We thank both reviewers for their comments, which we have addressed in this revised manuscript. (see below)

In addition to the changes suggested by the reviewers, we have incorporated two additional models into our analysis – NORESM (from BCCR) and CSIRO (from UNSW). As such, we have additional authors of Stephen Phipps, Petra Langebroek, Zhongshi Zhang, and Kerim Nisancioglu. All the tables, figures and text have been updated appropriately. We have also made several minor changes to the manuscript to improve readability.

This response consists of two parts: (1) a point-by-point response to the reviewers' comments, and (2) a copy of this revised submission with all the changes highlighted relative to the previous submission (blue shows additions, red shows deletions. Please note that the differences are calculated by latexdiff, which does mangle the file somewhat, including some references).

Additionally, in the final submission we will include Supplementary Information including netcdf files of all the model results.

Reviewer 1:

"one figure showing the difference between the models including dynamical vegetation and the models with stead-state/prescribed vegetation. In the final part of the discussion, the authors suggest that this could be important, however, part of the answer might already be included in the simulations considered in this paper. Since for some models, dynamical vegetation is also very resources-consuming to use/implement, it could give an idea on how this feedback is important for this particular time period. Personally, I am not convinced that this is the cause of majors discrepancies between the models at this particular time. That is why I just suggest to include such a figure, to show the importance of that perspective."

Although our results point at the possibility of vegetation being important, this can not be properly assessed without simulations from a single model, both with and without dynamic vegetation. As such, we would prefer not to overplay this aspect. Instead of a figure, we have added "As such, and without a set of comparable simulations from a single model, with and without dynamic vegetation, it is currently not possible to assess the impact of LIG vegetation feedbacks."

"one figure showing the difference in simulated temperatures between high and low resolution models. This can generate some large differences. I know quite well CCSM4 and I know that at low resolution for example, there is lack of oceanic heat transport towards the high latitudes generating a negative atmospheric temperature bias and a reduction in precipitation (Shield et al. 2012)."

We agree with this, and have strengthened this particular conclusion with a new figure (Figure 10), which shows the difference in response between the GCMs and EMICs.

Finally, since the title is "assessment of last interglacial temperatures" a last synthetic graph bars could be included, in the IPCC-like style, showing the range of seasonal temperature for each models. It could also makes Figure 5 more clear. All those points are only suggestions and I let the authors free to include them or not.

We are not sure what this would add to the figures already presented, so we choose not to include this plot. It would be interesting if there was data which could be used to evaluate this simple metric, but in the absence of such data we don't think such a plot would be informative.

page 3662 - line 13-15: this sentence contradict the main perspective of the paper suggested by the author: include dynamics vegetation feedback...consequently, this cannot be a "minor forcing" of the experiments

We have toned down the suggestions of vegetation being important, adding "As such, and without a set of comparable simulations from a single model, with and without dynamic vegetation, it is currently not possible to assess the impact of LIG vegetation feedbacks."

page 3666 - line 25-27: I am not sure I agree with this sentence. Most of the atmospheric variables at T31 or T42 horizontal resolution reach equilibrium after 100 to 200 years of simulations. This statement might be true for intermediate and deep ocean, but after a while, they do not influence significantly the atmospheric state. Therefore, the difference observed in Figure 4 may not be due entirely to the different in the length of simulations.

We chose to keep this statement because the deep ocean has been shown to affect the spinup of even surface variables (see Brandefelt and Otto-Bliesner. 2009 for an example from the LGM). However, we add a 'potentially'.

page 3666 - line 15-19: I think this statement is obvious since all the models are highly different. Also the fact that all the models exhibit a "similar behaviour" independently from which snapshot is used between 130k - 125k is quite obvious from Figure 1: precession does not vary much during this interval, the same for obliquity and eccentricity...

We agree that this is what is expected, but we don't see any harm in stating it, as the one of the zeroth-order findings.

page 3667 - line 5: may be due "to" models - add "to"

Done.

page 3667 - line 7-8: can you quantify the spread of the models and clarify what does "similar" mean for you?

We have clarified our meaning here: "It is also instructive to examine the ensemble mean response. In order to account for variations between models, and temporal variability through the LIG, we construct the ensemble mean as a straightforward average of all the simulations presented in Fig. 5"

page 3668 - line 8-9: do you mean that insolation forcing is negative relative to preindustrial?

Yes, added "relative to preindustrial".

page 3672 - line 6-12: see Shield et al. (2012) about the low versus high resolution CCSM4, as example of improvements of processes and resolutions

Although this paper is interesting, it is primarily reporting results about CCSM4, which is not included in this study; as such we choose not to reference it here.

page 3674 - line 3-6: I don't agree with this statement because MIS 11 and MIS 5 are very warm interglacial according to data, while observation suggest that MIS 7 is cooler than those two. I think that no particular interglacial can be considered for a specific calibration of the process. Also because the last interglacial during the Holocene was warm, but not as warm as during MIS 5. In that sense, it is very difficult to base models tuning and calibration on one specific interglacial.

We have re-worded this sentence to make our meaning clearer:

"Finally, this work indicates that although other interglacials, such as MIS 7 to MIS 11, could also be potentially useful targets for models (e.g. Yin et al), in terms of model-data comparison more benefit would probably be gained by improving aspects of the LIG data compilations first."

Table 1: there are some missing values for CLIMBER LSCE

Fixed.

Table 2: missing values for IPSL_LSCE

Fixed.

Figure 5: This figure is actually very hard to appreciate. The scale is cut at half, and one has to zoom in very much to see the dots from Turney and Jones (2010). Even in full screen, this figure is difficult ti analyse. You should remove the data from Turney and Jones (2010) since it's not visible. Maybe in landscape format, the figures would be more visible

We have cut this figure in half to enable it to be reproduced larger. We have also removed the spurious '129k' column, as there are no simulations of this time period. As such, we believe it should be large enough. At this larger size, we also prefer to keep the data points as they are.

Figure 6: How significant are those anomaly relative to Pre-industrial?

The significance of these anomalies is illustrated in the Figure by stippling. This shows the regions where less than 70% of the models agree on the sign of the change. This is the standard method as used by the IPCC in AR4, which we adopt here.

Reviewer 2:

General point 1. A main finding of the study is that, while the model simulations do adequately reflect the changes in LIG insolation pattern, they do not simulate the overall climate of the LIG very well. A possible implication from this is that although the model atmospheric dynamics are adequate to capture the 1st order insolation impacts, the models themselves are not yet adequate to the job of simulating past warm climates e.g. the coupled ocean-atmosphere dynamics are not up to the job / too highly tuned to present-day climate. If this is an accurate reflection of the authors own view, perhaps this could perhaps be stated more clearly in the conclusions/abstract.

We don't think that this is a conclusion that can be made from this paper. At the moment there are just too many uncertainties in the proxy data themselves, and as such it is premature to make this conclusion. Instead, we provide suggestions of ways both the models and data could be improved, in order to facilitate future model-data comparisons.

General point 2. Whilst it is difficult to know how to plot figures which adequately show the results from many simulations together, some of the figures are really difficult to read, with too many tiny panels. Fig 5 is particularly bad for this. Is there perhaps some means of making Fig 4 and esp Fig 5 a bit more readable?

We have cut Fig 5 figure in half to enable it to be reproduced larger. We have also removed the spurious '129k' column, as there are no simulations of this time period. As such, we believe it should be large enough.

Abstract L13 Occasionally 'model' is used where 'simulation' would be better (there are a few other instance of this).

Agreed – we have changed to 'models' to 'model simulations' in several places.

Abstract L16 Change phrase 'far from perfect'.

Changed to "Taking possible seasonal biases in the proxies into account improves the agreement, but only marginally".

P3660 L9 This is not correctly expressed as 'a gradient', perhaps a 'difference'?

Changed to "pole-to-equator temperature difference"

P3663 L29 I personally don't much like the use of 'flavours'. Can you use 'versions'? Flavours is very vague and seems a bit unhelpful re: climate modelling. (There are other instance of the 'flavours' terminology that should also ideally be changed.)

Changed "flavour" to "version" throughout.

P3665 L11-19 Can you add a line to say why you chose to use NCEP reanalysis as opposed to any other product?

Added "We choose NCEP as opposed to any other reanalyses product purely for pragmatic reasons in that we had it readily available."

P3665 L20- It would also be interesting to know what the errors are specifically at the sites where you have LIG proxy data. An error (model skill) score for the preindustrial simulation could usefully be added to Table 1 or Table 2.

We think that the most important and robust metric for assessing model performance is the global RMS error, which we have added to Table 1 as requested.

P3666 L4-7 Work on the use of multi-ensembles of simulations has previously tended to use a larger suit of purely GCM-based simulations. I think there is perhaps no precedent for a mixed EMIC/GCM ensemble and one which features a rather small number of independent GCM simulations. Some further consideration of this and comments and/or references would be useful. See also comment on Fig 7 below.

This sort of ensemble is actually not uncommon in the palaeoclimate modelling community. However, we have added "Previous comparisons of EMICs with GCMs (Stouffer et al, 2006} have not reported such large differences, and it is possible that our results are biased by the relatively few EMICs in this study"

P3666 L23 Can the statement on the reasons for differences between NCAR and BREMEN 'flavours/versions' be firmed up?

