Bartlein, P. J., Harrison, S. P., Prentice, I. C., Brewer, S., Friedlingstein, P., Harrison-Prentice, T. I., Inoue, J., Izumi, K., Marlon, J. R., Mooney, S., Power, M. J., Stevenson, J., Tinner, W., Andrić, M., Atanassova, J., Behling, H., Black, M., Blarquez, O., Brown, K. J., Carcaillet, C., Colhoun, E. A., Colombaroli, D., Davis, B. A. S., D’Costa, D., Dodson, J., Dupont, L.,

- Eshetu, Z., Gavin, D. G., Genries, A., Haberle, S., Hallett, D. J., Hope, G., Horn, S. P., Kassa, T. G., Katamura, F., Kennedy, L. M., Kershaw, P., Krivonogov, S., Long, C., Magri, D., Marinova, E., McKenzie, G. M., Moreno, P. I., Moss, P., Neumann, F. H., Norström, E., Paitre, C., Rius, D., Roberts, N., Robinson, G. S., Sasaki, N., Scott, L., Takahara, H., Terwilliger, V., Thevenon, F., Turner, R., Valsecchi, V. G., Vanni re, B., Walsh, M., Williams, N., and Zhang, Y.: Predictability of biomass burning in response to climate changes, *Global Biogeochem. Cy.*, 26, GB4007, doi:10.1029/2011GB004249, 2012.
- Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P. L., Wofsy, S. C., and Zhang, X.: Couplings between changes in the climate system and biogeochemistry, in: *IPCC, Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, UK and USA, 499–587, 2007.
- Deser, C., Phillips, A. S., and Alexander, M. A.: Twentieth century tropical sea surface temperature trends revisited, *Geophys. Res. Lett.*, 37, L10701, doi:10.1029/2010GL043321, 2010.
- de Vernal, A., Hillaire-Marcel, C., and Darby, D. A.: Variability of sea ice cover in the Chukchi Sea (western Arctic Ocean) during the Holocene, *Paleoceanography*, 20, PA4018, doi:10.1029/2005PA001157, 2005.
- de Vernal, A., Hillaire-Marcel, C., Solignac, S., Radi, T., and Rochon, A.: Reconstructing seaice conditions in the Arctic and Subarctic prior to human observations, in: *Arctic Sea Ice Decline: Observations, Projections, Mechanisms and Implications*, edited by: de Weaver, E., AGU Monograph Series, 180, 27–45, 2008.
- Dolan, A. M., Koenig, S. J., Hill, D. J., Haywood, A. M., and DeConto, R. M.: Pliocene Ice Sheet Modelling Intercomparison Project (PLISMIP) – experimental design, *Geosci. Model Dev.*, 5, 963–974, doi:10.5194/gmd-5-963-2012, 2012.
- Donnelly, J. P. and Woodruff, J. D.: Intense hurricane activity over the past 5,000 years controlled by El Ni o and the West African monsoon, *Nature*, 447, 465–468, 2007.
- Drost, F., Karoly, D., and Braganza, K.: Communicating global climate change using simple indices: an update, *Clim. Dynam.*, 39, 989–999, doi:10.1007/s00382-011-1227-6, 2012.
- Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F. F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Eth , C., Fairhead, L., Fichet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Jous-saume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lef vre, M.-P., Lef vre, F., L vy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., and Viovy, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, *Clim. Dynam.*, 40, 2123–2165, doi:10.1007/s00382-012-1636-1, 2013.
- Dyke, A. S. and Savelle, J. M.: Holocene history of the Bering Sea bowhead whale (*Balaena mysticetus*) in its Beaufort Sea summer grounds off southwestern Victoria Island, Western Canadian Arctic, *Quatern. Res.*, 55, 371–379, 2001.
- Edwards, T. L., Crucifix, M., and Harrison, S. P.: Using the past to constrain the future: how the palaeorecord can improve estimates of global warming, *Prog. Phys. Geogr.*, 31, 481–500, doi:10.1177/0309133307083295, 2007.
- Fasullo, J. T. and Trenberth, K. E.: A less cloudy future: The role of subtropical subsidence in climate sensitivity, *Science*, 338, 792–794, 2012.
