Clim. Past Discuss., 8, C1946–C1952, 2012
www.clim-past-discuss.net/8/C1946/2012/

Climate
of the Past
Discussions

Interactive comment on “Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid Holocene” *by* J. C. Hargreaves et al.

Anonymous Referee #1

Received and published: 17 October 2012

The manuscript of Hargreaves et al., presents a systematic data model comparison for the Last Glacial Maximum and mid-Holocene time slices. It builds on earlier studies of the author, which analyzed the consistency of the PMIP and PMIP2 climate model ensembles with the MARGO LGM data. In the new study, the model ensemble, the dataset and the statistical metrics were extended and most importantly, an analysis of the Mid Holocene was included. I welcome the effort of the authors in systematically comparing paleo-proxy datasets and climate model simulations. The paper is in the focus of Climate of the Past and I recommend it for publication after some revisions were made.

In the moment, at least from my viewpoint, the paper is unnecessary difficult to read and the presentation of the results could be improved. I added some suggestions in the line-by-line comments. I would further recommend including all the results either in figures or tables (summary statistics), if needed by adding an appendix. In the moment, most Holocene results (e.g. the GHOST comparison) are not shown.

Major points:

Holocene dataset and data uncertainty:

For the Holocene terrestrial dataset the error estimates seem unrealistic. The given minimum error of 0.04C (Fig.2) would be smaller than the error of a very good thermometer. I understand that a reanalysis of the terrestrial temperature proxy error is beyond the scope of the paper, but at least a discussion of this weakness and its potential effects on the results has to be included. A more realistic treatment of terrestrial Holocene temperature errors can be found for example in Zhang et al., 2010.

For the marine GHOST dataset, it is unclear how the anomaly between 6 and 0ka was formed. As the time series are highly variable (because of internal climate variability and "proxy" noise), taking the difference between the mean of two short timeslices will give highly variable estimates. This is very different than for the LGM in which the LGM value represents the mean over several thousand years and is taken relative to a well determined modern instrumental estimate. I would recommend checking if this uncertainty is already included in the 2C assumption.

My worries are that for the Holocene, the data error has a similar magnitude than the signal term, and the reliability test mainly tests if the error estimate was good (which it is likely not the case). I would therefore recommend a sensitivity study if the models would pass the reliability test if the data uncertainty would be more realistic.

For the Holocene, other "quantitative" studies found at least a significant pattern correlation (e.g. Schneider et al., 2010, Lohmann et al., 2012). How can this be reconciled with the results of this paper? At least the Taylor diagram for the GHOST dataset (which is not shown in the moment) should show some correlation and a significant correlation should also display as some skill relative to a uniform Holocene warming null hypothesis for the GHOST dataset.

Treatment of effective spatial degrees of freedom unclear: (Page 3498 in the manuscript)

1.) I have difficulties to understand how the ESDOF were estimated. The paper cites Annan and Hargreaves (2011) but this reference seems to use an estimate of decorrelation distance, whereas the present manuscript cites an EOF approach. For both, EOF's or decorrelation distance, one would need a time varying field. Here, only two scalar fields, one for preindustrial and one LGM fields are available. The alternative of estimating the ESDOF from the time varying equilibrium LGM simulation would be clearly wrong as the ESDOF depends on the time-scale (weather varies from place to place whereas glacials seem to be nearly synchronous globally).

2.) Thus for the analyzed timescale of changes between preindustrial, Holocene and LGM I could well imagine a very small ESDOF. How would this affect the results? If I understand it well, the smaller the ESDOF's, the easier it is to pass the rank histogram test of reliability but would this not also imply that we more often pass the test without being reliable (as passing or not passing gets more random)?

3.) What is the role of the observational error. To account for the observational error, random deviates (Line 6, 3489) are added to the model simulations. I assume iid normal distributed random variables? If the errors are a significant part of the variability, wouldn't this increase the ESDOF again?

Missing discussion:

I'm missing a discussion section in which the results (especially the non-existing skill of Holocene simulations) are tied into the literature and hypotheses are shortly discussed.

This includes the data error assumptions, the wrong attribution of the proxy data (they might represent another season than annual mean), an underestimation of the internal variability in the model simulations or an underestimation of the sensitivity to external forcing.

Line by line comments

3482:

Line 22: "representation error" explain or use other word Line 22: what about the possibility that internal climate variability is underestimated in the models?

Line 25: as the audience are largely "paleo people", better define what long-term means here Line 26: define GHG when it is first used

3483:

Line 6: believable; is there a better word? maybe reliable although I understand that reliable also has a specific statistical meaning.