We now reference Table 1 to clarify the inter-model differences, and have edited this section to read "This is probably related to the higher resolution of CCSM3_NCAR (T85 compared with T31), and the use of dynamic vegetation in CCSM3_Bremen (see Table 1). CCSM3_LLN appears to be more similar to CCSM3_Bremen than to CCSM3_NCAR. CCSM3_LLN has the same T31 resolution as CCSM3_Bremen, but similar to CCSM3_NCAR does not include dynamic vegetation, implying

that in CCSM3 the resolution has more of an effect on the climate than the inclusion of dynamic vegetation."

P3666-P3669 Some subheadings for this section would aid readability e.g. 'seasonal differences', 'mean annual differences', etc..

Agreed. We have divided this long section into two subsections, dealing with the individual models and then the ensemble mean.

P3670 L5- A table containing the simulation skill scores, or another column in the pre-existing Table 2, would be helpful.

Done.

P3671 L24-25 Reference required here re: evidence.

Done.

P3675 L5 Please summarise here the main improvement recommendations: they sounds like a main conclusion.

Agreed. Added "On the data side this includes the incorporation of error bars in the proxy datasets, and inclusion of seasonal proxies in order to capture the largest signals. On the model side, this includes more studies on the role of vegetation and fresh water forcing."

Table 1. Would be useful to have a it more detail for more models in column Other/Model resolution.

Done.

Fig. 4 and Table 1/2. There appears (on use of a microscope) to be possible discrepancies between model/simulation titles between panel headers and the table (e.g. ECHAM5?).

Done.

Fig. 7. Given some reservations about the compilation/use of a simulation ensemble, it would be nice if comparisons for individual (best case?) simulation could also be shown.

We have added an extra figure as Supplementary Information which shows the model-data comparison for each simulations from each model.

Manuscript prepared for Clim. Past
with version 5.0 of the LaTeX class copernicus.cls.
Date: 24 February 2013

A multi-model assessment of last interglacial temperatures

D. J. Lunt[1], A. Abe-Ouchi[2], P. Bakker[3], A. Berger[4], P. Braconnot[5], S. Charbit[5], N. Fischer[6], N. Herold[4], J. H. Jungclaus[6], V. C. Khon[7], U. Krebs-Kanzow[7], P. M. Langebroek[8], G. Lohmann[9], K. H. Nisancioglu[10], B. L. Otto-Bliesner[11], W. Park[12], M. Pfeiffer[9], S. J. Phipps[13], M. Prange[14], R. Rachmayani[14], H. Renssen[3], N. Rosenbloom[11], B. Schneider[7], E. J. Stone[1], K. Takahashi[15], W. Wei[9], Q. Yin[4], and Z. S. Zhang[8]

[1]School of Geographical Sciences, University of Bristol, UK
[2]Department of Earth and Planetary Sciences, University of Tokyo, Japan
[3]Department of Earth Sciences, VU University Amsterdam, The Netherlands
[4]Georges Lemaître Center for Earth and Climate Research,
Université catholique de Louvain, Belgium
[5]Laboratoire des Sciences du Climat et de l'Environnement, unité mixte CEA-CNRS-UVSQ, France
[6]Max Planck Institute for Meteorology, Germany
[7]Department of Geology, Kiel University, Germany
[8]UNI Research and Bjerknes Centre for Climate Research, Bergen, Norway
[9]Alfred Wegener Institute for Polar and Marine Research, Germany
[10]University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway
[11]National Center for Atmospheric Research, USA
[12]GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel, Germany
[13]ARC Centre of Excellence for Climate System Science and Climate Change Research Centre, University of New South Wales, Sydney, Australia
[14]MARUM – Center for Marine Environmental Sciences, University of Bremen, Germany
[15]JAMSTEC, Yokohama, Japan

Correspondence to: D. J. Lunt (d.j.lunt@bristol.ac.uk)

Abstract. The Last Interglaciation ($\sim$130 to 116 ka) is a time period with a strong astronomically-induced seasonal forcing of insolation compared to modern. Proxy records indicate a significantly different climate to that of the modern, in particular Arctic summer warming and higher eustatic sea level. Because the forcings are relatively well constrained, it provides an opportunity to test numerical models which are used for future climate prediction. In this paper, we compile a set of climate model simulations of the early Last Interglaciation (130 to 125 ka), encompassing a range of model complexity. We compare the simulations ~~models~~ to each other, and to a recently published compilation of Last Interglacial temperature estimates. We show that the annual mean response of the models is rather small, with no clear signal in many regions. However, the seasonal response is more robust, and there is significant agreement amongst models as to the regions of warming vs. cooling. However, the quantitative agreement of the model simulations ~~models~~ with data is poor, with the models in general underestimating the magnitude of response seen in the proxies. Taking possible seasonal biases in the proxies into account improves the agreement, but only marginally~~marginally, but the agreement is still far from perfect~~. However, a lack of uncertainty estimates in the data does not allow us to draw firm conclusions. Instead, this paper points to several ways in which both modelling and data could be improved, to allow a more robust model-data comparison.

1 Introduction

The last interglaciation (LIG, $\sim$130 to 116 ka) is the penultimate ~~most recent~~ interglaciation (period of reduced terrestrial ice cover relative to glacial periods) in Earth's history, prior to the current interglaciation (Holocene, $\sim$12 to 0 ka). In common with the Holocene, the early LIG (here, 130 to 125 ka) is characterised by a maximum in δD in Antarctic ice cores (**?**) and a minimum in benthic δ^{18}O in marine sediment cores (**?**), which qualitatively indicate a relatively warm climate and/or reduced terrestrial ice volume.

Palaeo data archives indicate that the climate of the LIG differed from that of the modern. A compilation of terrestrial and marine records (**?**) indicates a global mean warming relative to preindustrial of about 2 °C. A compilation of SST records (**?**) indicates a global mean SST warming relative to the late Holocene of 0.7 ± 0.6 °C. The maximum annual mean warming occurred in mid and high Northern Hemisphere latitudes, reducing the pole-to-equator temperature difference ~~meridional temperature gradient~~ by about 1.5 °C relative to preindustrial (**?**). This was associated with changes in vegetation patterns, notably a northwards shift of boreal forest across the Arctic (e.g. in Scandinavia, **?**, Alaska, **?**, and Siberia, **?**). Palaeo archives can also give an indication of seasonal changes in temperature; for example, records have been interpreted as representing Arctic summer temperatures about 5 °C warmer than present, with an associated decrease in summer sea ice (**?**). Ocean circulation also varied through the LIG, with North Atlantic δ^{13}C and ^{231}Pa/^{230}Th records indicating increasing AMOC strength in the early LIG, and maximum overturning in the middle of the LIG (**?**).

A compilation of global sea level records (**?**) indicates a LIG highstand of at least 6.6 m (95 % probability), and likely in excess of 8.0 m (67 % probability). Such records have been interpreted as representing contributions from reduced volume of both Greenland and West Antarctic ice sheets (**?**). A substantial contribution from the Greenland ice sheet at the LIG is supported by modelling evidence (**??**), which indicates a contribution from Greenland of 0.3 m to 3.6 m (80 % probability). A contribution from Antarctica is supported by benthic δ^{18}O and modelling evidence (**?**).

The principal driver of climatic differences between LIG and modern climate is the astronomical configuration of the Earth. The early LIG is characterised by relatively high obliquity and eccentricity compared with modern, and a precessional component with boreal summer coinciding with perihelion (**??**). This results in an insolation anomaly relative to modern consisting of a maximum in boreal summer and minimum in austral summer (Fig. **??**). A secondary driver is natural variations in greenhouse gases (**???**), which were fairly constant through the LIG, but with a maximum in all three gases (CO_2, CH_4 and N_2O) between 129 and 128 ka (Fig. **??**).

Because of the very different principal forcing mechanisms (seasonal astronomical variations compared with greenhouse gas changes), the LIG should not be considered an *analogue* for future climate change. However, because of its relative warmth and high sea level, the LIG could be considered as an appropriate test-bed for climate models developed for future climate prediction. Furthermore, modelling studies suggest that over Greenland, the summer warming is amplified by similar albedo and water feedbacks to those found in future climate simulations (**?**). As such, the LIG has begun to receive more attention from the modelling community, and the Palaeoclimate Model Intercomparison project (now in its third phase, PMIP3, http://pmip3.lsce.ipsl.fr) has recently extended its focus from the Last Glacial Maximum (LGM, 21 ka) and mid-Holocene (6 ka) to include the LIG (as well as another warm period, the Pliocene, 3 Ma).

This paper describes an ensemble of climate model simulations of the LIG, many of which have been carried out using guidelines developed by PMIP. The simulations are "snapshots", that is, each one is designed to represent equilibrium conditions during a ~ 1 ka "window" during the LIG. There are a number of snapshots covering the period 125 to 130 ka, and they have been carried out using a range of climate models, representing a range of model complexity.

The aims of the paper are twofold:

– Firstly, to catalogue the differences between the model simulations, determining which features are robust, and where there is uncertainty, and to provide some first-order hypotheses for the mechanisms behind the large-scale features.

– Secondly, to compare the simulations with the latest data compilations, determining to what extent the model simulations ~~models~~ and data are consistent.

The focus of this paper is on temperature, because there are more proxy records for temperature than any other variable, and it is generally one of the more robustly modelled variables. We consider the terrestrial and marine realm for our model-data comparisons, and investigate the seasonality of the model simulations and proxy records.