- Fern andez-Donado, L., Gonz lez-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., Garc a-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P., and Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Clim. Past*, 9, 393–421, doi:10.5194/cp-9-393-2013, 2013.
- Fischer, E. M., Seneviratne, S. I., L thi, D., and Sch r, C.: Contribution of land-atmosphere coupling to recent European summer heat waves, *Geophys. Res. Lett.*, 34, L06707, doi:10.1029/2006GL029068, 2007.
- Fraedrich, K. and Blender, K.: Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.*, 90, 108501, doi:10.1103/PhysRevLett.90.108501, 2003.
- Funder, S., Goosse, H., Jepsen, H., Kaas, E., Kj r, K. H., Korsgaard, N. J., Larsen, N. K., Linderson, H., Lys , A., M ller, P., Olsen, J., and Willerslev, E.: A 10,000-year record of Arctic Ocean sea-ice variability: View from the beach, *Science*, 333, 747–750, doi:10.1126/science.1202760, 2011.
- Gao, C., Robock, A., and Ammann, C.: Volcanic forcing of climate over the last 1500 years: An improved ice-core based index for climate models, *J. Geophys. Res.*, 113, D2311, doi:10.1029/2008JD010239, 2008.
- Giese, B. S. and Ray, S.: El Ni o variability in simple ocean data assimilation (SODA), 1871–2008, *J. Geophys. Res.*, 116, C02024, doi:10.1029/2010JC006695, 2011.
- Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, *J. Geophys. Res.*, 113, D06104, doi:10.1029/2007JD008972, 2008.
- Guiot, J., Boreux, J. J., Braconnot, P., and Torre, F.: Data-model comparison using fuzzy logic in paleoclimatology, *Clim. Dynam.*, 15, 569–581, 1999.
- Hall, A. and Qu, X.: Using the current seasonal cycle to constrain snow albedo feedback in future climate change, *Geophys. Res. Lett.*, 33, L03502, doi:10.1029/2005GL025127, 2006.
- Hansen, J., Sato, M., Ruedy, R., Kharecha, P., Lacis, A., Miller, R. L., Nazarenko, L., Lo, K., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, S., Baum, E., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Cohen, A., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Jonas, J., Kelley, M., Kiang, N. Y., Koch, D., Labow, G., Lerner, J., Menon, S., Novakov, T., Oinas, V., Perlwitz, J., Perlwitz, J., Rind, D., Romanou, A., Schmunk, R., Shindell, D., Stone, P., Sun, S., Streets, D., Tausnev, N., Thresher, D., Unger, N., Yao, M., and Zhang, S.: Climate simulations for 1880–2003 with GISS modelE, *Clim. Dynam.*, 29, 661–696, doi:10.1007/s00382-007-0255-8, 2007.

- Hansen, J. E., Sato, M., and Ruedy, R.: Perception of climate change, *P. Natl. Acad. Sci.*, 109, 14726–14727, doi:10.1073/pnas.1205276109, 2012.
- Hargreaves, J. C., Abe-Ouchi, A., and Annan, J. D.: Linking glacial and future climates through an ensemble of GCM simulations, *Clim. Past*, 3, 77–87, doi:10.5194/cp-3-77-2007, 2007.
- Hargreaves, J. C., Paul, A., Ohgaito, R., Abe-Ouchi, A., and Annan, J. D.: Are paleoclimate model ensembles consistent with the MARGO data synthesis?, *Clim. Past*, 7, 917–933, doi:10.5194/cp-7-917-2011, 2011.
- Hargreaves, J. C., Annan, J. D., Yoshimori, M., and Abe-Ouchi, A.: Can the last glacial maximum constrain climate sensitivity?, *Geophys. Res. Lett.*, 39, L24702, doi:10.1029/2012GL053872, 2012.
- Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A.: Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene, *Clim. Past*, 9, 811–823, doi:10.5194/cp-9-811-2013, 2013.
- Harrison, S. P. and Bartlein, P. J.: 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*, edited by: Henderson-Sellers, A. and McGuffie, K., Elsevier, Amsterdam, the Netherlands, 403–436, 2012.
