Backgrounder for McIntyre and McKitrick "Hockey Stick Project"

January 27 2005

Overview

We (McIntyre and McKitrick) are profiled in the cover story of the Feb. 1, 2005 edition of *Natuurwetenschap & Techniek (NWT)*, a prominent European science magazine (both Dutch and English versions are at www.natutech.nl). The cover story is based on two new peer-reviewed papers being published in the well-known science journals *Geophysical Research Letters* and *Environment and Energy* (see www.multiscience.co.uk).

Our article "Hockey Sticks, Principal Components and Spurious Significance" has been accepted for publication in Geophysical Research Letters, copyright 2005 American Geophysical Union (doi: 2004GL012750). A pre-publication version is at www.climate2003.com/pdfs/2004GL012750.pdf. Further reproduction or electronic distribution is not permitted.

Our article "The M&M Critique of the MBH98 Northern Hemisphere Climate index: Update and Implications" has been accepted for publication by Energy & Environment and is available at http://www.multi-science.co.uk/mcintyre-mckitrick.pdf.

Our research shows fundamental flaws in the "hockey stick graph" used by the Intergovernmental Panel on Climate Change (IPCC) to argue that the 1990s were the warmest decade of the millennium. The original hockey stick study was published by Michael Mann of the University of Virginia and his coauthors Raymond Bradley and Malcolm Hughes. The main error affects a step called principal component analysis (PCA). We showed that the PCA method as used by Mann et al. effectively mines a data set for hockey stick patterns. Even from meaningless random data (red noise), it nearly always produces a hockey stick.

This "backgrounder" provides a road map and summary of the 3 articles. While these papers have been under review, Mann et al. have opened up their own weblog and criticized some of our earlier work. We include some comments here on this commentary and some FAQ.

Natuurwetenschap & Techniek
The NWT article <http://www.natutech.nl/> presents a history of our original interest in the problem, proceeding through our first analysis of MBH98 in 2003, the responses by Mann et al., our counter-responses and so on up to the present. As part of their due diligence for the article, they showed our research to referees of their choosing for detailed assessment.

One was Dr. Rob van Dorland, IPCC Lead Author and climate scientist at the Dutch National Meteorological Agency. In the NWT article he is quoted saying it will "seriously damage the image of the IPCC." He added: "It is strange that the climate reconstruction of Mann has passed both peer review rounds of the IPCC without anyone ever really having checked it. I think this issue will be on the agenda of the next IPCC meeting in Peking this May."

Another was Dr Mia Hubert, a statistician at the Katholieke Universiteit Leuven in Belgium. She also agreed with our results, observing: "Tree rings with a hockey stick shape dominate the PCA with this method." Professor Hans von Storch, an IPCC Contributing Author and internationally-renowned expert in climate statistics at the Institute for Coastal Research in Geesthacht, Germany, concurred, calling McIntyre and McKitrick's criticism on this point "entirely valid."

Geophysical Research Letters

The GRL article, "Hockey Sticks, Principal Components and Spurious Significance" <http://www.climate2003.com/pdfs/2004GL012750.pdf> identifies what is almost certainly a computer programming error in the principal components method used in MBH98. The error causes their PC method to nearly always identify hockey stick shaped series as the "dominant pattern" in a data set (the so-called "first Principal Component" or PC1), even when the data are just random numbers. We carried out 10,000 simulations in which we fed "red noise", a form of trendless random numbers, into the MBH98 algorithm. In over 99% of the cases it produced hockey stick shaped PC1 series. The figure below shows 3 simulated PC1s and the MBH98 reconstruction: can you pick out the reconstruction?
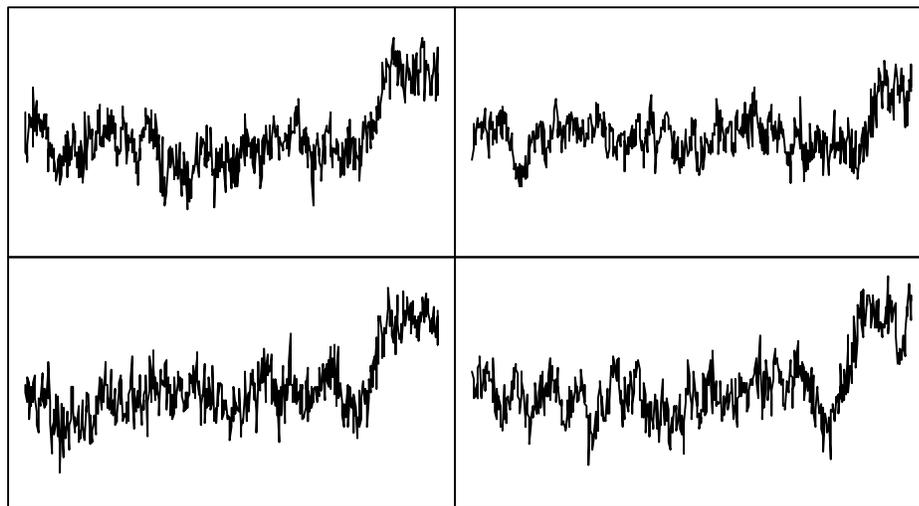