2 Model simulation descriptions

As part of the third phase of PMIP, a set of four Last Interglacial snapshot simulations were proposed, at 130 ka, 128 ka, 125 ka, and 115 ka. Here, we focus on the first three of these, which encompass the time of maximum anomaly in insolation in Northern Hemisphere summer; the fourth was designed to look at glacial inception processes at the very end of the LIG. PMIP laid out a set of boundary conditions for these snapshots. These consisted of astronomical and greenhouse gas parameters, as it was decided to leave possible smaller forcings, such as vegetation, ice sheet, sea level and aerosol changes, to subsequent sensitivity studies.

The PMIP3 LIG astronomical ~~and~~ greenhouse gas boundary conditions are illustrated in Fig. **??** and Fig. **??** (and also can be read off Table **??**). The astronomical constants were obtained from **?**. The greenhouse gas concentrations were derived from Antarctic ice core records: **?** for CO_2 (although note that this is a composite record), **?** for CH_4 and **?** for N_2O. The raw greenhouse gas data was interpolated onto a 100-yr timestep, and the values for each snapshot taken from the appropriate time in this interpolated record.

The simulations used in this paper are all those which were submitted to a call for model contributions to this intercomparison, following a PMIP meeting in Crewe, UK, in May 2012. Table **??** gives some details of the models included in this intercomparison, and Table **??** gives some key aspects of their experimental design, including boundary conditions. The models cover a wide range of complexity, from state-of-the-art GCMs used in the fifth assessment report of the IPCC (e.g. COSMOS, MIROC), through GCMs which featured in the fourth assessment report (e.g. CCSM3, HadCM3), to models of intermediate complexity ("EMICs", e.g. LOVECLIM, CLIMBER).

Not all simulations described in this paper follow the PMIP3 guidelines. Indeed, some were carried out before the guidelines were developed. As such, this is an "ensemble of opportunity", in that there is not complete consistency across all the model simulations. However, most of the model simulations from any one organisation are self-consistent; e.g. the simulations are all carried out with the same model version. A minor exception is CCSM3_NCAR, where the LIG simulations have a slightly greater solar constant than the preindustrial simulation (see Table **??**).

All groups used identical land-sea masks and terrestrial ice sheets in their LIG simulations as compared with their controls; as such, greenhouse gases and/or astronomical configuration were the main external forcings imposed in the LIG simulations compared with the controls. Although groups may have used slightly different astronomical solutions, these differences are minimal (e.g. **?** give insolation values which differ from those of **?** by less than 0.1 % for these time-slices). Therefore, different greenhouse gas concentrations were the main inconsistency in experimental design between different groups. The various greenhouse gas concentrations applied by the different groups are illustrated in Fig. **??**.

Simulations carried out using HadCM3_Bris, CCSM3_Bremen, COSMOS_AWI, LOVECLIM_Ams, ~~and~~ CLIMBER_LSCE, CSIRO SUBSCRIPTNBUNSW and NORESM SUBSCRIPTNBBCCR were all carried out using the greenhouse gas boundary conditions specified by PMIP3. Simulations carried out by KCM_Kiel, ~~and~~ COSMOS_MPI and IPSL SUBSCRIPTNBLSCE chose to keep the LIG greenhouse gases fixed at the control values, and as such just included astronomical variations. The other models developed greenhouse gas changes independently. Most are relatively consistent, but CCSM3_NCAR at 130 ka does have higher values of CO_2, CH_4 and N_2O (but note that the CCSM3_NCAR preindustrial greenhouse gas levels are also relatively high, see Table **??**).

Some of the models are similar to each other – the most obvious being three versions ~~"flavours"~~ of CCSM3, the two versions ~~"flavours"~~ of LOVECLIM, and the two versions ~~"flavours"~~ of COSMOS. In the case of CCSM3, the model versions are different – CCSM3_NCAR runs at a higher resolution (T85~~T42~~) than the other two (T31), and CCSM3_Bremen includes dynamic vegetation. In the case of LOVECLIM, although the model versions are identical, the two groups have contributed different snapshots (125 k and 130 ka from LOVECLIM_Ams, and 127 ka from LOVECLIM_LLN). In the case of COSMOS, COSMOS_MPI uses dynamic vegetation in all simulations, whereas for COSMOS_AWI the LIG simulation (130 ka) is forced by a fixed preindustrial vegetation that has been taken from the equilibrated control simulation, which itself is spun-up using a dynamic vegetation scheme (**?**). KCM_Kiel uses the ECHAM5 atmosphere model (**?**), an atmospheric component also used in COSMOS~~is a hybrid of the atmosphere model in COSMOS~~, and the NEMO ocean-sea ice model (**?**), an ocean component also used ~~ocean model~~ in IPSL_LSCE. NORESM SUBSCRIPTNBBCCR is a hybrid of an updated version of the atmospheric component of CCSM3 (CAM4 compared with CAM3), and an independent ocean model (MICOM)~~(OPA-9)~~.

3 Last interglacial SST and land temperature dataset

For the model-data comparison in Sect. **??**, we make use of the terrestrial and ocean annual mean temperature reconstruction of **?**. This consists of 262 sites, made up of 100 terrestrial temperatures and 162 SSTs (see Fig. **??**). The data are derived from a diverse range of proxies, including: Sr-Ca, U^k_{37}, Mg/Ca and diatom and radiolarian assemblage transfer functions for SSTs, pollen and macrofossils for terrestrial temperatures, and $\delta^{18}O$ for ice sheet temperatures. Sites are only included in the compilations if they have 4 or more data points through the LIG; the reconstruction consists of the average temperature of the period of plateaued $\delta^{18}O$ for marine sequences, and maximum warmth for terrestrial sequences. The data are presented as anomalies relative to modern (averaged over the years 1961–1990). **?** noted a pattern of early warming off the southern African coastline and Indian Ocean, that they interpreted as evidence for leakage from the Indian Ocean via an enhanced Agulhas current, consistent with southward migration of the Southern Ocean westerlies. Here, we consider all sites as contemporaneous, although in reality they represent average conditions over a time window which varies from site to site. However, as we shall see, the modelled variability across the time window of interest is relatively small compared to other uncertainties.

Unfortunately, **?** give no indication of the uncertainties in their SST or terrestrial reconstructions. It is possible that some of the LIG sites may be more representative of a seasonal change as opposed to an annual mean change (e.g. see discussion in **?** in the context of the Holocene). ~~.~~ This is because the calibration of many of the proxies used is based on modern analogues, which are by definition all under modern astronomical conditions; because the astronomical configuration of the LIG is significantly different, this could result in a seasonal shift being interpreted as an annual mean change.

4 Results and model-data comparison

Before turning to the simulations of the LIG, it is worthwhile to put these into context, by examining potential biases in the preindustrial control simulations. These are illustrated in Fig. **??**, which shows the simulated preindustrial annual mean temperatures from each model relative to those from the NCEP reanalysis product (**?**). We choose NCEP as opposed to any other reanalyses product purely for pragmatic reasons in that we had it readily available. It should be noted that the NCEP reanalyses themselves are not perfect. In particular, in regions of sparse observational input, such as over Antarctica, the model "error" should be treated with caution. Furthermore, the observations represent a 40-yr average which starts in 1948, whereas the model control simulations represent a "preindustrial" time, and assume a range of greenhouse gas concentrations (see Table **??**).

Every model has at least one gridbox where the "error" is at least 10 °C. The models with the smallest RMS error are HadCM3_Bris and CCSM_NCAR, both with 2.4 °C, and the model with the largest RMS error is CLIMBER_LSCE, with 4.9~~4.8~~ °C. However, note that because the differences are calculated after interpolating all simulations and observations to a resolution of 96×73 gridboxes (the resolution of HadCM3 SUBSCRIPTNBBris), this penalises those models, like CLIMBER, with relatively low resolution. Also note that the CSIRO SUBSCRIPTNBUNSW model uses flux adjustment for all simulations, so the control has relatively low biases over the ocean. As expected, similar models show similar anomalies; for example, all CCSM3-type models have a cold bias in the North Atlantic, and all models with ECHAM5 atmospheric components have a cold bias in the central Sahara. Because the control model simulations have been run for very different lengths of time (see Table **??**), any small cooling or warming trends could also potentially contribute to the differences between model results~~models~~. Figure **??**o ~~m~~ shows the model ensemble mean. This has a lower RMS error than any individual model, 2.2 °C, and also has a relatively low error in the global mean, having a mean error of $-$0.73~~0.75~~ °C (a fraction of which is likely related to the difference between modern and preindustrial temperatures due to recent warming). The strong relative performance of the ensemble mean has been observed in many other model ensembles, and **?** show that this is consistent with the model simulations ~~models~~ and observations being considered as being drawn from the same statistical distribution.

