Clim. Past Discuss.,
doi:10.5194/cp-2016-136-RC1, 2017

Climate of the Past
Discussions

Interactive comment on "Evaluation of PMIP2 and PMIP3 simulations of mid-Holocene climate in the Indo-Pacific, Australasian and Southern Ocean regions" *by* Duncan Ackerley et al.

Anonymous Referee #1

Received and published: 18 January 2017

General:

The article presents a (comprehensive) summary of proxy and model based results for investigating changes between the mid-Holocene and pre-industrial period for western Pacific and adjacent regions. The authors use a large number of proxy data to compare with temperature and precipitation, derived from the Paleoclimate Model Intercomparison Program (PMIP2 and PMIP3). In the second part of the manuscript they try to explain model-data (mis-) matches by dynamical processes.

The concept and outline of the article look promising but a closer inspection of the manuscript reveals that the authors present a collection of (raw) GCM model output data and compare them one-to-one with results based on published proxy literature.

Printer-friendly version

Discussion paper

In addition, they i) apply and interpret results in an inconsistent way and ii) for most cases don't use the variety of large-scale output from the climate models to test their hypothesized physical explanations. Moreover, the different sections are not very well organized into distinct results and discussion sections making it difficult to follow.

Although this kind of data-model comparison is often used, new methodological approaches are available in the context of model-data comparisons. These include Proxy system models (Dee et al. 2016 (incl. source code), application of numerical/statistical downscaling to improve regional precipitation characteristics (Fallah et al., 2015; Wagner et al., 2012), applying climate field reconstructions using a coherent network of proxy data (PAGES2k, 2015) and using pseudo proxy experiments in the virtual world of climate models to test the spatial representativeness of single proxies or their networks (Smerdon et al., 2012). These methods should be used and applied or at least been mentioned in the context of state-of-the-art proxy-model comparisons.

In all, I cannot suggest publication of the manuscript in its present form. Below I tried to include a number of suggestions how to improve the manuscript for a revised version to address my main concerns and to put the conclusions of manuscript in context of different sources of errors involved in direct proxy-(multi-)model comparisons.

Specific:

Abstract:

For my taste the first paragraph of the abstract should be included into the introduction because it is not related to any scientific advancement achieved with the manuscript.

Specific:

Introduction:

p.2 l.6: Proxies are not observationally based. At least the wording should be changed e.g. "Proxies give indications of past climate changes albeit with uncertainties associated to their individual recorder characteristics on changes on meteorological variables related to temperature/precipitation." same sentence: what's the rationale in comparing future simulations with reconstructions based on proxy data? I suggest re-phrasing the sentence that the comparisons may help to assess the general ability of models to simulate past climates which gives a certain degree of confidence in their ability to simulate potential future climate changes under specific forcing scenarios.

p.2 ll15 ff.: I suggest adding a few lines contrasting the merits/shortcomings of statistical versus numerical downscaling approaches. For instance the statistical downscaling assumes a constant large-scale/local scale relationship throughout time, whereas the numerical downscaling might account for those non-stationarities. Also numerical downscaling might be afflicted by the shortcoming that the driving GCM does not realistically simulate the large scale circulation and thus the RCM inherits errors from the GCM which in most cases cannot be compensated for.

p2. l. 23: Please add a short note what you understand by "effective precipitation" ?

p.2 ll. 28ff: How were those regions defined ? Are they purely related to some ad-hoc regionalization or have some statistical tools been applied to discriminate those temperate/hydrological different regions (e.g. Cluster analysis or some EOF-based approach). According to the figure 1 I assume however that the regions are rather related to some geographical lat-lon based criterion.

p. 3 l.3: In my opinion the authors should already indicate here that the simulated raw GCM precipitation is afflicted with a very high degree of uncertainty, especially over convective-active region and that current studies or compilations (e.g. IPCC) show large discrepancies between GCM-derived precipitation and observational/re-analysis/satellite derived precipitation. This would further motivate approaches to downscale GCM results with statistical/numerical approaches or/and use additional approaches (forward modelling) for proxy-model comparisons.

p.4 l. 1ff: I like the approach of developing seasonally resolved proxies. One might add here already some issues involved in addressing this point with the variety of proxies that are used over the study region (e.g. corals vs. speleothems vs. tree rings) and there pros/cons for seasonally resolved reconstructions.

p4. l 6ff: Maybe the authors can also state here that a congruence/disagreement between proxy and models does not necessary mean that in the real world this must be the case. Given the high degree of proxy and model uncertainty both "outcomes" can be right but for the wrong reason. This is just a hint towards a more general view of model-proxy comparisons and the ultimate need for i) a sound basis for comparisons and ii) the consideration of sound and robust dynamical mechanisms controlling the models' and proxies' mean state (and variability) for different time periods (which is already mentioned by the authors).

