

3rd Aug 2004

Dear Mr McIntyre

Thank you for your revised comment on the contribution by Mann et al., which I am afraid we must decline to publish. As is our policy on these occasions, we showed your revised comment to the earlier authors, and their response is enclosed. We also sent the exchange to 3 referees, whose comments are attached.

In the light of this detailed advice, we have regretfully decided that publication of this debate in our Brief Communications Arising section is not justified. This is principally because the discussion cannot be condensed into our 500-word/1 figure format (as you probably realise, supplementary information is only for review purposes because Brief Communications Arising are published online) and relies on technicalities that do not bring a clear resolution of the underlying issues.

Regarding your disagreement with the last sentence of the Corrigendum by Mann et al., I have consulted with my colleagues, who have now given the matter careful consideration. However, the errors Mann et al. refer to in the last sentence of their Corrigendum are errors in the listing of the proxy data sets in the original Supplementary Information, rather than errors in either the data sets used or the computational procedures.

Errors in the listing of data sets obviously do not affect the calculations or results, and we therefore feel that the sentence is appropriate and justified. Please note that it is our policy to inform readers how a corrigendum affects the conclusions of a manuscript where appropriate.

Yours sincerely

Rosalind Cotter  
Editor, Brief Communications

REVIEWS

Referee #1 (Remarks to the Author):

View of the comment by McIntyre et al. and reply by Mann et al. submitted to Nature

It seems interesting that in the comment not only the original publication (MBH98), but also, MBH99, a rebuttal by Mann et al. available from a CRU website (ref.3), an "unreported MBH calculation" available from a University of Virginia website (ref.5), another rebuttal (corrigendum) by Mann et al. now published in Nature (ref. 9), and a detailed critique of MBH by McIntyre and McKittrick published in Environment and Energy (ref. 14) are cited.

Additionally, the paper by Jones and Mann published in Reviews of Geophysics (ref. 4, response) already touches this issue.

Besides numerous technical and data-related issues, McIntyre and McKittrick also address a possible CO<sub>2</sub> effect on southwestern US strip-bark trees that was "corrected" using high-latitude tree-ring data. Whether it was at all useful to use these data or to apply this correction, seems not highly relevant, since MBH never hid this issue, but described it in detail. More relevant and pleasing would be if someone would find a way to assess the possible CO<sub>2</sub> fertilization effects that potentially influence growth at many sites. Additionally, the observation that some of the chronologies used in MBH98 and MBH99 have quite low sample replication during their early periods is also not new and was mentioned in a recent paper published in EOS.

To judge that the criticism by McIntyre and McKittrick is valid would require downloading all data and applying the seemingly differing approaches. Further, judgments would be needed on methodological decisions that were made by both McIntyre and McKittrick and by Mann et al. as two possibilities within the whole spectrum of methodological decisions on which chronologies to use, the calibration and computation of PC's over different time periods, special treatments to series, and so on. It could be seen as interesting, that the calculations as done by another operator with other perhaps reasonable alternative methodologies can have such a large effect on the resulting reconstruction.

Unfortunately, I have the impression that preconceived notions affect the potential "audit" by McIntyre and McKittrick. That would, of course, not mean that their assessment is necessarily wrong, but might explain the rather harsh and tricky wording used here and at other places by both parties, and I generally do not believe that this sort of an "audit" and rebuttal will lead to a better understanding of past climate variations.

Generally, I believe that the technical issues addressed in the comment and the reply are quite difficult to understand and not necessarily of interest to the wide readership of the Brief Communications section of Nature. I do not see a way to make this communication much clearer, particularly with the space requirements, as this comment is largely related to technical details. I also find it relevant that McIntyre and McKittrick already published a critique on MHB98 including some arguments similar to what is outlined in the current manuscript (ref. 14).