Figure 1. Three simulated PC1s and the MBH98 reconstruction.

The GRL article also examines the statistical significance of the MBH98 reconstruction in the controversial 15th century step. Although the Reduction of Error (RE) statistic is sometimes said to be "preferred" by paleoclimatologists, we show that it is necessary to look at more than one statistic before coming to a conclusion of statistical significance, because these simulated hockey sticks tend to have spuriously high RE statistics when compared to Northern Hemisphere temperatures during the calibration and verification periods of MBH98. The original article reported only one of the usual verification statistics used in paleoclimate studies. Mann et al. have not archived supporting data for this step and have refused to provide supporting calculations for this step. However, we report on emulations of MBH98, which indicate that the R^2 and other verification statistics for MBH98 lack statistical significance. Using more accurate benchmarks for the RE statistics under the actual conditions of MBH98, we show that the MBH98 lacks statistical significance in its early portion.

Elsewhere (<*http://www.realclimate.org/index.php?p=11*>) Mann et al. have insisted that climate reconstructions must pass a number of statistical verification tests or else they should not be considered. Applying that criterion, the MBH98 hockey stick graph should henceforth be considered invalid.

Energy & Environment

The *Energy & Environment* article, "The M&M Critique of the MBH98 Northern Hemisphere Climate Index: Update and Implications" <http://www.multi-science.co.uk/mcintyre-mckitrick.pdf > examines the effect of two methodological changes on the Northern Hemisphere climate index of MBH98: using the version of a Gaspé tree ring series as archived by the originating researchers rather than as emended by MBH; and using a conventional (centered) principal components methodology rather than the nonstandard methodology applied in MBH98. Our emulation methodology fully implements all the details reported in the July 2004 Corrigendum by Mann et al., including the actual steps for their stepwise principal components calculation and the gridcell standard deviations from the (now obsolete) HadCRU data previously unavailable. The result is, as we had found in our 2003 study, relatively high early 15th century values as shown in Figure 2 below.
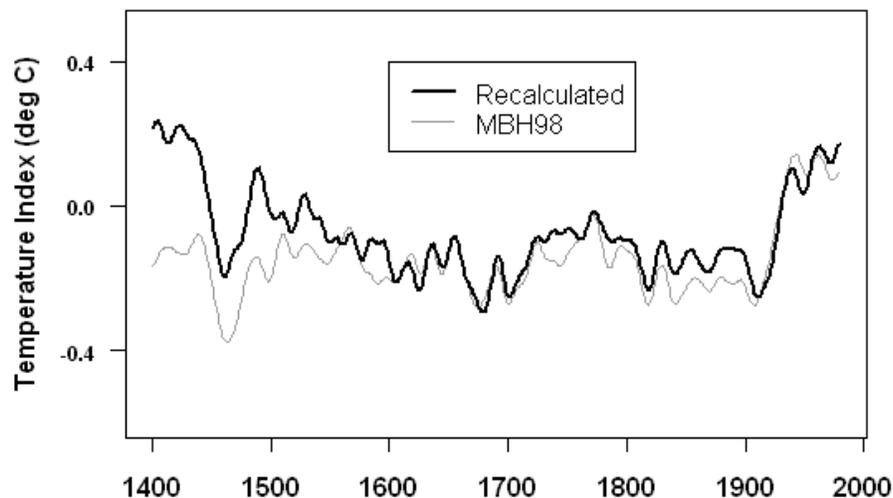

Figure 2. Northern Hemisphere temperature reconstructions: MBH98 and MM05 emulation of MBH98 using (a) centered principal components; (b) archived version of Gaspé series. Both series smoothed.