General comment on the Introduction:

Admittedly, given the manuscript is already very long I still think given the vast literature on the comparison between mid-Holocene and pre-industrial in various studies at least those publications that are most relevant in this context should be summarized in one paragraph and linked to the present study. This is also important to give the reader who is not so familiar with the topic a general idea about the basic climatic changes (e.g. intensified monsoon) and potential driving mechanisms (e.g. changes in earth' obliquity). In its present shape the intro is largely focused on an international consortium (OZ-Intimate) and a method used by Reeves et al. that might not provide the full spectrum on research that has been achieved in the last decade or so.

2 Data, Analysis and external forcings

2.1 Proxy data

p.5 How did the authors treat the different sources of proxies concerning their temporal resolution, their individual dating uncertainties, their seasonal biases and their different recorder characteristics in terms of meteorological sensitivity ?

2.2 Model simulations and boundary conditions

Printer-friendly version

Discussion paper

p.5 ll7 ff.: How many years are used for each time slice ? is it the same for all simulations ? The authors should also add a paragraph related to the large bandwidth of models they use for comparisons and that a separation of models showing a good/bad performance in the simulation of present day climate or specific variables is not possible in the subsequent multi model-mean analyses anymore.

p.5 ll 15-25: The authors might try to shorten this discussion to the most relevant insolation changes.

p. 6 For completeness the authors should add a sentence that the calendar was not changed between the MH and PI due to the precession changes. For periods with large gradients in insolation that might influence to a certain degree results based on the Gregorian calendar and the one based on a discrimination of seasons related to solely astronomical considerations (e.g. the periods between solstice and equinox).

p. 6 ll 9ff: Please add a short information on the target grid the models were interpolated.

2.3 Post 1750 C.E. datasets

p.6 l. 10: The authors might think of changing the wording of the header to "Re-analysis data sets". In this context I also suggest that authors add a few words on the reliability of the according data sets, especially the HadISST data set dating back to 1870 concerning the data availability, quality and coverage. Again it might be important to stress the fact that those data sets are not observational data. For the generation of re-analysis data sets, meteorological data are used in an assimilation scheme integrated into a comprehensive numerical model.

2.4 Analysis p 7 ll 15ff: Concerning the multi-model mean, I basically cannot support the approach because it represents a physically unrealistic state. Instead, an individual treatment of the models should be envisaged for all stages of the analyses. However, I acknowledge that the authors try to amalgamate and synthesize this kind of spread into “consensus” maps. A suggestion to further test the robustness of the model results is to use certain thresholds e.g. 2 times standard deviation for a more robust assessment not only in the sign of changes but also in terms of the magnitude. This is also motivated by statements further down that despite a certain agreement in the sign of the multi-model mean no statistical significant differences can be seen (e.g. p.9 ll 4ff).

3 Model and proxy synthesis

3.1 Annual mean

3.1.1 Surface temperature

p. 7 l. 24: Do the authors refer to surface temperature or 2m-(Near) surface temperatures ?

p 7 l 30ff: In the context of statistical significance it is important to at least mention the difference between statistical significance at a certain level and the physical relevance. I assume that the authors use the multi-model mean to carry out the t-test between MH and PI that is based on the number of years they use for their analysis. Given the usually low variance in tropical latitudes in the annual cycle also very small absolute changes in T2m become statistically significant just due to construction of the test algorithm. I further assume that the authors also don t take into account the serial correlation (SC) of the data that might also influence the level of statistical significance as the presence of SC can (profoundly) change the number of degrees of freedom (cf. effective sample size).

p. 9 ll. 6ff: The authors compare here their quantitative model estimates with empirical reconstructions – how robust are results based on the proxy records mentioned or at least what are uncertainties implicitly included in the studies cited ?

3.1.2 Precipitation

p9 ll. 19 ff: Taking raw precipitation from GCM output should be avoided. To circumvent this issue statistical and/or numerical downscaling methods should be applied first.

Here I also think it's important to have a look at the individual performance of the different GCMs. The situation gets even more complicated when authors compare their (continental-scale) regions with local information from proxy data. In contrast to temperature changes, precipitation may occur at spatial scales that are much more heterogeneous compared to temperature.

p9, 32 ff: This comment is related to my last one for temperature – how robust are results given uncertainty in the proxy and their ability to really record (solely) precipitation changes? The following sections are a repetition of the first one with seasonal focus but with very weak or vague statements concerning the representativeness of the respective proxy towards the respective season.

