

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

Anonymous Referee #1

Received and published: 29 August 2012

This is an interesting idea and analysis, but some aspects of it are rather unclear to me.

In general, throughout the manuscript, I often found it hard to follow exactly what was being tested against what. It would be very helpful if some clear terminology could be set out and consistently applied. It seems that the ensemble mean (of model simulations, or proxy reconstructions) is being tested for compatibility as a sample from the ensemble of reconstructions or model simulations respectively. Some of these analyses appear to have been done both ways round, but others (figs 4 onwards) only in one direction. It would be useful to have a clearer statement of the hypotheses being tested (eg, that a specific validation target was drawn from the same distribution as a particular ensemble).

C1363

The authors state in their abstract that "no status of truth can be assumed for climate evolutions...". But what is meant by a status of truth anyway? I'm afraid I don't grasp the point being made.

I think the authors also need to be more careful with their mention of a "true" distribution eg on line 25 of p 2411. What exactly is "truth" here? I think a more accurate description of the situation is that the historical climate system had a true state (or vector of truth through time), and any uncertainty, which may be represented through a distribution, is solely due to our limited knowledge. The concept of truth therefore is not applicable to the distribution itself. More generally, I suggest looking for instances of "truth" or "true" in the manuscript (which is usually presented in quotation marks, already suggesting some discomfort with the term) and considering whether the concept at hand can be better expressed. In the case of a verification target (eg Fig 3), it is probably uncontroversial, but as a description of a distribution, it seems inappropriate.

The use of ensemble means as verification target introduces a confounding factor due to smoothing of internal variability. This is briefly mentioned in the manuscript, but does not appear to be adequately treated. In the simple perfect case where all time series for both ensembles are drawn from the same generating method, the ensemble means will tend to show inadequate variability in a naive comparison. Perhaps this underlies the result summarised as: p2421 l13-15 "Figure 1 further shows that the considered ensembles of estimated temperature anomaly series generally enclose the verification data (Fig. 1a–c), but they often over- estimate inter-annual variability (Fig. 1d–f)." This issue should be addressed more clearly, eg by augmenting the ensemble mean verification target with internal variability "noise" (contrast with the use of noise on the ensemble members, to account for observational errors on the target).

Incidentally, the Hind et al manuscript is now published in two parts, as the authors may have already noticed.