- Harrison, S. P., Bartlein, P. J., Brewer, S., Prentice, I. C., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Model benchmarking with glacial and mid-Holocene climates, *Clim. Dynam.*, doi:10.1007/s00382-013-1922-6, in press, 2013.
- Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, *B. Am. Meteorol. Soc.*, 90, 1095–1107, doi:10.1175/2009BAMS2607.1, 2009.
- Haylock, M. R., Hofstra, N., Tank, A. M. G. K., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, D20119, doi:10.1029/2008JD010201, 2008.
- Haywood, A. M., Hill, D. J., Dolan, A. M., Otto-Bliesner, B. L., Bragg, F., Chan, W.-L., Chandler, M. A., Contoux, C., Dowsett, H. J., Jost, A., Kamae, Y., Lohmann, G., Lunt, D. J., Abe-Ouchi, A., Pickering, S. J., Ramstein, G., Rosenbloom, N. A., Salzmann, U., Sohl, L., Stepanek, C., Ueda, H., Yan, Q., and Zhang, Z.: Large-scale features of Pliocene climate: results from the Pliocene Model Intercomparison Project, *Clim. Past*, 9, 191–209, doi:10.5194/cp-9-191-2013, 2013.
- Hewitt, C. D. and Mitchell, J. F. B.: Radiative forcing and response of a GCM to ice age boundary conditions: Cloud feedback and climate sensitivity, *Clim. Dynam.*, 13, 821–834, 1997.
- Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orłowsky, B., and Stepanek, P.: Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, *Nat. Geosci.*, 4, 17–21, 2011.
- Huang, S. P., Pollack, H. N., and Shen, P.-Y.: Temperature trends over the past five centuries reconstructed from borehole temperatures, *Nature*, 403, 756–758, 2000.
- IPCC SREX: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, in: *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*, edited by: Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M., Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp., 2012.
- Izumi, K., Bartlein, P. J., and Harrison, S. P.: Consistent behaviour of the climate system in response to past and future forcing. *Geophys. Res. Lett.*, 40, 1817–1823, doi:10.1002/grl.50350, 2013.
- Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D.: 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*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, 2007.
- Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., van Ommen, T. D., Vinther, B. M., Luterbacher, J., Wahl, E. R., Zwiers, F. W., Mann, M. E., Schmidt, G. A., Ammann, C. M., Buckley, B. M., Cobb, K. M., Esper, J., Goosse, H., Graham, N., Jansen, E., Kiefer, T., Kull, C., Küttel, M., Mosley-Thompson, E., Overpeck, J. T., Riedwyl, N., Schulz, M., Tudhope, A. W., Villalba, R., Wanner, H., Wolff, E., and Xoplaki, E.: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects, *Holocene*, 19, 3–49, doi:10.1177/0959683608098952, 2009.
- Kageyama, M., Peyron, O., Pinot, S., Tarasov, P., Guiot, J., Jous-saume, S., and Ramstein, G.: The Last Glacial Maximum climate over Europe and Western Siberia: a PMIP comparison between models and data, *Clim. Dynam.*, 17, 23–43, 2001.
- Kageyama, M., Braconnot, P., Bopp, L., Caubel, A., Foujols, M.-A., Guilyardi, E., Khodri, M., Lloyd, J., Lombard, F., Mariotti, V., Marti, O., Roy, T., and Woillez, M.-N.: Mid-Holocene and Last Glacial Maximum climate simulations with the IPSL model, Part I: comparing IPSL\_CM5A to IPSL\_CM4, *Clim. Dynam.*, 40, 2447–2468, doi:10.1007/s00382-012-1488-8, 2013.
- Kang, S. M., Held, I. M., Frierson, D. M. W., and Zhao, M.: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM, *J. Climate*, 21, 3521–3532, doi:10.1175/2007JCLI2146.1, 2008.
- Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., and Klein Goldewijk, K.: Holocene carbon emissions as a result of anthropogenic land cover change, *Holocene*, 21, 775–791, doi:10.1177/0959683610386983, 2011.
- Kharin, V. V., Zwiers, F. W., and Zhang, X. B.: Intercomparison of near-surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations, *J. Climate*, 18, 5201–5223, 2005.
