Climate
of the Past

Open Access

Discussions

Interactive comment on "Consistency of the multi-model CMIP5/PMIP3-past1000 ensemble" *by* O. Bothe et al.

Anonymous Referee #1

Received and published: 2 September 2013

General comments This paper represents a development in probabilistic methods for assessing consistency between the latest available multi-model millennium ensemble of simulations (CMIP5/PMIP3) and leading climate proxy reconstruction data. The wide range of methods employed in this analysis suggest that the distribution of the simulation ensemble is not typically consistent with a global temperature reconstruction, though there is some consistency in certain area-averaged regions and the results, to some extent, depend on the frequency band of the analysis (i.e. multi-centennial or decadal).

I found this paper to be well written (particularly in the introduction and methodological setup) and with some very interesting and valuable results. This research builds on recent previous work by the author in a logical way and draws attention

to important emerging problems. Namely, how to compare simulated and reconstructed climates, considering the wide range of errors and complexities there-in. The main conclusion from a probabilistic consistency perspective is perhaps the general over-disperse nature of the simulations, which also contain differences between sub-periods. The main conclusion from a climatological consistency perspective is perhaps that the reconstructed quantiles have different respective shifts depending on the sub-period, with the reconstruction's transition from the Medieval Warm Period to the Little Ice Age (MWP-LIA) being more pronounced. The paper sensibly concludes that stronger conclusions regarding consistency deviations necessitate improvements in simulated estimates as well as reconstructed climates and forcing histories.

Generally I find that some changes and improvements to the structure of the results and discussion sections is required so they read more coherently and clearly. There are several important points emphasised in the discussion sections, which I feel could be highlighted. I will cover the specifics below, otherwise I heartily recommend this paper for publication.

Specific comments 1) I found the introduction to be highly informative as well as clearly and excellently written. I only ask for the sentence starting **p3791, l15** "This means..." to be expanded and clarified as I understand the principle but it's a bit awkward to read.

2) I realise that the Mann et al (2009) provides the only current reconstruction that includes field data and can be used throughout your analysis, but it is a little bit of a shame that the other major reconstructions can't be included to account more evidently for reconstruction uncertainty. They could perhaps give more information on the probabilistic conclusion of differences regarding (MWP-LIA) between the simulations and reconstruction.

3) Section 2.1 is finely written and concludes with some very important and informative

points!

4) Section 2.2 **p3798, l2**: Are these regions sensitive/robust?

5) **p3799, l13-l15:** good, an important point to stress!

6) In Section 3.2 and generally thereafter I became, probably through my own ignorance, confused by the discussions relating to over- and under-disperse in the text. This is in part because I was expecting a simulated or effectively "sample" distribution on the $y-axis$ and a reconstruction target or "theoretical" distribution on the $x-axis$ to be consistent along $y=x$. However, if this was the case, Fig. 3f or Fig. 2 top-left for example, would show high counts at the tails of the histogram counts distributions, when instead the early, middle and late periods all seem to be less weighted at the tails. I'm sure your arguments are right, I would just need a little more "hand holding" through this section.

7) I would appreciate **p3804, l16-l18** being fleshed out a little to explain which "deviations visualized" are contributing to the conclusion in this paragraph and which types of ensemble characteristics you are speaking of, especially as we don't have the COSMOS-Mill results at hand.

8) The first line of Section 4.4, I couldn't relate this statement with any evidence in Fig. 5. Is it actually Fig. 6 where any indication is made about the differences between the reconstructed and simulated start of the millennium?

