

Interactive comment on “On the verification of climate reconstructions” by G. Bürger

Anonymous Referee #1

Received and published: 22 March 2007

On the verification of climate reconstructions By G. Burger

This paper tries to assess the skill of proxy-based reconstructions of climate by comparing the results of various methods (forward and backward regression, truncated total least squares, ridge regression, expectation maximisation algorithm and RegEM) and various methods of preprocessing and postprocessing. The skill is measured by two validation statistics : RE and CE. It proposes a Monte-Carlo approach to assess the significance of the reconstructions. It brings a rigorous assessment of the quality of the reconstruction and bring objective arguments to the well known debate about the various millennial climate reconstructions. I think that this paper must be published. Nevertheless I have some specific comments which must be taken into account to provide an improved version of the paper.

Abstract : it should start with a summary of what has been done, explaining that sev-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

eral calibration methods have been compared with several pre- and post-processing methods. As it is written, it is difficult to understand on what is based the paper.

p.251, l.4 : replace last report of IPCC by third report of IPCC. At end of this paragraph, it should also be good to mention that the debate was reflected in the fourth report of IPCC (in press)

p.252. l14 : bootstrapping has been applied in more than one paper: e.g. Till and Guiot, 1990 (Quaternary Research), Guiot et al 2005 (The Holocene). Certainly in other papers but I have not the time to make a bibliographic research. The way to use bootstrapping is important. Guiot et al (2005) and Till and Guiot (1990) takes randomly, with replacement, a subset of n samples among the full dataset of also n samples (that means that in the subset, same sample can be repeated and that the calibration subset has always the same size than the full dataset); these samples are used for calibration and the remaining samples are used for independent validation (RE and CE can be calculated on the latter ones) ; Burger here uses a random extracting with NO replacement and with size of $n/2$. I do not know how much both approaches are different in practice, but conceptually they are. In the early papers of Efron, I had read that it was important to have a calibration set of same size (n), but it is true that he did not used the remaining samples for independent statistics.

p.252.l20-23 : this sentence is difficult to be understood.

p.255, l.15 : it is not so clear that ; I understand that it is true when $=R_c$, but if it larger than R_c , is it always true ?

p.256, l.15 : please provide the equation of adjusted R^2

p.258, eq.5. define late and early

p.258, l.13 : do you mean that you apply 500 times eq .5 itself based on 100 swapping ? Please try to better explain the method : all is confusing

p.259, l11-13 : not useful, it just adds difficulties of understanding

p.262, l10-11 : this sentence has nothing to do in a presentation of methods ; it must be reserved to a discussion of results.

p.263, l.23-24 : this test is based on the reconstruction of the average hemispheric temperature. The fact that PC1 is well reconstructed is a good point for that NHT, as likely PC1 is very close to that mean. But the counterpart is that with a single PC, it is not possible to reconstruct regional gradient. I know that it is not the objective but as it is presented, the reader could believe that the use of a single PC is recommended.

p.265, l.11 : what do you mean by “ allowance for nonsense regressors ” ? do you mean rather “ comparison with nonsense regressors ” ?

p.265, l.14, under a significan(c)e level of 0% : I have never seen of probability level of 0 in statistics

p.266, an alternative method is the bootstrap confidence interval of RE. If the value of 0 does not belong to the interval, its means that the reconstruction is significantly different from the climatology. It has the advantage to avoid the generation of nonsense predictors and then to avoid the risk of nonsense predictors with characteristics different from those of the true predictors. This true bootstrap method is used for the R2 on validation samples (not used by the random extraction) by Guiot et al (2005).

p.266, point 5 : add a few words to precise which characteristics of the predictors are kept.

p.268, l.13-14, they become more and more dependent if the memory of the time series gets comparable to the series length : I think that high autocorrelation of order 1 is enough to make validation dependent on the calibration, as a sample is in the calibration set, there is a large chance that the next or previous sample (according to chronology) is in the validation set.

p.269. end of conclusion : I agree that statistical significance of RE and CE will not help very much to decide if the Hockey stick is realist or not and it is clear that a single

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

reconstruction cannot help to decide. To decide, two points must be stressed here : (1) several independent approaches, i.e. using different proxies, must be compared, (2) the reconstructions must be done on a regional basis to be able to analyse the spatial coherency of the results and overall to understand the effect of unequally distributed proxies. The fact that the flavors involving PC1 outperform the other ones shows the interest to remove the noise from the predictands, but the counterpart is a loss of variance as this PC1 is usually the common signal between all the local temperature series. As Medieval Warm Period is likely not a global phenomenon, its amplitude may be attenuated according to the 20th century warming which is much more global.

Interactive comment on Clim. Past Discuss., 3, 249, 2007.

CPD

3, S96–S99, 2007

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper