**Climate
of the Past
Discussions**

Interactive comment on "Climate of the last millennium: ensemble consistency of simulations and reconstructions" *by* O. Bothe et al.

Anonymous Referee #2

Received and published: 11 September 2012

The authors admirably approach an increasingly important area of research involving the comparison of both the simulated "model" climate and that reconstructed from proxy information, however the article has a few shortcomings regarding the layout, structure and language which make it, at times, difficult to follow. I have attempted to discuss my desire for the clarification of a few things below in more detail though I want it noted that I find the analysis technically sound and commendable for stepping away from, as stated by the authors themselves, the "wiggle matching" or "by eye" approaches traditionally used in comparing simulated and reconstructed data.

I often found myself confused by what was meant by consistency, though I am aware that these terms are comprehensively dealt with elsewhere. Although a detailed discussion occurs in section 2.1 to inform the reader, you also discuss "reliability of a

C1508

probabilistic ensemble" which I also found confusing. For example, the authors could reference more clearly that this sentence stems from forecast analysis and is now a methodology being applied to climate science (if I have understood this correctly).

I would like help with understanding how "deviations in climatological distributions" (page 2414, line 28) can be interpreted in the reconstructed and simulated worlds. I would also welcome a more thorough discussion of what the probabilistic and climatological components of consistency are in the analysis. Perhaps a brief clarification of the similarities and differences between the rank histogram analysis and the r-q-q plots (from Marzban et al. 2010) would also link well with these clarifications, where you could translate these forecast ideas and methodologies into this climatological comparison framework to help the reader understand why you use both methods here. This discussion could be where we find the discussion to help interpretation of the rank histograms (page 2413, line 22).

I have seen the previous referee's comments and the author's reply, however I would reiterate that whilst your definition of "truth" or "true" is discussed somewhat and referenced towards Annan and Hargreaves (2010) for example, there are instances where things are simply not clear such as page 2413, line 23: "If the truth is sampling from a distribution narrower (wider) than the ensemble". I realize these ideas and definitions have already been represented in previous research but is it not possible somehow to distinguish between the "observable truth" and the "true" mean or distribution of a data set by using a more distinct and explicit terminology? i.e. simulated ensemble mean target / reconstructed ensemble mean target?

Regarding section 2.2, it is not clear to me why the approach is reversed only for the Frank et al, (2010) data and not in the Central European and global temperature field consistency analyses. Note here, I find the inclusion of the "single forcing simulations as a valid hypotheses about the pre-industrial climate trajectory" page 2415, line 11, as likely to be useful in broadening uncertainty estimates regarding pre-industrial forced components.

C1509

In regards to the order of some results discussed in the main text, there is a good results summary at the start of section 3.1.2, but I wondered if it might be better to discuss Fig. 1 earlier in the text? I felt the same for Fig. 4c which is discussed in great detail in section 3.2.2 after later figures are expanded upon earlier on in the results discussions.

Page 2423: Here you refer to the "first" or "last" records, forgive me if I have missed something but I don't understand what these are or how they are defined. You could expand upon line 1-3 describing the different periods. You also mention "early" and "late" records, how do these fit in with the first, second, third and last described in the caption of Fig. 5? I think it would help if you expanded upon your description of the Fig. 6 results generally here as well. I.e. page 2424, line 20: Here you refer Fig. 7e as being "about 1595 to 1845". This specific terminology is helpful and could be applied earlier in your results.

Page 2424, line 25-29: These statements feature significant conclusions, perhaps the motivation for these statements should be shown in Figures? or at least in supplementary material.

Page 2424, line 10: What is meant by a "moderate random error"?

Page 2425, line 6: Could you expand upon what "an error estimate" is defined as here?

Section 4: This introduces a lot of new results into the analysis which are in addition to those discussed earlier. It is good the authors contextualize their work here but there is a great deal of new analysis introduced here that is not shown (perhaps it could be included in supplementary material?). Is there some way you can integrate these additional results more clearly with the earlier analysis so the reader can see how and where in your analysis these additional results (and the work by other authors) can be compare?

Some further comments about the otherwise excellent Figures:

Figure 2: You have results for both an ensemble mean simulation Fig. 2d for one area averaged NH series, and for the ensemble mean reconstruction equivalent Fig. 2f next to one another. What are the implications for the results in using these different approaches? I was also unclear what "sub-ensemble" refers to in regard to the results of Fig. 2e and 2f.

Figure 4: I would appreciate further discussion regarding the "mapped ranks" and how they were derived.

Figure 5: I think it would help if you highlight in the caption of Fig. 5 that when you say "left to right for the first, second.." you are discussing the coloured box legend in each individual histogram.

Incidental remarks:

Page 2418, line 16: "Contrarily," could be phrased "In contrast,"

Page 2418, line 23: add an "a" here: "but originates from only a few ensemble"

Page 2420, line 10: becomes "simulation ensembles and reconstructions."

Page 2423, line 20: you can remove the "n" from "neither" and from "nor".