9) Here I am thinking specifically of **p3808, l20-l22:** where reference is made to a result that is not shown comparing two groups of simulations separated by different volcanic forcing data-sets; and the paragraph from **p3810, l9-l20** where multi-model and single-model ensembles are compared (past1000 vs earlier COSMOS analysis). I consider these to be results of great interest in the article but they are rather swamped by way the article is structured in Section 4 and 5. I feel that these results in Section 4 particularly, could be shown in the article explicitly and much of the fragmented results

in Section 4 about "sources of uncertainty" could be moved to Section 3 where they appear to me to be complimentary side-by-side rather than two separate analyses. I mean, no mention is made of these rather interesting conclusions in the abstract for example. In the case of the distinct volcanic groups of simulations, I would like more discussion of how the sub-ensemble variations can be compared with the differences in simulations with different volcanic forcings, such as if you use several of the GISS ensemble members for example.

10) **p3810, l26** Do the "long-term trends" refer to when the analysis is split into sub-periods?

11) Table 1 is very useful!

Technical corrections

p3791, l6: "set agrees **with** the"
general comment the Southwestern North America region of (Wahl and Smerdon, 2012), is it not possible for simplicities sake to refer to this throughout the paper by the authors assignment of 'North American Southwest' NASW?
p3792, l11: "allows consider**ation of** the consistency"
p3792, l12: "not only **with** initial and **forcing uncertainties** (**as** for the COSMOS-Mill ensemble) but **also with** different parametric choices**, physical** models **and different** structural"
p3793, l7 and l9: perhaps the "see, e.g. Author" can just be replaced with "see, Author" in both cases
p3794, l19: "target quantiles**, indicating an over-dispersive data set. **"
p3794, l21: "data **set, which we refer to, as under-dispersive.**"
p3795, l26: "reconstruction**s**"
p3796, l3: "re-consider**ation of**"
p3796, l7: "or biased **in terms of** uncertainty, further"
p3796, l10: "that, **with** the present"

p3796, l25: "reliable **during** this period"

p3799, l4: "We will **return** to this point"

p3799, l7 and l9: Question to the Editor, should it be "centre", i.e. British English here?

p3799, l10-l12: "full stud**y** period, and thereby shift focus **on to** the comparability of the variability **over pre-industrial times** only."

p3799, l13: "neither allow **the ranking of** the various **simulated realizations**" against

general comment you use the word "relation" (e.g. **p3801, l7**) often where "relationship" is perhaps better suited

p3801, l7: "between **the** reconstruction and **the** simulation'

p3801, l8: "simulations **rarely** represent the"

general comment I rather recommend "see Fig. X" instead of "compare Fig. X" throughout the article

p3802, l10: "for the AMO **in this period** (Fig. 3f and h)"

p3802, l14: "the simulation data do**es** not reproduce"

p3803, l22: "their variability **in** the simulation ensemble"

p3804, l24: you don't have to agree with me but may I suggest "We consider correlation analysis as a universal method in studies comparing simulations and reconstructions"

p3804, l26: "reconstruction indicates that **to some extent both data-sets** feature a similar signal"

p3806, l25: I was a bit unsure if this was a similar statement to the one earlier about correlation methods. "Again we discuss correlations as example for common practices". Is this a justification of the method?

p3809, l5: "Furthermore, **it is possible the simulations** do not fully"

p3809, l6: perhaps remove the 'e.g. the' so it reads "influence of, solar forcing"

p3809, l11-l12: "also showed that **a** larger agreement and consistency closer to the present **should not be expected.**"

p3810, l22: You can remove "only" from this sentence

C1913

p3811, l8: "the necessity **for** improvements **of** simulated estimates"

p3819, caption remove "'the range between' on line 4 which becomes "in steps of 0.2 within 0.1 and 0.9"

p3819, caption for clarification, does 'single deviation tests' mean tests done on each individual grid point?

general comment regarding figures with residual qqplots, I realise that axis labels are impractical in Fig. 2, but an explanation of the axes would be useful for the reader. This also relates to Fig. 3 and Fig. 4 where "Simulated Target" is used in the former, whilst "Projection Target" is in the latter. Is there any significance to this difference?

Interactive comment on Clim. Past Discuss., 9, 3789, 2013.

C1914