Referee #2 (Remarks to the Author):

Nature manuscript 2004-01-14277B Referee 1

Referee 2's comments on the original versions in this exchange suggested that there was insufficient time to understand all the technicalities involved. This is even more the case with these revised versions, their supplementary material and the various replies to replies and comments on comments. The amount of material, often contradictory, is simply too complex and lengthy to resolve all the rights and wrongs in a realistic length of time. I can only give some general impressions and home in on two or three points of detail.

Regarding publication, I think it is all or nothing. Either you publish neither, or both. In the latter case, the main thing that would be achieved is to highlight that a serious disagreement exists. Only a reader with several days to spare (longer if they are unfamiliar with the area), to chase references and probably the authors, could hope to come close to a full understanding of the arguments.

I started my original review by saying that I found merit in the arguments of both MBH & MM. To rewrite this, I believe that some of the criticisms raised by each group of the other's work are valid, but not all. I am particularly unimpressed by the MBH style of 'shouting louder and longer so they must be right'.

I do not have the days of time needed to fully get to the bottom of the arguments, so I look briefly at just three here.

1. I think I understand better than before what the MBH98 PCA is doing, namely centering the data about the mean of the 1902-80 period rather than of the whole series. The question is why, and what properties and interpretation does such a procedure have? Given the non-stationarity of the series, it is certainly not successively maximising variance as in PCA, and talking about 'explained variance' therefore makes little sense. I don't feel I can comment on whether or not this procedure is appropriate without understanding its properties and interpretation.
2. Continuing this theme, the original MM article said that using MBH's PCA on 10 red noise simulations produced a 'hockey stick' (hs) shape in all 10. MBH's response says they have repeated the simulations and 'shown that the claimed result is not true'. It is very unlikely that 10 of MM's simulations all show the hs effect and MBH's do not, simply by chance. Either the two sets of simulations are constructed differently, or there is a mistake in someone's code. This is not something that a referee can resolve.
3. The advocacy of RE in preference to  $r$  by MBH is a bit extreme. The correlation coefficient certainly has drawbacks, but no verification measure is perfect, and I see no evidence in the verification literature (or Wilks) that RE is the standard preferred measure. Indeed the only one of the 3 references (7) cited in the revised response that was available to me is somewhat critical of RE. My preference would be not to rely on a single measure, but to look at contributions from bias, differences in variances and departures from linear dependence.

Referee #3 (Remarks to the Author):

Comments on the manuscripts

Global-scale temperature patterns and climate forcings over the past centuries: a comment, by McIntyre and McKritik (hereafter MM04) followed by comments on "Reply" by Mann, Bradley and Hughes (hereafter MBH04)

After going through both revised manuscripts and the accompanying,

voluminous supplementary material, I have now a much clearer idea about the points of disagreement between both manuscript. I must confess that this has been one of the most difficult reviews that I have been confronted with. Both manuscript plus supplementary material are dense and methodological questions, data quality questions are entangled and both refer heavily on previous information on published papers that has to be scoured beforehand.

The comment by MM04 underlines two apparent errors in the original work of MBH98: the incorrect use of the Principal component methodology to reduce the dimensionality of the NOAMER tree ring data set, and the inclusion of a time series (Gaspe), that seems to be very influential on the final temperature reconstruction.

Considering the changes relative to the first version of MM04, it seems to me that the case presented by MM04 has weakened considerably. The main claim presented in MM04 is now that the main features of MBH98 reconstruction (the hockey stick) derive from two methodological aspects. Now, no preeminence is given to the 16th century being warmer than the 20th century. MM04 have emulated the MBH98 reconstruction improving the methodological aspects that they think are flawed, and arrive to another reconstruction that yields rather low values of RE as verification statistics. They claim that MBH98 reconstruction has also low values of another verification statistics ( $R^2$ ), and therefore conclude that MBH98 is therefore on equal footing as theirs. Accepting this line of reasoning, however, a reader of these manuscripts will be led to think that both reconstructions are not trustworthy (at least, I would not trust any reconstructions with such low values of verification statistics table 2 in MM04 supplementary material). This only conclusion seems to me rather weak for a manuscript. As a reader, I would rather see a more substantial contribution, such as an alternative reconstruction with a sound validation, that may offer some further interesting points- comparison with reconstructions, comparison with the different estimations of forcings, etc..