4.1 Inter-model LIG comparison

4.1.1 Individual model responses

Figure **??** shows the annual mean surface air temperature (at ~ 1.5 m height) change, LIG minus preindustrial control, for each snapshot carried out by each model (although note that for NORESM SUBSCRIPTNBBCCR, the surface temperature is shown, as the surface air temperature was not available). There are several points worth noting here. Firstly, for nearly all models and ~~for all~~ snapshots, the maximum warming occurs in the mid to high latitudes of the Northern Hemisphere. The spread in predicted temperature change as a function of snapshot for any particular model, is less than the spread in predicted temperature as a function of model for any particular snapshot. In other words, which model is used has more of an influence on the predicted LIG climate than which snapshot is used (in the range 130 ka to 125 ka). Some of the models show similar behaviour. For example, as expected, different versions ~~flavours~~ of a model show similar behavior (see for example COSMOS_AWI, COSMOS_MPI, and KCM_Kiel, which share a common atmospheric component, ECHAM5). However, there are also strong similarities between HadCM3_Bris and COSMOS_MPI at 125 ka, and between MIROC_Tokyo and CCSM3_NCAR at 125 ka. Perhaps surprisingly, CCSM3_NCAR and CCSM3_Bremen at 125 ka are not very similar. This is probably ~~possibly~~ related to the higher resolution of CCSM3_NCAR (T85 compared with T31), and the use of dynamic vegetation in CCSM3_Bremen (see Table **??**). CCSM3_LLN appears to be more similar to CCSM3_Bremen than to CCSM3_NCAR. CCSM3 SUBSCRIPTNBLLN has the same T31 resolution as CCSM3 SUBSCRIPTNBBremen, but similar to CCSM3 SUBSCRIPTNBNCAR does not include dynamic vegetation, implying that in CCSM3 the resolution has more of an effect on the climate than the inclusion of dynamic vegetation. The LOVECLIM EMIC has a different response to many of the GCMs, with a greater Arctic warming (especially at 127 ka), and reduced cooling in the Sahel. However, it is interesting to note that although this cooling is absent in the surface air temperature response, it is present in the surface temperature response (not shown). CLIMBER_LSCE also exhibits different behaviour, with a lack of geographical structure. Amongst the GCMs, the IPSL_CM4 model is an outlier in that it does not exhibit cooling in the Sahel at 126 ka. Possible reasons for these differences are discussed later in the context of the DJF and JJA changes. One point to note is that the length of the different LIG simulations could be playing a role; for example, Herold et al. (2012, QSR) show that the Nordic Sea cooling in CCSM3_LLN is only manifested after 800 yr of simulation. Other inconsistencies may be due to models using differing dates of vernal equinox or calendar definitions (**?**).

4.1.2 Ensemble mean response

It is also instructive to examine the ensemble mean response. In order to include variations between different models, and temporal variability through the LIG, we construct the ensemble mean as ~~Because of the similar climate response in the different snapshots, it is possible to treat all time periods independently when constructing an ensemble . As such, our LIGensemble consists of~~ a straightforward average of all the simulations presented in Fig. **??**. This will weight higher those models which have more than one simulation, and treat different versions ~~flavours~~ of models as independent.

The model ensemble mean annual mean temperature change, LIG minus preindustrial (Fig. **??**a) is characterised by maximum warming at high latitudes, especially in the Arctic. However, there is disagreement amongst the models as to the sign of the change in the Southern Ocean and Antarctica. There is little temperature change in the tropics except for in the Indian and African monsoon regions, where there is a cooling.

The ensemble mean temperature change in DJF (Fig. **??**b) is more consistent across models. There is a warming in the Arctic Ocean, and a cooling over most of the rest of the globe, with maximum cooling occurring in the tropical regions. The models generally agree about the sign of the change, except in the region between warming and cooling in the Northern Hemisphere mid latitudes, and in the Southern Ocean. The large winter warming of the Arctic in response to insolation forcing was highlighted by **?** in the context of the LOVECLIM_LLN model, who related it to the “summer remnant effect”. Their analysis of the surface heat balance components shows that the excess of solar radiation over the Arctic during summer is transferred directly into downward ocean heat flux, and it enhances the melting of sea ice and increases the warming of the upper ocean preventing any important warming of the model surface atmospheric layer. The additional heat received by the upper ocean delays the formation of sea ice and reduces its thickness in winter. This reduction of the sea ice thermal insulation allows the ocean to release heat which finally leads to a significant warming of the surface atmospheric layer in winter. **?** ~~Otto-Bliesner et al. (2012)~~ also attribute the DJF Arctic warmth in the CCSM3_NCAR model to seasonal lags in the system associated with sea-ice; this region still feeling the effects of the preceding summer warming. This warming is not likely due to local insolation forcing (Fig. **??**), because the DJF Arctic signal is weak owing to this being polar night in both LIG and modern, and the CO_2 contribution is relatively ~~realtively~~ small.

The cooler LIG temperatures at other latitudes can be related to the insolation forcing, which is negative relative to preindustrial in DJF at all latitudes south of 65° N. The maximum cooling occurs in the ensemble mean in monsoon regions; however, the cause of this is different to cooling in JJA in these regions, because in DJF there is also a decrease in precipitation compared with preindustrial. Little previous work has focused on this DJF ~~focussed on this~~ monsoon-region cooling, but it is consistent with an increase in north-easterly winds in the Sahara seen in HadCM3_Bris (not shown), advecting relatively cold air from the Eurasian continental interior, and associated with a modelled increase in DJF sea level pressure across much of North Africa. This is also consistent with the fact that this maximum in cooling is not as strong in the CLIMBER model (not shown) – the relatively simple CLIMBER ~~statistical-dynamical~~ atmosphere is unlikely to capture these dynamical changes in the tropics.

The ensemble mean temperature change in JJA (Fig. **??**c) exhibits warming in most regions, apart from the subtropical Southern Hemisphere oceans, and the monsoon regions. There is also good agreement amongst the models in most regions of warming. The maximum warming occurs in the Northern Hemisphere mid latitude continental regions, especially in central Eurasia. The general warming is consistent with the seasonal insolation signal, including the fact that in the Arctic the signal is slightly weaker, due to a negative forcing in August (Fig. **??**). The maximum warming over continents as opposed to over oceans is consistent with the lower heat capacity of the terrestrial surface, and reduced potential for latent cooling. Many models exhibit JJA cooling in the monsoon regions. Previous studies (e.g. **?**) have attributed this to enhanced monsoon circulation, driven by greater land-sea contrasts, leading to enhanced precipitation, cloud cover and evapotranspiration. The models which do not simulate cooling in JJA are CLIMBER, LOVECLIM, and IPSL_CM4. For CLIMBER, the signal is large enough that it should be visible even at the low model resolution, which indicates the relatively simple ~~simple statistical-dynamical (SD)~~ atmosphere may be responsible. For LOVECLIM, clouds are prescribed in all LIG simulations to be the same as modern (**?**), and so the summer monsoon cooling feedback is weaker (but still present to an extent due to increased precipitation, **?**). For IPSL_CM4, this is due to a more limited response of monsoon precipitation in this model (Pascale Braconnot, personal communication, July 2012).

It can be seen that the lack of clear signal in the annual mean response over the Southern Ocean and Antarctica is due to the balancing of seasonal positive and negative forcings. The annual mean cooling in the tropics is due to dominant DJF cooling, the annual mean warming in Northern Hemisphere high latitudes is due to dominant JJA warming, and the annual mean Arctic warming is due to year-round warming.

The warm-month mean (WMM, the temperature in the warmest month, at any one gridcell) temperature change (Fig. **??**d) exhibits warming in the Northern Hemisphere~~Hemipshere~~, and cooling in the Southern Hemisphere. This is effectively an amalgam of the DJF signal in the Southern Hemisphere, and a JJA signal in the Northern Hemisphere. In this case, the only major region of equivocal sign is in the tropics.

4.2 Model-data comparison

The terrestrial model-data comparison as a function of latitude for the annual mean surface air temperature is shown in Fig. **??**a. Although the very fundamental pattern of maximum warming at mid and high latitudes is present in both model simulations ~~models~~ and **?** data, it is clear that the ensemble mean fails to capture the same magnitude of change as in the data. In particular, the data indicates warming of up to 15 °C in Eurasia at the LIG, but the ensemble mean is only about 2 °C. Also in Antarctica, the data is interpreted as indicating warmth of up to 5 °C, whereas the model simulations ~~models~~ are less than 1 °C. The agreement is actually worse than this considering that the data represents anomalies relative to modern (1961–1990), whereas the model simulations are relative to the (cooler) preindustrial. This mismatch is highlighted in Fig. **??**b, which shows a point-by-point comparison of the ensemble mean and the data (see Supplementary Information for this figure for each individual simulation from each model). It can be informative to quantify the degree of model-data agreement by defining a "skill score", σ. In this case, we use a very simple measure of skill, σ, equal to the RMS difference between the proxy values (T_p) and the modelled values (T_m) at the same location, so that

$$\sigma = \frac{1}{N}\sqrt{\sum (T_\mathrm{m} - T_\mathrm{p})^2} \qquad (1)$$

where N is the number of data points ($N = 100$ in the case of terrestrial data, and $N = 162$ in the case of SSTs). The skill score is not ideal, due to uneven data coverage, including some regions with no data. As such, the metric gives high weighting to model errors in the Mediterranean region, where there is the greatest density of data. However, it does give a first order estimate of the models' ability to replicate the data.

For the ensemble mean, $\sigma = $ 3.6~~3.5~~ °C. This lies approximately at the center of the distribution of all the model σ's – the lowest ("best", but note caveats above) being MIROC_Tokyo at 125 k, with $\sigma = 3.0$ °C, and the highest being NORESM~~CCSM3~~_BCCR at 130~~remen at 125~~ k, with $\sigma = $ 4.6~~4.2~~ °C. It is interesting to note that for three ~~two~~ of the models (CCSM3_Bremen, ~~and~~ CCSM3_LLN and NORESM SUBSCRIPTNBBCCR), the LIG σ is actually worse (higher) than the equivalent σ obtained by assuming that the LIG climate is identical to that of preindustrial ($\sigma = 4.0$ °C).