- Kharin, V. V., Zwiers, F. W., Zhang, X. B., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *J. Climate*, 20, 1419–1444, 2007.
- Kohfeld, K. E. and Harrison, S. P.: How well can we simulate past climates? Evaluating earth system models using global palaeoenvironmental datasets, *Quaternary Sci. Rev.*, 19, 321–346, 2000.
- Köhler, P., Bintanja, R., Fischer, H., Joos, F., Knutti, R., Lohmann, G., and Masson-Delmotte, V.: What caused Earth's temperature variations during the last 800,000 years? Data-based evidence on radiative forcing and constraints

- on climate sensitivity, *Quaternary Sci. Rev.*, 29, 129–145, doi:10.1016/j.quascirev.2009.09.026, 2010.
- Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., and Mearns, L.: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections, in: Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., and Midgley, P. M., IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, 2010a.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A.: Challenges in combining projections from multiple climate models, *J. Climate*, 23, 2739–2758, 2010b.
- Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, *Nat. Clim. Change*, 3, 369–373, 2012.
- Koenker, R.: *Quantile regression*, Cambridge University Press, Cambridge, 2005.
- Kutzbach, J. E.: Monsoon climate of the early Holocene: Climate experiment with the Earth's orbital parameters for 9000 years ago, *Science*, 214, 59–61, 1981.
- Lainé, A., Kageyama, M., Braconnot, P., and Alkama, R.: Impact of greenhouse gas concentration changes on the surface energetics in the IPSL-CM4 model: Regional warming patterns, land/sea warming ratio, glacial/interglacial differences, *J. Climate*, 22, 4621–4635, 2009.
- Lambert, F. H., Webb, M. J., and Joshi, M. M.: The relationship between land–ocean surface temperature contrast and radiative forcing, *J. Climate*, 24, 3239–3256, doi:10.1175/2011JCLI3893.1, 2011.
- LeGrande, A. N. and Schmidt, G. A.: Global gridded data set of the oxygen isotopic composition in seawater, *Geophys. Res. Lett.*, 33, L12604, doi:10.1029/2006GL026011, 2006.
- Le Roy Ladurie, E.: *Histoire humaine et comparée du climat, Canicules et glacier, XIII<sup>e</sup>–XVIII<sup>e</sup> siècle*, Fayard, Paris, 2004.
- Le Roy Ladurie, E.: *Histoire humaine et comparée du climat, Disettes et révolutions 1740–1860*, Fayard, Paris, 2006.
- Ljungqvist, F. C.: A new reconstruction of temperature variability in the extra-tropical Northern Hemisphere during the last two millennia, *Geograf. Ann. A*, 92, 339–351, 2010.
- Lorius, C., Jouzel, J., Raynaud, D., Hansen, J., and Treut, H. L.: The ice-core record: Climate sensitivity and future greenhouse warming, *Nature*, 347, 139–145, 1990.
- Lovejoy, S. and Schertzer, D.: Scale invariance in climatological temperatures and the spectral plateau, *Ann. Geophys.*, 4B, 401–410, 1986.
- Lovejoy, S. and Schertzer, D.: Haar wavelets, fluctuations and structure functions: convenient choices for geophysics, *Nonlin. Processes Geophys.*, 19, 513–527, doi:10.5194/npg-19-513-2012, 2012a.
- Lovejoy, S. and Schertzer, D.: Stochastic and scaling climate sensitivities: solar, volcanic and orbital forcings, *Geophys. Res. Lett.*, 39, L11702, doi:10.1029/2012GL051871, 2012b.
- Lovejoy, S. and Schertzer, D.: *The Weather and Climate: Emergent Laws and Multifractal Cascades*, Cambridge University Press, Cambridge, 496 pp., 2013.
- Mahlstein, I. and Knutti, R.: September Arctic sea ice predicted to disappear near 2 °C global warming above present, *J. Geophys. Res.*, 117, D06104, doi:10.1029/2011JD016709, 2012.
- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *P. Natl. Acad. Sci.*, 105, 13252–13257, doi:10.1073/pnas.0805721105, 2008.
- MARGO Project Members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, *Nat. Geosci.*, 2, 127–132, doi:10.1038/NGEO411, 2009.
