

Interactive comment on “Evaluating climate field reconstruction techniques using improved emulations of real-world conditions” by J. Wang et al.

Anonymous Referee #1

Received and published: 12 July 2013

The paper compares 4 different climate field reconstruction methods using pseudo proxies based on a millennium long forced climate experiment. In contrast to most previous studies of field reconstruction methods the authors include the spatial inhomogeneity of proxy quality and the temporal decline in proxy availability.

The paper is interesting and contributes significantly to the field. However, I have some comments both the methodology and the presentation that the authors must address before I can recommend that the paper is accepted.

Major comments:

- 1) I think the arguments in section 2.1 leading to the conclusion that "proxies .. are
C1446

not indicative of local temperatures .." are wrong. The problem is that with increasing distance between the proxy and the temperature more and more grid-points will be scanned looking for the maximum correlation. While a correlation of 0.6 would be hard to find by chance in a single grid-point it would be less rare to find as the largest correlation in 10 different grid-points. So the explanation for the plot in Fig. 2 could be a combination of two effects: 1) as the distance increases the correlations decrease because the proxy/temperature relation is local. 2) As the distance increases the correlations increase because more and more grid-points are included. If the statistical significance was shown in Fig. 2 is would be around 0.4 for a distance of 0 and fall off quickly with distance.

Thus, the form of Fig. 2 could be a simple consequence of the procedure and is not necessarily connected to teleconnections. Therefore, the max SNR experiments may be useful as sensitivity studies (like a best case scenario) but they have nothing to do with reality.

I am not quite sure how the max SNR procedure works. The SNR is calculated using the maximum correlation. But is the noise added to the local temperature (where the proxy is positioned) or to the local temperature from the grid-point with maximum correlation?

- 2) The manuscript lacks many details about the methodologies. Are annual means of temperatures and proxies used? What is the calibration period? Are the time-series centered to zero over the calibration interval. Is the trend in the calibration period removed before the correlations are calculated?

It is stated on page 3020 that many of the proxies are of decadal resolution. Is this reduction of degrees of freedom taken into account when the significance of the correlations are calculated? And how are the corresponding pseudo proxies treated? These should somehow reflect the resolutions of the real proxies. In Christiansen and Ljungqvist (J. Clim., 24, 6013-6034, 2011 and Clim. Past. 8, 765-786, 2012) the local

temperatures were low-pass filtered with a cut-off at 5 or 10 years to match the proxy resolution. Is something similar done here? There is a single sentence regarding this in section 2.3 but I can not understand what actually was done. I think the non-annually resolved proxies can be the source of much uncertainty in the reconstructions

The authors should also give some values for the different truncations used with the RegEM and CCA methods. These might depend on the specific ensemble (do they?) but at least some average numbers should be given.

3) A major conclusion is that the reconstructions are best in periods with forced variability (section 4.3.1). It is argued that this is because the temperature anomaly is more spatial coherent in these periods. However, I don't think this argument is very strong. The patterns in Figs. 7 and 8 could be a consequence of the bias being large/small in periods with large/small temperature anomalies. It would be nice to see the CE split into bias and variance. Also the authors should give some kind of measure of the spatial coherence in the 100 year slices.

Minor comments:

p3017: What is meant by "analytical uncertainties of proxies" and "each methods inherent risk properties"?

p3023 and Fig. 4: In the text it is the "sum of the SNR" in the figure caption it is the "average SNR". I can understand that the number of proxies becomes smaller going back in time and therefore the sum of the SNR falls accordingly. But does this also happen for the average SNR? Is it the worst proxies that reach the longest back in time? Some clarification is needed here.

Section 2.3. Regarding the persistence of the noise there is a discussion of this issue in Christiansen and Ljungqvist, J. Climate, 25, 7998-8003. Here the uncertainties are compared when the noise is white, ar_1 , or "realistic".

p3026, top: I think all the methods are regression based. Where they differ is in the

C1448

method of inference. Most studies use some kind of least squares with some kind of regularization. A few authors, e.g. Tingley, use Bayesian inference. But even here the underlying model is based on linear regression.

p2025: Strictly speaking the matrix P in Eqs. 3 and 4 must be a pruned version that match the length of the calibration interval. Also "suboptimal" is a weak word to use here where $P \# P$ is non-invertible.

Section 3.1.1: It should be specified which field μ and σ are calculated from. There is a discussion of the relative merits of RegEM Ridge and RegEM TTLS in the Comment/Reply (J. Climate, 23, 2832-2838 and 2839-2844) that could be cited here.

Section 3.1.4: It is probably difficult to describe GraphEM in a few lines, but I did not become much wiser from reading this section.

Section 4.1: The authors state that the CE and RE statistics are related to the MSE. This is probably correct but it is not easy to see just how. In Fig. 6 the authors also present the MSE, the bias, and the variance. I find these statistics much more useful than the CE and CR and will urge the authors to present the bias and the variance throughout the paper. This will also help interpreting the conclusion that the CE statistics depends on the type of variability (beginning of section 4.3.1).

Figure 5: What is the calibration period and what is the baseline? It is my experience that reconstructions usually are biased towards zero (relative to the mean over the calibration period) as also mentioned by the authors in the end of the discussion. But here the bias is positive also when the target is positive.

I notice that the target in the reconstruction period is well inside the range of the target in the calibration period. This might lead to conservative (low) estimates of the bias (see the Comment/Reply mentioned above) and should be discussed. I think this is a more serious limitation than the mere low internal variability mentioned in the beginning of the discussion.

C1449

It is not easy to see the tick-marks on the x-axes.

p3035: "This is encouraging ". Yes, but as I argue in the major comment above I think these high correlations in the max SNR network are spurious. The discussion should be changed to reflect this (perhaps the max SNR experiments should be removed from the paper).

p3039, end of discussion: Much of the bias in the reconstruction methods comes from expressing the temperatures as a sum of the proxies plus noise (as in Eq. 3). This is unphysical and will lead to biased estimators. The solution is to go to forward models which express proxies as sums of temperatures plus noise. This will remove the bias but may increase the variance. The variance can then be reduced by temporal and or spatial smoothing. This is explained in Christiansen, J. Clim. 674-692, 2011 which should be cited here. Forward models are also used in the Bayesian settings of Tingley et al. Ammann et al. deals with the different but closely related problem of errors-in-variables models where noise is both on the dependent and independent variable.

Table 1, caption: What is the calibration period and the verification period? Is the mean spatial or temporal?

Interactive comment on Clim. Past Discuss., 9, 3015, 2013.