The E&E article canvasses various permutations and combinations of methodology now advocated by MBH98 in an attempt to avoid our criticism, showing how slight variations in methodology sometimes lead to high early 15th century results and sometimes to low early 15th century results. For example, the presence or absence of the PC4 in centered calculations determines early 15th century results. We point out that this is inconsistent with MBH98 claims that their method is "robust" to the presence or absence of dendroclimatic indicators.

The common factor in the bifurcation of 15th century temperatures is traced to the effect of bristlecone pines. There is an undoubted growth pulse for a small network of bristlecone pines in the Western USA in the late 19th and 20th centuries, and their ring width "chronologies" thus have a hockey stick shape. While these series represent only a small fraction of the MBH98 data base the resulting hockey stick shape is essentially an imprint of this bristlecone pine growth pulse. If world climate history is held to be uniquely characterized by the growth of bristlecone pines, we think that there should be a clear exposition of all the issues involved. We survey specialist literature on bristlecone pines and show that the original authors of the sites which dominate the MBH98 PC1 stated clearly that the 20th century growth pulse was **not** due to temperature. Even Mann's co-author, Hughes, has stated in print that the bristlecone pine growth pulse is a "mystery".

MBH98 data mining methods resulted in the selection of these hockey stick shaped series into the PC1, thereby ranking them as the "dominant pattern of variability", whereas they appear in neither the PC1 nor the PC2 under a centered calculation. Regardless of the role of the data mining method, it is important to recognize and fully assess the role of these series in imprinting a hockey stick shape and to completely assess whether they are the correct way of measuring world climate history.

FAQs

What is the connection of the present studies to your original article in 2003?

The present articles build on the first article. In that article, we pointed out that there were serious problems with the data set in MBH98 and, in particular, with the tree ring principal component series. We also showed that quite different results could be obtained for the 15th century under reasonable assumptions using the MBH98 method. Since then, and largely because of the effect of the original article, a great deal of new information about MBH98 has been made available. In July 2004, as required by Nature, Mann et al. published a Corrigendum, which included a voluminous archive on data and methods used in MBH98. Earlier, Mann et al. made public the address for the data actually used in MBH98, rather than the address which they had previously provided us. The present articles reflect detailed study of this new material. In particular, we are now able to precisely diagnose the problems with the principal component series in MBH98, which previously were simply noted as being incorrect.

We characterized our first article as raising "audit issues"; without co-operation from the original authors, it could not be a complete audit. In retrospect, the issues raised have proved to be very important. In MM03, we reported that we were unable to replicate the MBH98 principal component series. Given the importance of Mann's PC methods, one would think that this replication would have been attempted by someone else. In MM03, we were not in a position to fully diagnose the problems, but we are now. Also, the simple comparison of archived series versions to versions actually used revealed the unreported editing of the Gaspé series, which also had important consequences. This effect was not specifically analyzed in MM03, but was analyzed here.

In short, we believe that the present articles are a definitive resolution of issues first raised in MM03.

Has Michael Mann accepted any part of your arguments?

There is surprising agreement between Mann and ourselves on the effect of differing assumptions on the NH temperature reconstruction – if the assumptions are specified exactly. For example, both of us get high early 15th century results with centered PC calculations and 2 PCs in the AD1400 North American network and both of us get low early 15th century results with centered PC calculations and 5 PCs in the North American network. We have tried to canvass these matters in an evenhanded way in our E&E article to show what is agreed and what is not agreed.

Mann has categorically denied that his PC method generates hockey stick shaped series from red noise, but we see no way that he will able to sustain this argument, in the face of the compelling evidence to the contrary in our *GRL* paper.