It is possible that some of the proxies used in the compilation of **?** may be more indicative of changes in seasonal temperature, as opposed to annual mean temperature. If this were the case, then better agreement may be achieved by comparing the proxy temperatures with seasonal modelled changes. In particular, it is possible that some proxies may be biased towards warm growth-season changes. The equivalent plots as for Fig. **??** are shown for DJF, JJA, and the warm-month-mean (WMM), in Fig. **??**. The JJA and WMM simulations are "better" in the sense that they have a wider range of anomalies (i.e. the greatest warming is larger for the WMM than for the annual mean), which is closer to the range of the data, but they are "worse" in that they all have a higher value of σ. As such, considering possible seasonal biases in the proxies does not substantially improve the model-data agreement.

? also provide a compilation of LIG SSTs. The SST data is less geographically biased than the terrestrial data, but there is still an over-sampling of data in the Atlantic, coastal, and upwelling regions. We compare these with the modelled SSTs (as opposed to surface air temperatures in the previous sections) in Fig. **??**. Note that because the CLIMBER model has a 2-D ocean, for that model we use the global surface air temperatures in place of SST. Many of the findings from the analysis of surface air temperature are supported by the SST analysis. Namely, that the model ensemble does not exhibit the same range of warming as the proxy data, and that this is also the case for each individual model within the ensemble. In particular, the model simulations do not show as much warming ~~models do not warm as much~~ as the data in the north Atlantic, and on the northward margins of the Antarctic Circumpolar Current. The σ for the ensemble mean SST ~~SSTs~~ is 2.6 °C. In a similar way as for surface air temperatures, looking at the JJA or WMM temperature does improve the range of modelled warming, but does not have a substantial effect on the σ values.

5 Discussion

There are several ways in which the model simulations, and the ensemble, presented in this paper could be improved.

Firstly, an attempt could be made to use more realistic boundary conditions. In particular, evidence for relatively high LIG sea level (e.g. **?**) suggests that a reduced Greenland and/or West Antarctic ice sheet would be more realistic than the unchanged-from-modern ice sheets used here, and could result in an improved model-data agreement in the North Atlantic SSTs. Evidence for shifts in Arctic treelines (see Section **??**) suggests that a modified vegetation could be imposed in the models, or more widespread use made of dynamic vegetation models. The combination of vegetation with ocean and sea-ice feedbacks could transform the seasonal insolation forcing into an stronger annual mean warming (**?**). MIROC_Tokyo has a particularly strong JJA response in terrestrial Northern Hemisphere high latitudes compared with many other models, which may be related to its use of dynamic vegetation; however, other models with dynamic vegetation (CCSM3_Bremen, COSMOS_MPI, and LOVECLIM_LLN) do not have this same response (Fig. **??**).

Secondly, many of the models included in this intercomparison are not “state-of-the-art”. It is possible that higher resolution, improved atmospheric and ocean dynamics, more complex parameterisations, and additional “Earth system” processes, could lead to better simulations of the LIG. Such simulations would be computationally challenging, but the LIG has the advantage over some other time periods, such as the LGM and Pliocene, in that the boundary conditions are very easy to implement (if modern ice sheets are assumed, as has been done for all the simulations in this paper). An indication of the impact of increasing complexity can be obtained by comparing the response from the EMICs in the ensemble (LOVECLIM and CLIMBER) with that of the GCMs (see Figure **??**). This shows a more complex response from the GCMs, even considering the difference in resolution; the cooling in the African monsoon region is not seen as strongly in the EMICs as in the GCMs (at least partly related to the simplified rerpesentation of clouds in the EMICs). However, the warming in the Arctic is stronger in the EMICs than in the GCMs. Previous comparisons of EMICs with GCMs (**?**)

Thirdly, in order to examine more closely the range of climates across the interglaciation, and to make the most of the many sites which have well dated time-varying proxy records, it is desirable ~~desireable~~ to carry out transient simulations across the LIG. As computational power increases, such simulations become more feasible, although not necessarily with the very latest models. Some such simulations exist already, mostly with models of intermediate complexity, low resolution GCMs, or with accelerated boundary conditions. A companion paper to this one, Bakker et al. (submitted to Climate of the Past, 2012) is carrying out an initial review of existing LIG transient simulations. Evaluation of these simulations with transient proxy records is an exciting and challenging prospect.

There are also ways in which the data synthesis could be modified, in the context of making model-data comparison more robust.

Because the LIG climate signal is driven primarily by a seasonal forcing, the annual mean response of the models is relatively small, and model-dependent, as shown in Fig. **??**a. But, the seasonal response is large. As such, a synthesis of seasonal, or WMM/CMM proxy indicators would be much more useful than annual mean indicators for evaluating models.

On a similar note, proxy indicators are perhaps most useful when they show a large signal, as the signal-to-noise ratio will likely be higher. Figure **??**b–d show clearly the regions of large modelled seasonal signals. Although there are some data located in northern Eurasia in the **?** compilation, there are none in central North America, or the Africa and Eurasian monsoon regions, where there are strong summer and winter modelled signals respectively. This is a similar approach to that suggested by **?** in the context of the Miocene.

Probably the most important improvement would be an assessment of the uncertainties in the various proxy estimates. A single value from a proxy, without an error estimate, is almost meaningless in the context of model-data comparison. For example, a model-data disagreement of 5 °C, on a proxy with an uncertainty estimate of 5 °C, has a very different implication to a model-data disagreement of 2 °C, on a proxy with an uncertainty estimate of 0.5 °C. One way in which proxy uncertainty can be tested, is to aim for multi-proxy assessments ~~asssessments~~ at all sites. Such an approach can radically change the interpretation of proxy data, such as was found by the MARGO group for the LGM (**?**), and by the PRISM group for the Pliocene (**?**).

The LIG clearly has potential as a test-bed of climate models, due to its large seasonal signal, and relative abundance of proxies with sufficient age control. However, this paper has shown that there is still some way to go before its potential can be realised, both in the development of a robust proxy dataset, and in the use of state-of-the art models.

Future work should also look at other aspects of these and other model simulations, such as the hydrological cycle and ocean circulation. In addition, it would be very interesting to look at the response of the Greenland and West Antarctic ice sheets to a range of modelled climates; previous work in this field (e.g. **??**) has focused ~~focussed~~ on a single model and so ignored this potentially important aspect of uncertainty. The simulations here have implied that the CO_2 and other greenhouse gas contribution to LIG warmth is small compared to the seasonal astronomical signal, but this could be confirmed by carrying out sensitivity studies.

Finally, this work indicates that although other interglacials, such as MIS 7 to MIS 11, could also ~~could~~ be potentially useful targets for models (e.g. **?**), ~~but that,~~ in terms of model-data comparison ~~,~~ more benefit would probably be gained by improving aspects of the LIG data compilations first.

6 Conclusions

In this paper, we have assembled a set of climate model simulations of the Last Interglacial, spanning 12 models of varying complexity, and 5 time-slices. We have compared the temperature anomalies predicted by the models with those reconstructed by **?**.

The main findings are that:

– The annual mean signal from the ensemble is small, with robust changes largely limited to warming in the Arctic and cooling in the African and Indian monsoon regions.

– The seasonal signal is stronger and more robust, with clear JJA warming across the mid-high latitudes of the Northern Hemisphere, and DJF cooling globally except for warming in the Arctic, and equivocal signal in the Southern Ocean.

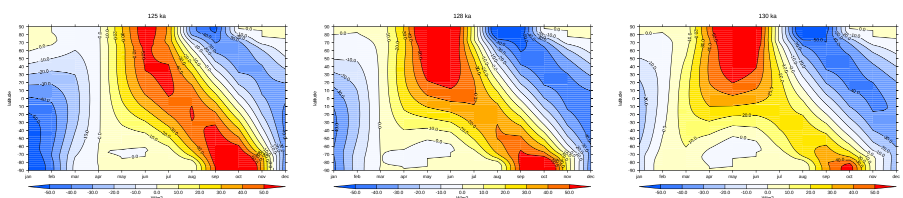
– There appears to be a difference in signal from the models of intermediate complexity compared with the GCMs (see Figure ??), which can not just be explained by resolution, but this should be confirmed with further analysis.

– The model simulations ~~models~~ and data do not show good agreement, for all individual models and for the ensemble. In particular, the large LIG ~~values of~~ annual mean temperature anomalies in the data are not replicated by the models.

– The range of seasonal warming in the model simulations is closer to that of the data, but there is still very little skill in the seasonal model predictions, with, in some cases, a better model-data agreement being obtained if it is assumed that the LIG were identical to modern.

– This study points the way to several improvements in both the modelling and data strategy, which could be employed to provide a more robust model-data comparison. On the data side this includes the incorporation of error bars in the proxy datasets, and inclusion of seasonal proxies in order to capture the largest signals. On the model side, this includes more studies on the role of vegetation, and ice sheet change and associated fresh water forcing.