On the other hand the RE statistics is one that is commonly accepted, not only by MBH98, but by a number of authors working in this field. To argue that other statistics, such as  $R^2$ , may be more meaningful than RE requires in my opinion a strong justification, which is missing. MBH04 offers, furthermore, a plausible reason for the low values of verification  $R^2$  in MBH98 found by MM04. Being a crucial point in MM04, the authors do not provide enough information to assess if these calculations have been performed properly, and for that matter how they have been performed. My own calculations with the data available to me of the validation statistics in the 19th century for the full proxy network tend to support the numbers indicated by MBH04, but this is of course a limited test.

This notwithstanding, I see some merit in MM04 and I would encourage them to pursue their testing of MMB98, and by the way other reconstructions. As I wrote in my first evaluation, this should be a normal and sound scientific process that should not hampered. For instance, questions that seem to be quite critical, such as the sensitivity of the MBH98 reconstructions in more remote periods to changes or omissions in the proxy network or the dependency of the final results to the rescaling of the reconstructed PCs, have become clearer to me now. From the reply in MBH04 I am now afraid that they were not sufficiently described in the original MBH98 work. In particular the PCs renormalization, could have been included as clarification in the recent Corrigendum in Nature by MBH.

At the moment, my opinion is that the present MM04 manuscript could be of interest just for the bunch of specialist working exactly in the area of statistical methods for climate reconstructions, and this only after several hours of considerable work to understand all technical details properly. Perhaps this is caused by the tight constrained imposed to the Communications Arising category.

In summary, judging from the present version of the manuscript and the response by MBH04, I now think that basis for MM04 has wavered and that further work , or further convincing evidence, would be needed to present a more solid case.

Just in case that the editor decides to publish MM04, I would suggest to reformulate the first half of the second paragraph, describing the calculation of principal components of the NOAMER data set in MBH98. The present version can be hardly understood, even by specialists. At the end of the manuscript, I would avoid the term "goodness of fit", since this has another meaning in the framework of standard linear regression methods (related to the linearity or non-linearity of the fit).

#### Comments on Reply MBH04

The reply by MBH04 on the previous comment by MM04 addresses in my opinion both points raised by MM04 in a convincing way. Although it is for a reviewer impossible to check all the technical details involved in this reply, they arguments used by MBH04 seem plausible, and I would say they are probably correct. This is of course no guarantee that the entirety of MBH98 work and conclusions are free of error.

Therefore, if the editor decides to go ahead with the publication of MM04, I would recommend to publish MBH04 as well.

I would have some minor comments on this reply:

In general, in a scientific text I would rather avoid as much as possible disqualifying formulations (e.g. demonstrate a lack of familiarity,,, without merit..) . This is of course a matter of taste, but I think that science and scientist benefit if it the same thing is said in a neutral way.

in page 2, in the middle:  $RE=-1$  is the average value for a random estimate.. This is correct only when the estimate has the same variance as the true values. However, I think that a more useful bench mark value for RE would be actually zero, since a poor man's prediction using just climatology would yield zero in the case of a stationary process. Certainly, negative values of RE indicate a quite poor skill.

Second to last paragraph: MBH 04 refer to other published reconstructions to support the lack of 16th century warming. I think this reference is to some extent bowed to match the authors intentions. Some of these

reconstructions are their own, and others (e.g. Esper et al) show considerably disagreement with MBH98. In any case, the 16th warmth has been dropped in this version of MM04.

Supplementary material : I have revised the original MBH98 publication and I could not find any description of the renormalisation of the reconstructed PCs. If I am correct, one could not recriminate MM04 for not having included this step in their protocol. This renormalisation seems to me somewhat awkward. If I understood properly this amounts to a statistical inflation, which would not yield the best estimations in the sense of Least Square Errors. I do not think that this invalidates the method, but some readers will perhaps be surprised to find out that the MBH98 reconstruction method included this step.