Clim. Past Discuss., 2, S182–S184, 2006
www.clim-past-discuss.net/2/S182/2006/

Climate of the Past Discussions

Interactive comment on "On the verification of climate reconstructions" *by* G. Bürger and U. Cubasch

J. Guiot

guiot@cerege.fr

Received and published: 10 July 2006

Rev#2 has done a detailed review of the manuscript of Burger and Cubash (BC) with many interesting comments. But some of them are given to try to close a debate in which, in my sense, there is still some things to say. The research considered by the paper is indeed eight year old, but the matter continues to be discussed in many recent papers. Even if some studies and audits have validated the conclusions of the Mann et al (1998) paper (including the US National Academy of Sciences), I think that it is still appropriate to discuss the statistical bases of climate reconstructions especially when there is an effort of generalization as in the BC paper. Rev#2 recognises that " the Journal of Climate would of course publish any legitimate criticisms of the Rutherford et al (2005) or Mann et al (2005) findings " but why not Climate of the Past ?

An important point raised by BC, is the performance of regEM when there is a large percentage of missing values. It is not really the situation for which Schneider (2001) has applied his method and in this sense it is legitimate to test various flavors of the regEM method which are maybe better than this used by Rutherford et al (2005) in the extreme situation where missing values are numerous. Additional “flavors” should be even defined to take into account the increased number of missing data and its effect on the final skill.

Another point raised by rev#2 is that “the authors’ statistical significance estimates are meaningless, as they are based on random permutations of subjective “flavors” that are simply erroneous in the context of the correctly implemented RegEM procedureĚ ” (beginning of page S146). I have the impression that rev#2 is confused by the permutation tests used by BC : as all the statistics done on each flavor are independent, there is then no effect of “ subjective ” flavors on the “ correct ” implementation of regEM. The fact to test additional flavor has no effect on the other tests. Nevertheless a potential problem I see about the significance of the results is that BC assume that the predictions are white noise and not red noise as done by Mann et al (2005).

By the way, I would like to point out the fact that the use of RE is standard in dendroclimatology but not in other disciplines. With this parameter, there is no way to decide if a score is significant or not (it is just admitted by Lorenz and other subsequent users that it must be positive). The term of significance is then not properly used. In the statistical literature, RMSEP is largely preferred. The probability distribution of this parameter is not known (as for RE), but it has the advantage to be more largely known to be directly comparable with the error bars of the reconstructions. Its significance is tested with bootstrap or leave-out-one methods (see Ter Braak, 1995).

A key point raised in the paper is that verification and calibration samples must be taken from the same population to have a sense. It is true and the fact that the 20th century has a strong tendency in the temperature series contradicts this assumption when one dataset is taken in the first half and the second in the second half. It is true

CPD

2, S182–S184, 2006

also that such trend decreases the actual number of degrees of freedom, and that a predictor having accidently such trend could be taken as good predictor, even if this is a nonsense. Rev#2 claims that it has been tested by Mann et al (2005), but I do not see in that paper any test where the calibration and verifications periods have very different properties. Indeed they use the 850-1855 period for verification which is too large to be very different from the various calibration periods tested (all starting after 1855). This is then an open problem.

Another criticism of rev#2 is that " The authors moreover appear to be unaware of the large number of model simulations that have been performed that independently agree with these various empirical reconstructions within the estimates uncertainties, reinforcing conclusions regarding the anomalous nature of recent warmth " (end of page S150). It is not the object of this paper. It is likely true that the recent warming is anomalous, but it does not prevent any discussion on the statistical methods used.

In conclusion, I think that the objectives of this paper are legitimate. It proposes a extended method to evaluate the results of climate reconstructions (not only the reconstructions of the Mann's collaborators). Even if this method is not completely new, it has some interest. This way to define flavors of one or several methods is instructive. Independently on that, the paper could be much clearly written. As it is, it is reserved to people knowing very well the debate around these reconstructions. I suggest to the authors to think to a figure to summarize the main findings about the various flavors (something like an anova of the RE) and to be more explicit in the discussion.

Citation: ter Braak, C.J.F., 1995. Non-linear methods for multivariate statistical calibration and their use in palaeoecology: a comparison of inverse (k-nearest neighbours, partial least squares and weighted averaging partial least squares) and classical approaches. Chemometrics and Intelligent Laboratory Systems 28, 165-180.

Interactive comment on Clim. Past Discuss., 2, 357, 2006.