I suggest that the authors re-structure the whole section 3 into one section and discuss results only for robust proxy-model comparisons. Also the comparison between circulation and proxy should be restricted to model based results and if possible for individual models in their relation to the individual temperature/precipitation (hydrological) changes.

4 Mechanisms responsible for agreement and disagreement

4.1 Model-proxy agreement

4.1.1 Tropical north-west (TNW)

p. 16 ll. 8ff. The authors state that "but the modeled change is not statistically significantâĂŤalthough 65% of the models simulate higher rainfall at 6 ka." – I am wondering why there is surprise given the large spread and the very weak coherence in the sign of the models. If anything, I would call this a very weak tendency towards higher rainfall. So I think it's really important to stress the effect and meaning of statistical versus physical significance in the context of the multi-model mean differences for the various parameters.

In all, I'm not really convinced in the usefulness looking for statistical significance given

Printer-friendly version

Discussion paper

the issues mentioned above related to absolute differences and the effect of the number of degrees of freedom on the level of significance.

p. 18 l. 6f: This comment relates to the one on page 16: In this context the authors state that “only 62%” of the models agree. In an earlier statement they state this is two thirds agreement and exploit this number being significant in terms of the consensus. There should at least be a consistent nomenclature when the authors speak about model agreement being “large” or “rather evenly distributed”. There are more occurrences of the inconsistent use, particularly if the percentage lies between 60 and 70 %.

p. 18 l.9 ff: The authors try to explain the changes based on dynamical reasoning. Unfortunately, they don t use the simulated and model based output. Instead some vague mechanisms including effects of sea breeze is suggested which can by far not be simulated by the climate models the authors use.

4.1.3 The arid zone

I still don t understand why the authors don t use the models to test the physical consistence of the mechanisms they hypothesize based on proxy evidence. In addition, it’s very hard to distinguish the actual results section from an alone standing discussion section or information that might be important to know earlier e.g. the shortcomings of PMIP2 and PMIP3 presented on p 24 ll 4ff and other model deficiencies.

4.2 Model-proxy conflict

p25, l. 31ff: I find it a quite strong argument that just because the authors see “more confidence” in the proxies the models are being flagged wrong. The authors use a very large bandwidth in the complexity of models ranging from EMIC-type to comprehensive Earth System Models. I would expect a more detailed discussion if one could discriminate differences in between the models concerning their ability to simulate ENSO and if so, whether there is conceptual/technical reasoning, for instance related to the resolution of the according ocean model.

p 26 ll1 ff: Here again the authors begin a discussion about model deficiencies that should be placed elsewhere (e.g. in the general introduction) but not in the results section. Moreover, a distinction on the different complexity levels and resolution of the models should be clearly taken into account in their evaluation.

Future directions:

This section mirrors the (mostly) conservative nature of the authors' team to maintain their strategies for future directions, neglecting innovative methods to robustly and consistently compare data and models. For instance, it does not make any difference running longer simulations with the same models on their performance – their biases will still remain, also for ENSO dynamics. Even dynamical downscaling can only improve results over regions where there is some confidence in the driving model to realistically simulate the large-scale circulation at the lateral boundaries. Also the mentioned calibration of proxies will often fail because of the coarse temporal resolution of the proxy and it's ability to record to a high degree meteorological entities. Even if it would be nice to have this at hand, most of proxies presented in the study are not suited given i) the short length and availability of meteorological observations over the study region for calibration and ii) the complexity (and partly inability) for the inverse modelling of meteorological data based on proxy data.

References:

Bijan Fallah, Ulrich Cubasch, Kerstin Prömmel and Sahar Sodoudi (2015), A numerical model study on the behaviour of Asian summer monsoon and AMOC due to orographic forcing of Tibetan Plateau, Climate Dynamics, DOI:10.1007/s00382-015-2914-5.

Dee, S. G., N. J. Steiger, J. Emile-Geay, and G. J. Hakim (2016), On the utility of proxy system models for estimating climate states over the common era, J. Adv. Model. Earth Syst., 8, doi:10.1002/2016MS000677.

PAGES 2k Consortium (2013): Continental-scale temperature variability during the past two millennia (2013) Nature Geoscience, 6, 339-346, doi:10.1038/ngeo1797.

Smerdon, J.E. (2012), Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Climate Change, 3:63-77, doi:10.1002/wcc.149.

Wagner, S., Fast, I., and F. Kaspar (2012): Climatic changes and forcing mechanisms between 20th century and pre-industrial times over South America in regional model simulations, Climate of the Past , 8, 1599-1620, doi:10.5194/cp-8-1599-2012

Interactive comment on Clim. Past Discuss., doi:10.5194/cp-2016-136, 2016.