- Masson, D. and Knutti, R.: Climate model genealogy, *Geophys. Res. Lett.*, 38, L08703, doi:10.1029/2011GL046864, 2011.
- Masson-Delmotte, V., Kageyama, M., Braconnot, P., Charbit, S., Krinner, G., Ritz, C., Guilyardi, E., Jouzel, J., Abe-Ouchi, A., Crucifix, M., Gladstone, R., Hewitt, C., Kitoh, A., LeGrande, A., Marti, O., Merkel, U., Motoi, T., Ohgaito, R., Otto-Bliesner, B., Peltier, W., Ross, I., Valdes, P., Vettoretti, G., Weber, S., Wolk, F., and Yu, Y.: Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints, *Clim. Dynam.*, 26, 513–529, doi:10.1007/s00382-005-0081-9, 2006a.
- Masson-Delmotte, V., Kageyama, M., Braconnot, P., Charbit, S., Krinner, G., Ritz, C., Guilyardi, E., Jouzel, J., Abe-Ouchi, A., Crucifix, M., Gladstone, R., Hewitt, C., Kitoh, A., LeGrande, A., Marti, O., Merkel, U., Motoi, T., Ohgaito, R., Otto-Bliesner, B., Peltier, W., Ross, I., Valdes, P., Vettoretti, G., Weber, S., Wolk, F., and Yu, Y.: Past and future polar amplification of climate change: climate model intercomparisons and ice-core constraints, *Clim. Dynam.*, 27, 437–440, doi:10.1007/s00382-006-0149-1, 2006b.
- Masson-Delmotte, V., Braconnot, P., Hoffmann, G., Jouzel, J., Kageyama, M., Landais, A., Lejeune, Q., Risi, C., Sime, L., Sjolte, J., Swingedouw, D., and Vinther, B.: Sensitivity of interglacial Greenland temperature and  $\delta^{18}\text{O}$ : ice core data, orbital and increased CO<sub>2</sub> climate simulations, *Clim. Past*, 7, 1041–1059, doi:10.5194/cp-7-1041-2011, 2011.
- Massonnet, F., Fichet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland, M. M., and Barriat, P.-Y.: Constraining projections of summer Arctic sea ice, *The Cryosphere*, 6, 1383–1394, doi:10.5194/tc-6-1383-2012, 2012.
- McKay, J. L., de Vernal, A., Hillaire-Marcel, C., Not, C., Polyak, L., and Darby, D.: Holocene fluctuations in Arctic sea-ice cover: dinocyst-based reconstructions for the Eastern Chukchi Sea, *Can. J. Earth Sci.*, 45, 1377–1397, 2008.
- Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C., Watterson, I. G., Weaver, A. J., and Zhao, Z.-C.: Global climate projections, 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*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.
- Moberg, A., Sonnechkin, D. M., Holmgren, K., and Datsenko, N. M.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, 433, 613–617, 2005.

- Moros, M., Andrews, J. T., Eberl, D. D., and Jansen, E.: The Holocene history of drift ice in the northern North Atlantic – evidence for different spatial and temporal modes, *Palaeoceanography*, 21, PA2017, doi:10.1029/2005PA001214, 2006.
- Mueller, B. and Seneviratne, S. I.: Hot days induced by precipitation deficits at the global scale, *P. Natl. Acad. Sci.*, 109, 12398–12403, doi:10.1073/pnas.1204330109, 2012.
- Müller, J., Wagner, A., Fahl, K., Stein, R., Prange, M., and Lohmann, G.: Towards quantitative sea ice reconstructions in the northern North Atlantic: A combined biomarker and numerical modelling approach, *Earth Planet. Sc. Lett.*, 306, 137–148, 2011.
- Nigam, S. and Ruiz-Barradas, A.: Seasonal hydroclimate variability over North America in global and regional reanalyses and AMIP Simulations: Varied representation, *J. Climate*, 19, 815–837, doi:10.1175/JCLI3635.1, 2006.
- Nogaj, M., Yiou, P., Parey, S., Malek, F., and Naveau, P.: Amplitude and frequency of temperature extremes over the North Atlantic region, *Geophys. Res. Lett.*, 33, L10801, doi:10.1029/2005GL024251, 2006.