Line 12, 23, 28, next page line 5 and the whole remaining text. The text would be easier to read if the explanations in parentheses could be either omitted or included in the text.

Line 14: good point ...but one also has to ensure that the performance of the models was not used in the construction of the forcing datasets... e.g. on the last millenium timescale I would suspect that this might happen.

3484:

Line 1: "Intercomparison" Line 10: "and global mean insolation forcing is rather small".

Line 11: this is true at least in the climate model world, in the data/proxy world, also other region experienced strong changes; I would therefore propose to either remove this sentence or extend the description of MH climate changes + including more references.

Line 19: again, avoiding parentheses would improve reading Line 23: remove second "we" Line 22: replace comma by full stop before "Secondly" Line 26: typo "statistics" Line 28" omit "In this paper"

Page 3485 The description of models (AOGCMs, AOVGCMs...), Table 1 etc. applies for both LGM and MH and is confusing under the headline Last Glacial maximum.. Therefore, either combine LGM and MH into one section or separate into three sections.

Line 26: "from centres, which have contributed more than one model": this is confusing at this point of the text; 1.) are the "problem" really the centres or are these just very similar models? Was this considered or done? (Later = Page 3494 one can see that it was done as one specific case).

3486

The data section is hard to follow. Maybe add a table in which data sources, references, baseline (against core-rop or modern instrumental data), error estimation method, #data points left for comparison, season definition. Try to avoid the use of too many "shortcuts" H11, B11. e.g. B11 is only used two times and can therefore be just left as pollen dataset (Bartlein...).

Line 4:" remove "plus"

Line 23: grid point area?

Line 29: H11 was defined before but the reader has likely forgotten it. I propose to fully write it out once more here. Unclear what "scaled by 1C" means.

Page 3487:

Line 9: this is not exactly true; 1.) there is a very small change in global annual mean (by eccentricity)... so better write ... changes in global annual mean ... are negligible 2.) not the global mean but the regional changes matter... so better argue that even

regionally the annual mean forcing changes are small and therefore, only a small signal is expected

Line 12: which representations; already describe them here Line 14 and Table 1: Is the AOGCM really just called ECHAM (as ECHAM is the name of the atmosphere model)? Line 16: better the full reference instead of B11

3489 Line 6: "random deviates of appropriate magnitude" what exactly? iid normal distributed values with a standard deviation given by the data uncertainty?

3492: Line 19-21: It is unclear here if this approach is used in the paper (which it is as one can see later) or just a possible method. Also write out the "simple calculations" or cite a reference in which they are described.

3493: Results and discussion are mixed together. 16-26 should be moved in a discussion section.

3494: The discussion about creating an ensemble and similar models should be moved in the Models and Data section and just the results should be included here.

3495: Line 21: "not only reproduce"

3497: please shift the discussion of which season (or season difference) to analyse and which ensemble to use in the data and model section and only focus on the results here. In the present version, it takes half a page to come to the results. Again, all results have either to be shown in figures, or summarized (by summary statistics) in a table. The title of this subsection could be "reliability, skill and Taylor diagrams" to be consistent with the LGM subsections.

3498: remove the discussion part here and move it to a separate discussion. Line 20: What means "high frequency mismatch"? What is "representation error" missing representativeness of the data for gridbox scale variability?

3499: What is meant by "high frequency noise": internal variability? Even "low frequency e.g. millenial internal variability could cause such a result.

Line 12: "finer spatial scales" Line 18 and 27: The study just showed that in the models, vegetation feedbacks didn't help... and also did not show evidence for vegetation feedbacks in the data; so either present evidence for the vegetation case in the main text, or remove this conclusion.

Figures:

Figure 2: a and b have different sizes; in b, the colors of 2 and 3 degree are hard to separate. A finer color scale would be more informative

References used in the Review:

Zhang, Q., HS Sundqvist, A. Moberg, H. Körnich, J. Nilsson, and K. Holmgren. 2010. “Climate Change Between the Mid and Late Holocene in Northern High latitudes–Part 2: Model-data Comparisons, Clim.” Past 6: 609–626.

Schneider, B., G. Leduc, and W. Park. 2010. “Disentangling Seasonal Signals in Holocene Climate Trends by Satellite-model-proxy Integration.” Paleoceanography 25 (4): PA4217.

Lohmann, G., M. Pfeiffer, T. Laepple, G. Leduc, and J.-H. Kim. 2012. “A Model-data Comparison of the Holocene Global Sea Surface Temperature Evolution.” Climate of the Past Discussions 8 (2) (March 29): 1005–1056. doi:10.5194/cpd-8-1005-2012.

Interactive comment on Clim. Past Discuss., 8, 3481, 2012.