

9th Mar 2004

Dear Mr McIntyre

Your comment entitled "Global-scale temperature patterns and climate forcings over the past six centuries: A comment" on the earlier Nature paper by Mann et al. has now been seen by those authors, and their response is attached.

We also sent the exchange to two referees, whose comments are enclosed. You will see from this advice that, while our referees are essentially in favour of publication of the comment and the reply in some form, they make important criticisms that we would like you to address before we reach a final decision on publication.

Please accompany your revised manuscript with a short letter explaining the changes you have made in response to the comments from the authors and the referees.

We ask you to use the following link to submit your revised manuscript:

Referee #1 (Remarks to the Author):

I find merit in the arguments of both protagonists, though Mann et al. (MBH) is much more difficult to read than McIntyre & McKittrick (MM). Their explanations are (at least superficially) less clear and they cram too many things onto the same diagram, so I find it harder to judge whether I agree with them.

There are two main points of dispute:

1. The principal component technique used.
2. The quality of the early data.

I deal with 1. first. It is an area where I have expertise, but it is not at all clear what exactly is being done. MM talk about 'scaling to 1902-1980 mean and standard deviation', whereas MBH's phrase is 'restandardisation by detrended standard deviation'. These suggest different things to me. The latter seems more appropriate than the former, but I am still uneasy about applying a standardisation based on a small segment of the series to the whole series, if that is what is being done. MBH seem to be too dismissive of MM's red noise simulations. Even if red noise is not the best model for the series, they should have reservations about a procedure that gives the 'hockey stick' shape for all 10 simulations, when such a shape would not be expected. Having said this, MM's corrected series in their Figure 4c still has the upward trend towards the end of the series, so this trend is not just an artifact of MBH's PCA procedure.

I am not qualified to say much on 2. but it seems to be the crucial point. Both sets of authors agree that the omission of some early data changes the early reconstruction considerably. MBH say that the omitted data are reliable; MM say they are not. Does anyone know who is correct? If there is disagreement among experts, then the true behaviour of the series must be very uncertain.

Incidentally, I am not entirely convinced by MBH's dismissal of the MM model reconstruction on the basis of RE. I suspect that a lot of the difference is due to the much larger variance in the MM model reconstruction compare to MBH's. This is probably inevitable, given the reduced sample size for the early data.

Referee #2(Remarks to the Author):

The technical criticisms raised by McIntyre and McKritik (MM) concerning the temperature reconstructions by Mann et al (MBH98), and the reply to this criticism by Mann et al is quite difficult to evaluate in a short period of time, since they are aimed at particular technical points of the statistical methods used by Mann et al, or at the use of particular time series of proxy data. A proper evaluation would require to redo most of the calculations presented in both manuscripts, something which is obviously out of reach in two weeks time. Furthermore, both manuscripts seem to contradict each other in some basic facts.

Therefore, my comments are based on my impression of the consistency of

the results presented, but there is a wide margin of uncertainty that could be resolved only by looking in detail into the whole data set and the whole software used by the authors.

In general terms I found the criticisms raised by McIntyre and McKritik worth of being taken seriously. They have made an in depth analysis of the MBH reconstructions and they have found several technical errors that are only partially addressed in the reply by Mann et al.

1)Mann et al assert that important features in the reconstruction by MM, for instance the increased warmth in the 15th century, is due to the fact that they completely ignore the time series from the NOAM tree ring data sets. However, MM explicitly state that they have used the two leading PCs of this data sets. Of course, it is impossible to ascertain who is right and who is wrong in this particular point, but I feel that Mann et al should have taken into account in their reply the statement by MM concerning the NOAM time series.

2)My doubts expressed in point 1 are strengthen by Figure 1b in Mann et al. reply. Mann et al. have tried to replicate in this figure the MM reconstruction (MM04c, green line). But one can clearly see that the variance of this reconstruction is much larger than the observations even in the calibration period. Although it might be possible, it seems a very awkward result. Any linear regression method that I am aware of must produce

a reconstructed predictand with less variance than the observed predictand (the rest of the variance being the residuals). It seems to me that the something is technically not correct in the replication by Mann et al of the MM reconstruction. If the main difference between the original MBH98 and the MM04c reconstruction is just the elimination of the NOAM data set, why is the variance of the reconstruction in the calibration period inflated by a factor of 2-3?.

3) The reply by Mann et al is in my opinion correct when requiring MM to present some validation statistics in a validation period, and the RE statistics seems to me adequate in this context. Since this is the main argument in Mann et al reply I would urge MM to address this criticism. The low value of the RE statistics in the replicated MM reconstruction (MM04c) indicated by Mann et al. seems to be due to the erroneous replication of the MM reconstruction. An inspection by eye of the MM04c reconstruction seems to indicate that both reconstructions, the replicated MM04c and the MBH98 method (1400-1500 model) are more or less equally correlated with the instrumental temperature in the calibration period, and that the low RE value of the MM04c reconstruction stems from its much larger variance. I think that this large variance is unrealistic and therefore the real RE value should be positive.

4) The MM reconstruction presented by MM (Fig 4, bottom) should be in principle very similar to the replicated reconstruction MM04c, but it seems to me that they are not. For instance in Fig4, I do not see this excess variance compared to the original MBH98 reconstruction. This suggests again that something is not correct in the calculation of MM04c by Mann et al.

In summary, my recommendation is that MM offer validation statistics for their reconstruction and that they make their original reconstruction (Fig 4, bottom) available to MBH, so that these authors can also compute validation statistics with the original MM reconstruction. Furthermore, it should be explicitly cleared up if MM are using or not the NOAM data set. Should this validation be successful, I would recommend the publication of both manuscripts. Further evaluation requires months of work and should be left, in my opinion, to the scientific community. This should be the normal scientific process.

At this stage, I think any Correction or Retraction by MBH98 is premature and really not required.