Acknowledgements. This is Past4Future contribution no. 27. The research leading to some of these results has received funding from the European Union's Seventh Framework programme (FP7/2007-2013) under grant agreement no. 243908, "Past4Future. Climate change – Learning from the past climate". This paper arose from discussions held at the PMIP3 2nd General Meeting: Crewe Hall, Crewe, Cheshire, UK, 6–11 May 2012. DJL acknowledges funding from RCUK as a research fellow. The research of LLN group is supported by the European Research Council Advanced Grant EMIS (No. 227348 of the Programme 'Ideas'). Q. Z. Yin is supported by the Belgian National Fund for Scientific Research (F. R. S.-FNRS). ZSZ is supported by the Earth System Modeling project financed by Statoil.



(a) (b) (c)

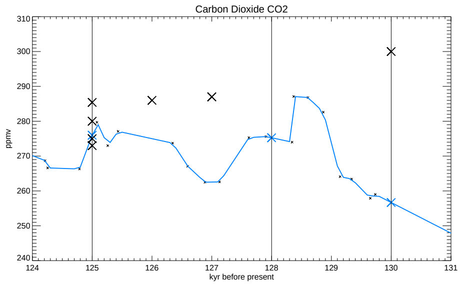
Fig. 1. Insolation at the top of the atmosphere [Wm^{-2}]for (a) 125 ka, (b) 128 ka and (c) 130 ka, relative to modern, as a function of month of the year and latitude, as calculated by the radiation code in HadCM3. The calculation assumes a fixed calendar, with vernal equinox on 21st March; as such, the anomalies in October in the Southern Hemisphere and September in the Northern Hemisphere are largely an artefact (**?**).

Table 1. Summary of models in this intercomparison. ‘Type’ refers to the atmospheric component of the model: General Circulation Model (GCM) or ~~,~~ Earth system Model of Intermediate Complexity (EMIC). ‘RMS’ gives the RMS ‘error’ of the preindustrial simulation surface air temperature ~~, or statistical-dynamical~~ (°C~~SD~~) relative to the NCEP climatology (see Figure **??**). Note that the RMS error is not area-weighted.

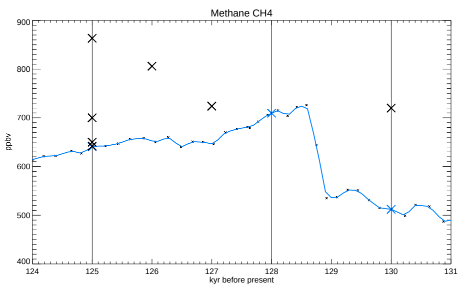
Model ~~name~~	Institution	Model name	Model reference	Type	Other	RMS
HadCM3 SUBSCRIPTNBBris	University of Bristol **?**	HadCM3 GCM	$3.75° \times 2.5°$	2.4 ~~n/a~~		
CCSM3 SUBSCRIPTNBBremen	MARUM, University of Bremen **?**	CCSM3 GCM	T31, land model hydrography improved compared to original CCSM3 release (**?**)	2.9		
CCSM3 SUBSCRIPTNBLLN	Louvain la Neuve **?**	CCSM3 GCM	T31	3.9		
CCSM3 SUBSCRIPTNBNCAR	NCAR **?**	CCSM3 GCM	T85, land model hydrography improved compared to original CCSM3 release (**?**)	2.4 ~~T42~~		
COSMOS SUBSCRIPTNBAWI	AWI **?**	COSMOS GCM	T31	2.8 ~~n/a~~		
COSMOS SUBSCRIPTNBMPI	MPI-M **?**	COSMOS GCM	T31	2.9 ~~n/a~~		
KCM SUBSCRIPTNBKiel	CAU-GEOMAR, Kiel **??**	KCM GCM	T31	3.9 ~~n/a~~		
LOVECLIM SUBSCRIPTNBAms	Amsterdam **?**	LOVECLIM EMIC	T21	4.2 ~~n/a~~		
LOVECLIM SUBSCRIPTNBLLN	Louvain la Neuve ~~Louvain-la-Neuve~~ **?**	LOVECLIM EMIC	T21	4.7 ~~n/a~~		
MIROC SUBSCRIPTNBTokyo	University of Tokyo **?**	MIROC GCM	n/a	2.5		
CLIMBER SUBSCRIPTNBLSCE ~~??~~	LSCE **?** ~~SD~~	CLIMBER EMIC	CLIMBER-2, version AOV PSI0	4.9		
IPSLCM4 SUBSCRIPTNBLSCE	LSCE **?**	IPSL GCM	n/a	2.8		

Table 2. Summary of simulations in this intercomparison. For the greenhouse gas concentrations, a '*' indicates that the value is that specified by PMIP3. CO_2 is in units of ppmv, CH_4 and N_2O are in units of ppbv. The LIG skill score, $sigma$, is relative to the terrestrial data of ?, and is defined in Equation ??. Note that CO_2 is the only greenhouse gas considered by CLIMBER.

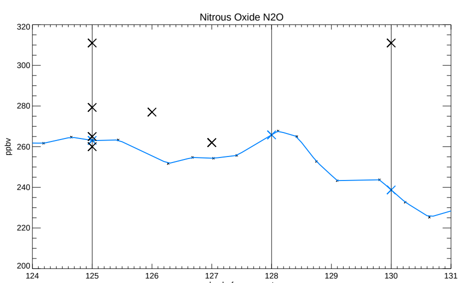
Model name	Snapshot	CO_2	CH_4	N_2O	length	notes	publication	σ
HadCM3_Bris	0	280	760	270	>1000	n/a	n/a	4.0
	125	276*	640*	263*	550	n/a	n/a	3.7
	128	275*	709*	266*	550	n/a	n/a	3.7
	130	257*	512*	239*	550	n/a	n/a	4.1
CCSM3_Bremen	0	280	760	270	1000	dynamic veg	n/a	4.0
	125	276*	640*	263*	400	dynamic veg	n/a	4.2
CCSM3_LLN	0	280	760	270	1300	n/a	Herold et al (submitted, QSR)	4.0
	127	287	724	262	1000	n/a	Herold et al (submitted, QSR)	4.0
CCSM3_NCAR	0	289	901	281	950	sol const 1365 W/m2	**?**	4.0~~Otto-Bliesner et al. (submitted, Phil. Trans. Roy. Soc.)~~
	125	273	642	311	350	sol const 1367 W/m2	**?**	3.5~~Otto-Bliesner et al. (submitted, Phil. Trans. Roy. Soc.)~~
	130	300	720	311	350	sol const 1367 W/m2	**?**	3.4~~Otto-Bliesner et al. (submitted, Phil. Trans. Roy. Soc.)~~
COSMOS_AWI	0	280	760	270	3000	dynamic veg	**?**	4.0
	130	257*	512*	239*	1000	same veg as 0k	n/a	3.6
COSMOS_MPI	0	280	700	265	>1000	dynamic veg	**?**	4.0
	125	280	700	265	>1000	dynamic veg	**?**	3.2
KCM_Kiel	0	286	806	277	1000	n/a	**?**	4.0
	126	286	806	277	1000	n/a	**??**	3.8
LOVECLIM_Ams	0	280	760	270	>1000	n/a	n/a	4.0
	125	276*	640*	263*	2000	n/a	n/a	3.4
	130	257*	512*	239*	2000	n/a	n/a	3.4
LOVECLIM_LLN	0	280	760	270	1000	dynamic veg	**?**	



(a)



(b)



(c)

Fig. 2. Atmospheric concentrations of (a) CO_2, (b) CH_4 and (c) N_2O through the Last Interglaciation. Vertical lines show the PMIP-defined snapshots of 125 ka, 128 ka, and 130 ka. Small black crosses show the raw gas concentrations from the Dome C ice core: **?** for CO_2 (although note that this is a composite record), **?** for CH_4 and **?** for N_2O. Blue line shows this raw data interpolated onto a 100-year resolution. Large blue crosses show the PMIP3 gas concentrations at the time of the snapshots. Large black crosses show the greenhouse gas concentrations used by those groups which did not use the PMIP3 guidelines.

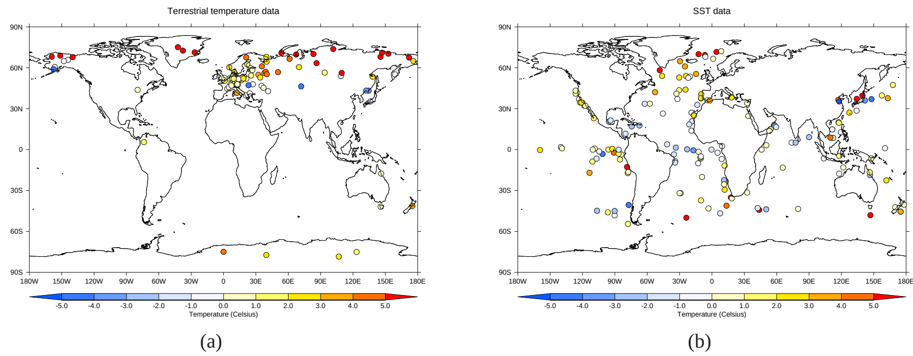


Fig. 3. Data compilation of **?**, showing the LIG temperature anomaly relative to modern (1961-1990) for (a) terrestrial temperatures (100 sites) and (b) SSTs (162 sites).