- O’ishi, R. and Abe-Ouchi, A.: Polar amplification in the mid-Holocene derived from dynamical vegetation change with a GCM, *Geophys. Res. Lett.*, 38, L14702, doi:10.1029/2011GL048001, 2011.
- Oppo, D. W., Schmidt, G. A., and LeGrande, A. N.: Seawater isotope constraints on tropical hydrology during the Holocene, *Geophys. Res. Lett.*, 34, L13701, doi:10.1029/2007GL030017, 2007.
- Osborn, T., Raper, S., and Briffa, K.: Simulated climate change during the last 1,000 years: comparing the ECHO-G general circulation model with the MAGICC simple climate model, *Clim. Dynam.*, 27, 185–197, 2006.
- PALAEOSENS Project Members: Making sense of palaeoclimate sensitivity, *Nature*, 491, 683–691, doi:10.1038/nature11574, 2012.
- Parey, S., Dacunha-Castelle, D., and Hoang, T. T. H.: Mean and variance evolutions of the hot and cold temperatures in Europe, *Clim. Dynam.*, 34, 345–359, 2010a.
- Parey, S., Hoang, T. T. H., and Dacunha-Castelle, D.: Different ways to compute temperature return levels in the climate change context, *Environmetrics*, 21, 698–718, 2010b.
- Pausata, F. S. R., Li, C., Wettstein, J. J., Kageyama, M., and Nisancioglu, K. H.: The key role of topography in altering North Atlantic atmospheric circulation during the last glacial period, *Clim. Past*, 7, 1089–1101, doi:10.5194/cp-7-1089-2011, 2011.
- Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M., Jennings, A. E., Miller, G. H., O’Regan, M., Savelle, J., Serreze, M., St. John, K., White, J. W. C., and Wolff, E.: History of sea ice in the Arctic, *Quaternary Sci. Rev.*, 29, 1757–1778, doi:10.1016/j.quascirev.2010.02.010, 2010.
- Pongratz, J., Reick, C. H., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, *Global Biogeochem. Cy.*, 22, GB3018, doi:10.1029/2007GB003153, 2008.
- Prado, L. F., Wainer, I., Chiessi, C. M., Ledru, M.-P., and Turcq, B.: A mid-Holocene climate reconstruction for eastern South America, *Clim. Past*, 9, 2117–2133, doi:10.5194/cp-9-2117-2013, 2013.
- Prentice, I. C. and Harrison, S. P.: Ecosystem effects of CO<sub>2</sub> concentration: evidence from past climates, *Clim. Past*, 5, 297–307, doi:10.5194/cp-5-297-2009, 2009.
- Quesada, B., Vautard, R., Yiou, P., Hirschi, M., and Seneviratne, S. I.: Asymmetric European summer heat predictability from wet and dry southern winters and springs, *Nat. Clim. Change*, 2, 736–741, doi:10.1038/nclimate1536, 2012.
- Ramstein, G., Kageyama, M., Guiot, J., Wu, H., Hély, C., Krinner, G., and Brewer, S.: How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison, *Clim. Past*, 3, 331–339, doi:10.5194/cp-3-331-2007, 2007.
- Reichler, T. and Kim, J.: How well do coupled models simulate today’s climate?, *B. Am. Meteorol. Soc.*, 89, 303–311, 2008.
- Rind, D. and Peteet, D.: Terrestrial conditions at the last glacial maximum and CLIMAP sea-surface temperature estimates: are they consistent?, *J. Geophys. Res.*, 94, 12851–12871, 1985.
- Russon, T., Tudhope, A. W., Hegerl, G. C., Collins, M., and Tindall, J.: Inter-annual tropical Pacific climate variability in an isotope-enabled CGCM: implications for interpreting coral stable oxygen isotope records of ENSO, *Clim. Past*, 9, 1543–1557, doi:10.5194/cp-9-1543-2013, 2013.
- Rybski, D., Bunde, A., and von Storch, H.: Long-term memory in 1000-year simulated temperature records, *J. Geophys. Res.*, 113, D02106, doi:10.1029/2007JD008568, 2008.
