

Interactive comment on “Refining error estimates for a millennial temperature reconstruction” by M. N. Juckes

M. Juckes

martin.juckes@stfc.ac.uk

Received and published: 2 March 2010

Firstly I should thank the reviewer for his careful attention and detailed comment. I am particularly grateful for his drawing my attention to the Li et al paper, which is clearly of great relevance, though there are also significant differences.

1 Focus

Concerning the focus of the paper, I feel that the additional data can be included without great effort and it is therefore appropriate. This makes the results interesting both as a new reconstruction and from a methodological point of view. There is also an important

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



point to be made from table 1: increasing from $N_c = 13$ to 15 results in a reduction in the preliminary error estimate. That is, the increased data volume is leading to a decreased uncertainty at a rate which is broadly consistent with expectations (in fact there is a 10% decrease where one would expect 7.5%). Thus the inclusion of additional data serves both to bring the data selection up to date and to provide a test of the performance of the uncertainty estimate.

2 Li et al.

The Li et al. paper is, as the reviewer says, very relevant. Though there is much overlap, there is also a surprisingly large divergence of approach. Firstly, Li et al. use a parametric bootstrap rather than a resampling approach. That is, they first fit a noise model to the proxies and then analyse an ensemble of reconstructions obtained by sampling the noise model. They do not consider correlations among the proxies (they use the R 'gls' function, which allows for "within group" correlation, but not between group correlation); they use 13 year periods for calibration of the error estimate (through an inflation factor) – an approach which is incompatible with slowly varying proxies used in this study; they look at over-fitting, but not at structural uncertainty in general.

They use 14 Mann et al. 1999 proxy series, of which 4 are in the southern hemisphere, to estimate N. hemispheric temperature.

A more detailed discussion will be included in a revised paper.

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

3 Bootstrap

I had, prior to submission, done some bootstrap estimates and found that the method gave a smaller initial error estimate – that is, the estimated uncertainty variance of the reconstructed temperature given by the bootstrap algorithm, averaged over the 11th century, is $\overline{\hat{\sigma}_{boot}} = 0.125K$ compared with to $\overline{\hat{\sigma}_{d:jack}}$ (or, in the reviewer's preferred notation, $\hat{\sigma}_{d:jack}$) in the range 0.139 to 0.147K in table 1 for $N_c = 15$. The structural uncertainty adjustment would be the same for both approaches, but I do not have a quantitative estimate for the effect of proxy-proxy correlations on the bootstrap estimate. In the submission I decided to omit reference to the bootstrap results but, given the reviewers comments, it is clear that it needs to be discussed. I will add the $\hat{\sigma}_{boot}$ values to table 1 with a brief discussion. The primary reason for adopting the Jackknife is that it provides (in this instance) a more conservative error estimate. In order to understand these differences a little I have checked how the two algorithms deal with the simpler problem of estimating the variance of the mean of N iid Gaussian variables. The bootstrap method has a clear low bias at small N , though both methods are consistent as $N \rightarrow \infty$.

A second reason for using the Jackknife approach here is the availability of a correction factor, derived in the paper, to allow for between group error correlations (i.e. correlations among the errors of the different proxies). I am not aware of a comparable method for the sampling Bootstrap algorithm. It would, however, be possible to represent this effect in the parametric Bootstrap, and I will, for completeness, investigate that option (with the caveat that the parametric Bootstrap is not well suited to representing the significant variation in error characteristics between different proxies in the collection used here).

Many authors expressing preference for the bootstrap do so on the basis of Efram (1979) result concerning the variance of the median as $N \rightarrow \infty$. The standard Jackknife ($d = 1$) is not consistent for this problem, but Duttweiler (1973) [see Shao and Tu

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



(1996), p53] show that the delete- d version is consistent if $d > \sqrt{N}$.

I will respond to the remaining issues after revising the paper.

4 References

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.*, 7 (1): 1-26. doi:10.1214/aos/1176344552

Duttweiler, D.L. (1973). The mean-square error of Bahadur's orderstatistic approximation. *Ann. Statist.*, 1, 446-453.

Interactive comment on *Clim. Past Discuss.*, 5, 2631, 2009.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