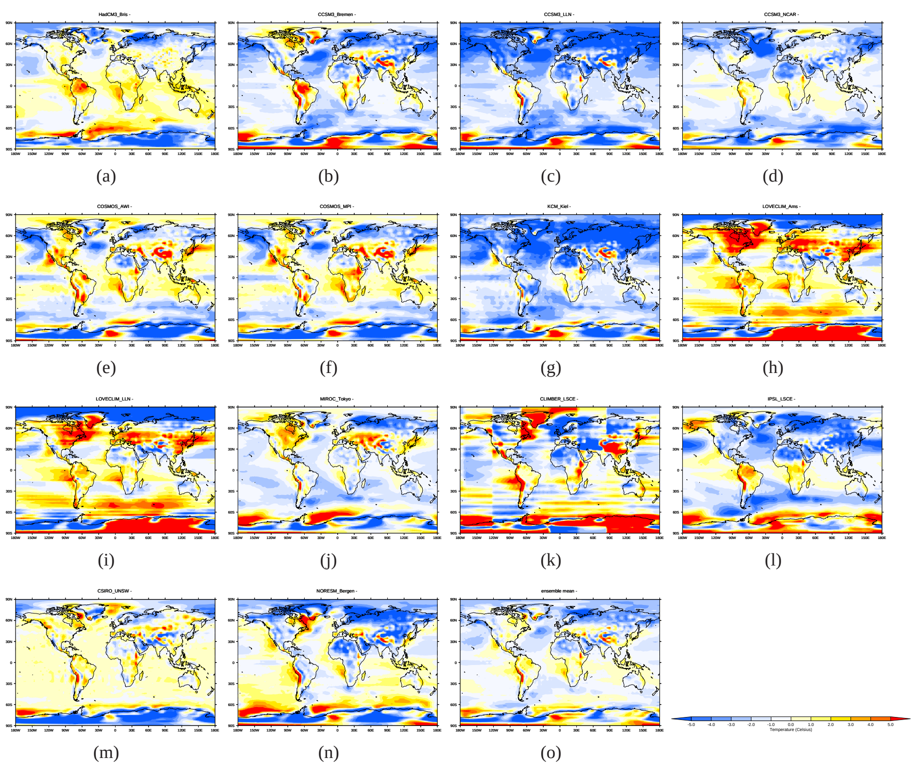


Fig. 4. 'Error' in the preindustrial control simulation of each model, relative to NCEP reanalyses (**?**), for surface air temperature. (a) HadCM3_Bris, (b) CCSM3_Bremen, (c) CCSM3_LLN, (d) CCSM3_NCAR, (e) COSMOS_AWI, (f) COSMOS_MPI, (g) KCM_Kiel, (h) LOVECLIM_Ams, (i) LOVECLIM_LLN, (j) MIROC_Tokyo, (k) CLIMBER_LSCE, (l) IPSL_LSCE, (m) CSIRO SUBSCRIPTNBUNSW, (n) NORESM SUBSCRIPTNBBCCR, (o) ensemble mean of models (a)-(n~~l~~). Note that the observations are for modern (1948-1987), whereas the models are designed to represent preindustrial. All data is interpolated onto a 96×73 resolution before calculating the difference, model minus data. The RMS values for each model simulation are given in Table **??**.

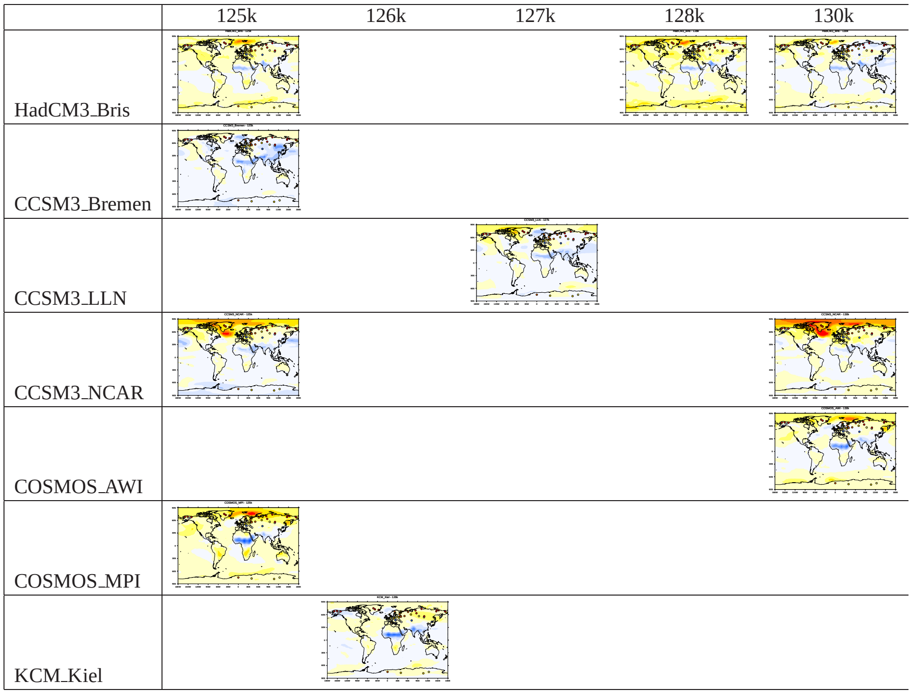


Fig. 5. Simulated annual mean surface air temperature change, LIG minus preindustrial, for each model and each snapshot carried out. Also shown are the terrestrial data points of **?**. [Continued on next page].

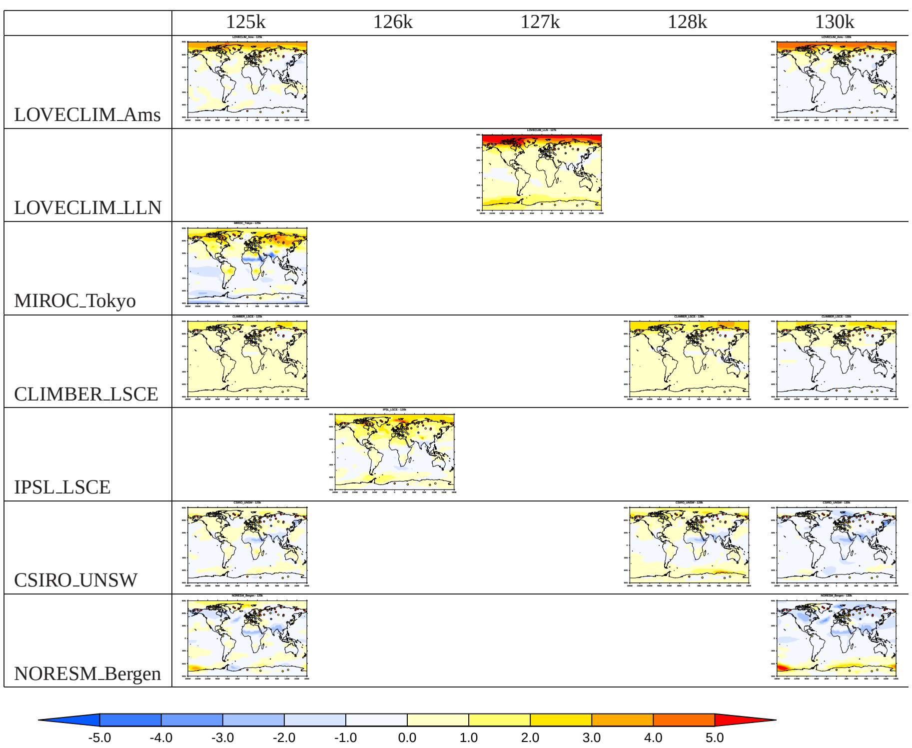
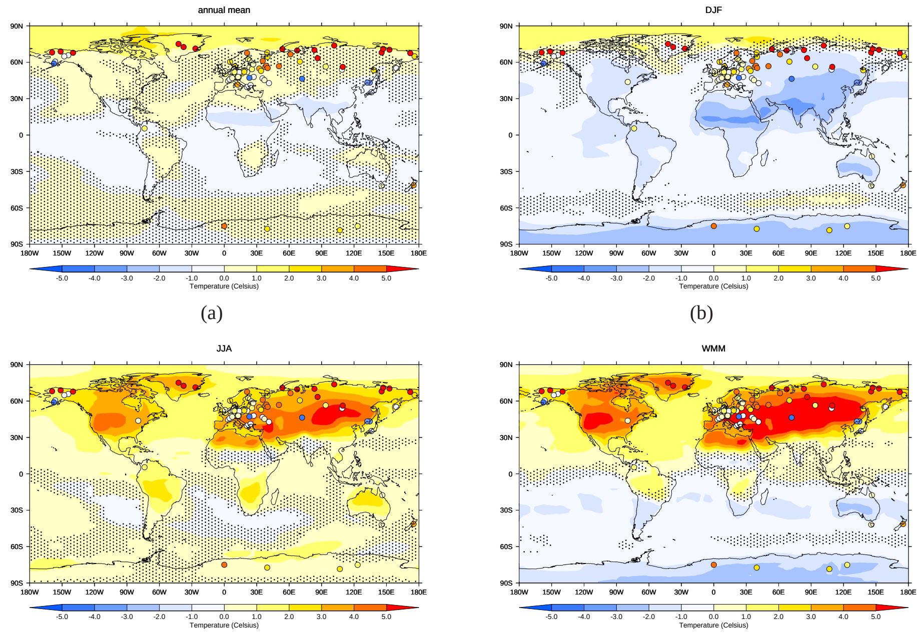
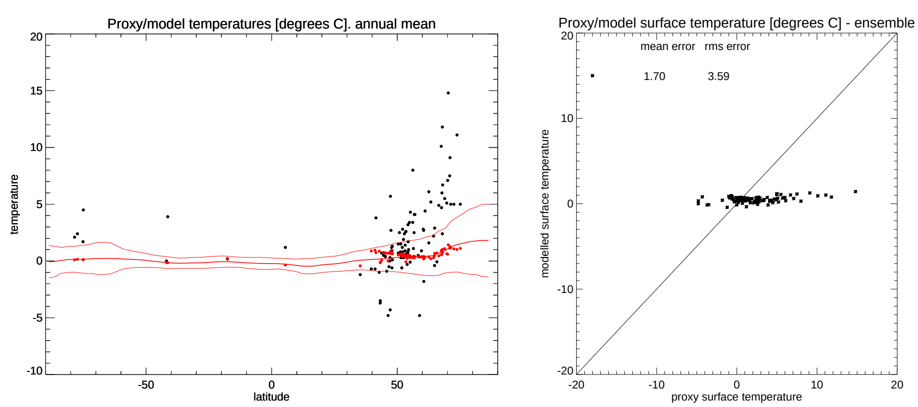