- Santer, B. D., Taylor, K. E., Gleckler, P. J., Bonfils, C., Barnett, T. P., Pierce, D. W., Wigley, T. M. L., Mears, C., Wentz, F. J., Brüggemann, W., Gillett, N. P., Klein, S. A., Solomon, S., Stott, P. A., and Wehner, M. F.: Incorporating model quality information in climate change detection and attribution studies, *P. Natl. Acad. Sci.*, 106, 14778–14783, 2009.
- Schär, C., Lüthi, D., Beyerle, U., and Heise, E.: The soil-precipitation feedback: A process study with a regional climate model, *J. Climate*, 12, 722–741, 1999.
- Schmidt, G. A.: Enhancing the relevance of palaeoclimate model/data comparisons for assessments of future climate change, *J. Quaternary Sci.*, 25, 79–87, doi:10.1002/jqs.1314, 2010.
- Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., Bauer, S., Cairns, B., Canuto, V., Cheng, Y., Del Genio, A., Faluvegi, G., Friend, A. D., Hall, T. M., Hu, Y., Kelley, M., Kiang, N. Y., Koch, D., Lacis, A. A., Lerner, J., Lo, K. K., Miller, R. L., Nazarenko, L., Oinas, V., Perlwitz, Ja., Perlwitz, Ju., Rind, D., Romanou, A., Russell, G. L., Sato, M., Shindell, D. T., Stone, P. H., Sun, S., Tausnev, N., Thresher, D., and Yao, M.-S.: Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data, *J. Climate*, 19, 153–192, doi:10.1175/JCLI3612.1, 2006.
- Schmidt, G. A., LeGrande, A., and Hoffmann, G.: Water isotope expressions of intrinsic and forced variability in a coupled ocean–atmosphere model, *J. Geophys. Res.*, 112, D10103, doi:10.1029/2006JD007781, 2007.
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Bracco, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, 4, 33–45, doi:10.5194/gmd-4-33-2011, 2011.

- Schmidt, G. A., Jungclauss, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), *Geosci. Model Dev.*, 5, 185–191, doi:10.5194/gmd-5-185-2012, 2012.
- Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., DelGenio, A., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A., Perlwitz, J., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, doi:10.1002/2013MS000265, in press, 2014.
- Schmittner, A., Urban, N. M., Shakun, J. D., Mahowald, N. M., Clark, P. U., Bartlein, P. J., Mix, A. C., and Rosell-Melé, A.: Climate Sensitivity Estimated from Temperature Reconstructions of the Last Glacial Maximum, *Science*, 334, 1385–1388, doi:10.1126/science.1203513, 2011.
- Schneider von Deimling, T., Ganopolski, A., Held, H., and Rahmstorf, S.: How cold was the Last Glacial Maximum?, *Geophys. Res. Lett.*, 33, L14709, doi:10.1029/2006GL026484, 2006.
- Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B., and Phipps, S. J.: Separating forced from chaotic climate variability over the past millennium, *J. Climate*, 26, 6954–6973, doi:10.1175/JCLI-D-12-00826.1, 2013.
- Seneviratne, S. I. and Koster, R. D.: A revised framework for analyzing soil moisture memory in climate data: Derivation and interpretation, *J. Hydrometeorol.*, 13, 404–412, 2012.
- Seneviratne, S. I., Lüthi, D., Litschi, M., and Schär, C.: Land-atmosphere coupling and climate change in Europe, *Nature*, 443, 205–209, 2006.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, *Earth Sci. Rev.*, 99, 125–161, 2010.
- Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C. M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC SREX Report), edited by: Field, C. B., Barros, V., Stocker, T. F., Qin, D., Dokken, D. J., Ebi, K. L., Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Allen, S. K., Tignor, M., and Midgley, P. M., Cambridge University Press, Cambridge, 2012.
- Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, *Nature*, 491, 435–438, doi:10.1038/nature11575, 2012.
- Steinhilber, F., Beer, J., and Fröhlich, C.: Total solar irradiance during the Holocene, *Geophys. Res. Lett.*, 36, L19704, doi:10.1029/2009GL040142, 2009.
