**Climate
of the Past
Discussions**

Interactive comment on "Clustering climate reconstructions" *by* G. Bürger

Anonymous Referee #1

Received and published: 18 May 2010

The paper compares 10 reconstructions of the NH temperature. All reconstructions cover the last millennium but are based on different methodologies and proxy compilations. The author defines a "similarity index" based on the coherence between pairs of reconstructions. Based on this index the reconstructions are clustered into groups of reconstructions that are "similar". Additionally a multidimensional scaling is used to map the reconstructions onto the plane so that the "similarity" is transformed into distances.

The author emphasizes that many reconstructions are not consistent and he raises some methodological and philosophical issues. These issues are related to how the inconsistent reconstructions can be interpreted in a meaningful way.

I think the paper is interesting and that it deals with an important topic. Different reconstructions are often only visually compared and the more analytic approach of this

paper is an important step forward. I am also quite confident that the cluster analysis and the multidimensional scaling are done correctly - in fact the results of these analyses are not very surprising given the values in Table 2. However, I have some concerns that the author must resolve before I can recommend that that the paper is accepted.

Major comments:

1) My major concern is about the significance levels. Not much is mentioned about this in the paper, only in the caption to Table 2 is there a brief comment (0.47 as the 99 % level). It seems that the results of the paper depend very much on this value and it is therefore important that the method to calculate it is described in detail. One important parameter is the number of degrees of freedom used in the calculation of the threshold value (0.47). There is a lot of serial correlations in the reconstructions so the number of degrees of freedom is smaller than the number of years. How this is dealt with should be described in details in the paper. Also, I guess that only the reconstructions themselves and not their uncertainty bands are used when the significance level is calculated. But the reconstructions would probably appear more coherent if the uncertainty bands are included.

2) I am also confused by the way the significance threshold is used. You can reject the null-hypothesis of non-coherent reconstructions if the coherence is above the threshold. But does that mean that two reconstructions are non-coherent if their mutual coherence is below the threshold? If the reconstruction period had been short enough the threshold value would be large enough for none of the coherences to be significant. How should that be interpreted?

3) The author uses exclusively an integrated value of the coherence as a measure of "similarity". I would like to see this described in more detail. How is the integrated coherence related to the cross-correlation coefficient of smoothed data (as used in Juckes 2007)? The author mentions that coherence "present a better protection against e.g. spurious significance". Better than for the cross-correlation? The author should be

much more detailed here.

I guess that the coherence does not depend on the amplitude and offset of the reconstructions (this is also not very clear from the description in the paper). But many reconstruction methods differ very much in these aspects (Christiansen et al. 2009). I think the author should repeat his analyses with other indices of "similarity" such as the cross-correlation and indices based on the low-frequency amplitude. Perhaps that will give different results. It could be that the coherence is not a good measure. From Fig. 3 it seems that a large part of the low-frequency variability is common to most reconstructions. Most of them have the coldest period in the beginning of 17th century and the warmest in the beginning of the millennium.

Minor comments:

p660,l5: Multiple proxies -> proxies of mixed types?

p660,l13: This example does not hold if the +-5 % is interpreted as one or two standard deviations as usual in climate science.

p660,l18: What is "reasoning under uncertainty: and "paraconsistent logic"? The author should be more explicit here.

p660,l26: "Weighted accordingly" to what?

p662,l25: Note that Christiansen et al. 2009 showed that the reconstruction methods all have serious biases regarding the low-frequency variability. So even if the methods are consistent they may all miss the target. Christiansen et al. 2009 also found that the low-frequency correlation (with the target) was well reconstructed by most methods in agreement with the speculations in the present paper that the major part of the problem of the weak coherence lies in the different proxy compilations.

Fig. 3: There are only four panels to show the five clusters.

p666,l13: But they do seem to agree about the low-frequency variability. Perhaps

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

coherence is not the best measure (see above).

p666,l15: A verification procedure could be based on pseudo-proxy experiments. Such experiments could give information about the influence of the number of proxies, their positions, the choice of regression method etc.

Interactive comment on Clim. Past Discuss., 6, 659, 2010.