Fig. 5. [Continued from previous page] Simulated annual mean surface air temperature change, LIG minus preindustrial, for each model and each snapshot carried out. Also shown are the terrestrial data points of **?**. .



(a) (b) (c) (d)

Fig. 6. Simulated surface air temperature change, LIG minus preindustrial, for the model ensemble. (a) annual mean, (b) DJF, (c) JJA, and (d) warm month mean (WMM). Stippled regions show regions where less than 70% of the model simulations agree on the sign of the temperature change. Also shown are the terrestrial data points of **?**.



(a) (b)

Fig. 7. Comparison of ensemble mean surface air temperatures with data from **?**. (a) Latitudinal distribution of proxy data (black dots), compared with the ensemble mean model (red dots, from the same locations as the proxy data), with the zonal model ensemble mean (thick red line), and ± 1 standard deviation of the zonal model ensemble mean (thin red lines). (b) Ensemble mean vs. proxy data for each datapoint. All units are $^{\circ}$C.

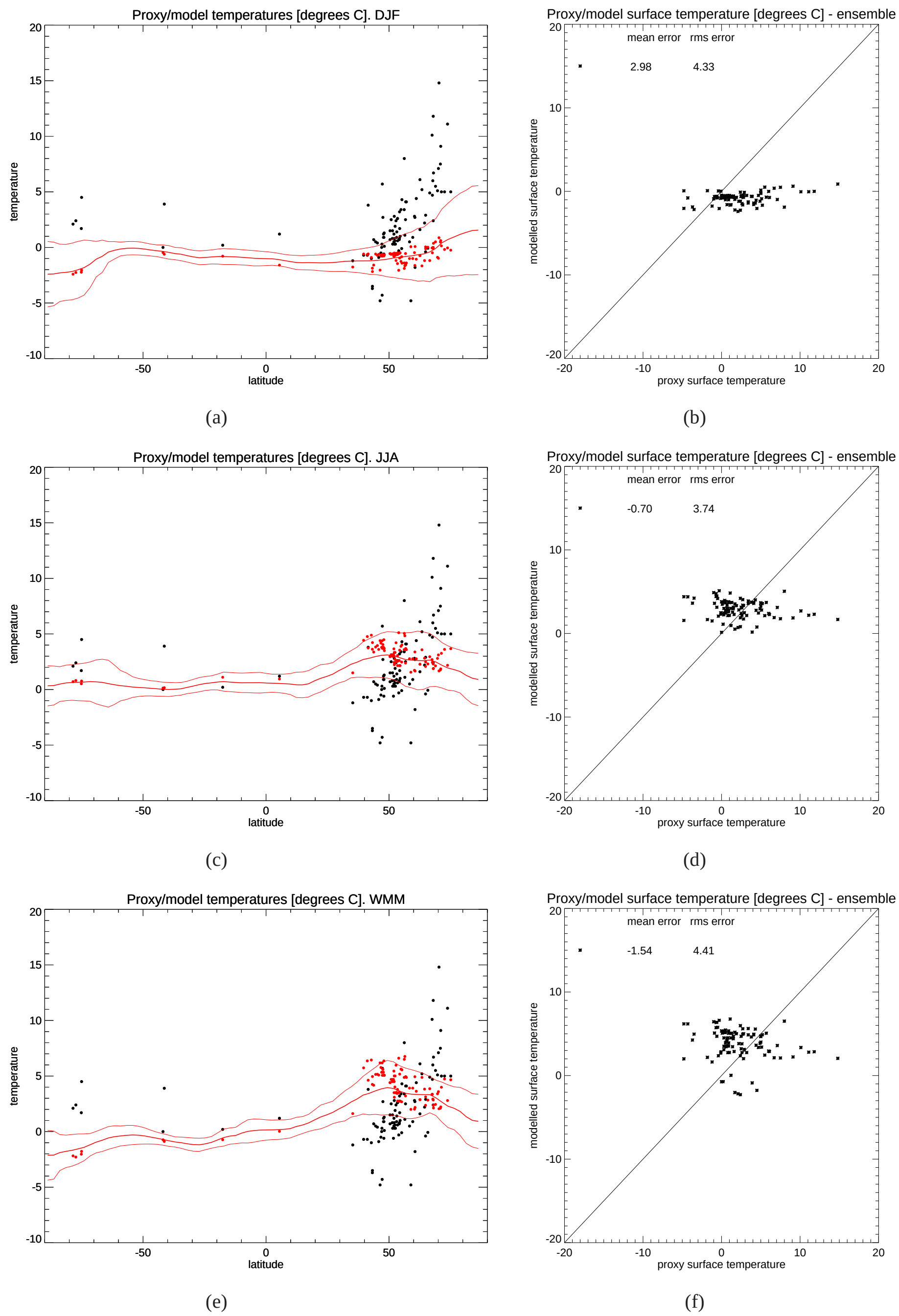
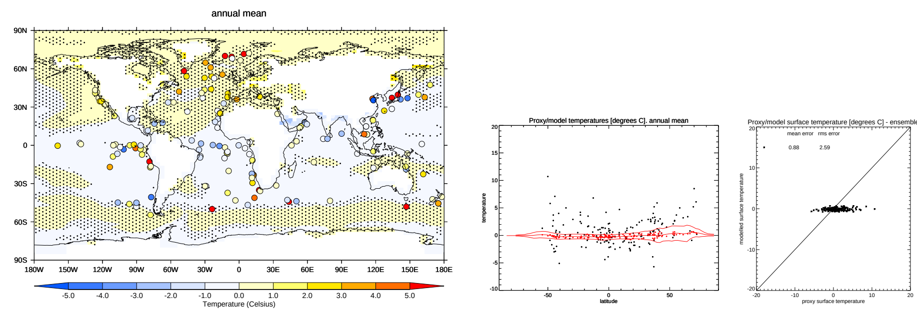
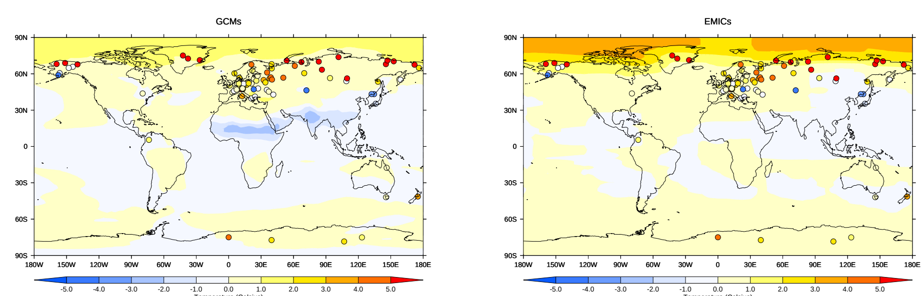


Fig. 8. Comparison of ensemble mean surface air temperatures with data from **?**. (a,c,e) Latitudinal distribution of data (black dots), compared with the ensemble mean model (red dots, from the same locations as the proxy data), with the zonal model ensemble mean (thick red line), and ± 1 standard deviation of the zonal model ensemble mean (thin red lines). (b,d,f) Ensemble mean vs. proxy data for each datapoint. (a,b) are for DJF, (c,d) are for JJA, and (e,f) are for WMM. All units are $^{\circ}$C.



(a) (b) (c)

Fig. 9. (a) Simulated annual mean SST change, LIG minus preindustrial, for the model ensemble. Stippled regions show area where less than 70% of the model simulations agree on the sign of the temperature change. Also shown are the ocean data points of **?**. (b,c) Comparison of annual mean SSTs with data from **?**. (b) Latitudinal distribution of data (black dots), compared with the ensemble mean model (red dots, from the same locations as the proxy data), with the zonal model ensemble mean (thick red line), and ±1 standard deviation of the zonal model ensemble mean (thin red lines). (c) Ensemble mean vs. proxy data for each datapoint. All units are °C.



(a) (b)

Fig. 10. Simulated surface air temperature change, LIG minus preindustrial, for (a) the GCMs in the ensemble, (b) the EMICs in the ensemble. Also shown are the terrestrial data points of **?**.