- Stroeve, J. C., Kattsov, V., Barrett, A., Serreze, M., Pavlova, T., Holland, M., and Meier, W. N.: Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophys. Res. Lett.*, 39, L16502, doi:10.1029/2012GL052676, 2012.
- Sundberg, R., Moberg, A., and Hind, A.: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium – Part 1: Theory, *Clim. Past*, 8, 1339–1353, doi:10.5194/cp-8-1339-2012, 2012.
- Sutton, R. T., Dong, B., and Gregory, J. M.: Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations, *Geophys. Res. Lett.*, 34, L02701, doi:10.1029/2006GL028164, 2007.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, 106, 7183–7192, 2001.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the experiment design, *B. Am. Meteorol. Soc.*, 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.
- Telford, R. J., Li, C., and Kucera, M.: Mismatch between the depth habitat of planktonic foraminifera and the calibration depth of SST transfer functions may bias reconstructions, *Clim. Past*, 9, 859–870, doi:10.5194/cp-9-859-2013, 2013.
- Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., and Emile-Geay, J.: Comparison of observed and simulated tropical climate trends using a forward model of coral  $\delta^{18}\text{O}$ , *Geophys. Res. Lett.*, 38, L14706, doi:10.1029/2011GL048224, 2011.
- Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., Emile-Geay, J., and LeGrande, A. N.: Coral-CGCM comparison highlights role of salinity in long-term trends, in: El Niño Southern Oscillation: observation and modeling, edited by: Braconnot, P., Brierley, C., Harrison, S. P., and von Gunten, L., *PAGES News*, 21, 60–61, 2013.
- Tremblay, L.-B., Huard, D., Schmidt, G. A., and de Vernal, A.: Mid Holocene constraints on future Arctic climate change, in preparation, 2014.
- Trenberth, K. E. and Fasullo, J. T.: Simulation of present-day and twenty-first-Century energy budgets of the Southern Oceans, *J. Climate*, 23, 440–454, doi:10.1175/2009JCLI3152.1, 2010.
- Vautard R., Yiou, P., D’Andrea, F., de Noblet, N., Viovy, N., Cassou, C., Polcher, J., Ciais, P., Kageyama, M., and Fan, Y.: Summer-time European heat and drought waves induced by wintertime Mediterranean rainfall deficit, *Geophys. Res. Lett.*, 34, L07711, doi:10.1029/2006GL028001, 2007.
- Vieira, L. E. A., Solanki, S. K., Krivova, N. A., and Usoskin, I.: Evolution of the solar irradiance during the Holocene, *Astron. Astrophys.*, 531, A6, doi:10.1051/0004-6361/201015843, 2011.
- Vyushin, D. I., Kushner, P. J., and Zwiers, F.: Modeling and understanding persistence of climate variability, *J. Geophys. Res.*, 117, D21106, doi:10.1029/2012JD018240, 2012.
- Werner, M., Mikolajewicz, U., Heimann, M., and Hoffmann, G.: Borehole Versus Isotope Temperatures on Greenland: Seasonality Does Matter, *Geophys. Res. Lett.*, 27, 723–726, 2000.
- Wilby, R. L. and Wigley, T. M. L.: Downscaling general circulation model output: a review of methods and limitations, *Prog. Phys. Geogr.*, 21, 530–548, 1997.
- Wittenberg, A. T.: Are historical records sufficient to constrain ENSO simulations?, *Geophys. Res. Lett.*, 36, L12702, doi:10.1029/2009GL038710, 2009.

- Yiou, P., Dacunha-Castelle, D., Parey, S., and Hoang, T. T. H.: Statistical representation of temperature mean and variability in Europe, *Geophys. Res. Lett.*, 36, L04710, doi:10.1029/2008GL036836, 2009.
- Yiou, P., Servonnat, J., Yoshimori, M., Swingedouw, D., Khodri, M., and Abe-Ouchi, A.: Stability of weather regimes during the last millennium from climate simulations, *Geophys. Res. Lett.*, 39, L08703, doi:10.1029/2012GL051310, 2012.
- Zhu, X., Fraederich, L., and Blender, R.: Variability regimes of simulated Atlantic MOC, *Geophys. Res. Lett.*, 33, L21603, doi:10.1029/2006GL027291, 2006.