We had thought that both Mann and ourselves agreed that a valid reconstruction should pass a range of statistical verification tests – this is a position which we endorse and a position which Mann seemed to endorse at realclimate.org. However, it seems that Mann et al. are inconsistent on this matter and are now arguing that you should only look at one verification test (the RE statistic) to establish significance. We doubt others would endorse this position, but in any case we show in the *GRL* paper that their reconstruction fails the RE test anyway.

Both Mann and ourselves have done calculations showing that the MBH98 reconstruction is not robust to the presence or absence of bristlecone pines: this is not under dispute.

We presume that Mann has done calculations showing that the MBH98 reconstruction is affected by their editing of the Gaspé tree ring series or else they wouldn't have done the editing. Strangely, the Gaspé series is used twice in the MBH98 model: it appears simultaneously in the North American ("NOAMER") network and in the "northern treeline" network. But only in the second usage is it edited, whereas in the NOAMER network the archived version was used.

Are you saying the 15th century was warmer than the present?

No, we are saying that the hockey stick graph used by IPCC provides no statistically significant information about how the current climate compares to that of the 15th century (and earlier). And notwithstanding that, to the extent readers consider the results informative, if a correct PC method and the unedited version of the Gaspé series are used, the graph used by the IPCC to measure the average temperature of the Northern Hemisphere shows values in the 15th century exceed those at the end of the 20th century.

Does your work disprove global warming?

We have not made such a claim. There is considerable evidence that in many locations the late 20th century was generally warmer than the mid-19th century. However, there is also considerable evidence that in parts of the Northern Hemisphere, the mid-19th century was exceptionally cold. We think that a more interesting issue is whether the late 20th century was warmer than periods of similar length in the 11th century. We ourselves do not opine on this matter, other than to say that the MBH results relied upon so heavily by the Intergovernmental Panel on Climate Change in its 2001 report are invalid.

Where can we get further information on principal components methodology?

In all our discussions, a principal component series is weighted combination of up to 70 individual tree ring series – some readers may find it helpful to think of the Dow-Jones Index, which is a weighted average of individual stock prices. Principal component series can include negative weights, which result in showing a *contrast* between different series – picture a series with positive weights for finance stocks and negative weights for tech stocks. In principal components discussions, the weights have forbidding names like eigenvectors or empirical orthogonal functions, but, at the end of the day, these are just weights. The decomposition is prescribed by the matrix algebra. There are canned programs in high level languages so that the principal components decomposition of a matrix X of time series can be obtained in one line. As we discuss in our articles, these decompositions can be highly sensitive to transformations of the data – even if the transformation only seems to be a "standardization". A slightly more technical exposition is available for download at http://www.uoguelph.ca/~rmckitri/research/trc.html.

Who paid for your research?

We have neither sought nor received funding for this work. For McKitrick, undertaking the project has required considerable time away from his own economics research. For McIntyre, undertaking this project has required an unpaid leave of absence from his career in mineral exploration financing, at the cost of over a year's foregone earnings so far.

Mann's Responses and Our Counter-Arguments

While the above articles have been under review and in press, Mann et al. have established a weblog at www.realclimate.org and have published a number of criticisms of our work. As described in the NWT article, there has been somewhat of an ongoing dialogue between Mann et al. and ourselves, since our original article in late 2003. The comments at realclimate.org are directed at our original article and towards a submission to Nature in early 2004 and re-state positions previously made in an Internet response by them to our first article and in correspondence with Nature. We were accordingly familiar with all of these positions by August 2004. Our new articles directly rebut the positions at realclimate.org. Since we anticipate that these positions may be re-iterated, we summarize comments on these responses below. Many of the points are quite technical.

1.   *Mann acknowledges that, with centered PC calculations, there is no longer a hockey stick pattern in the PC1, but points out that there is a hockey stick shape in the PC4 [where the bristlecone pine series get weighted](see http://www.realclimate.org/index.php?p=8). Mann argues that a "standard selection rule (Preisendorfer's Rule N)" supposedly used in MBH98 entitles them to use a larger number of PCs (5) in the AD1400 North American network with a centered PC calculation than with the uncentered method (2 PCs actually used in MBH98). Thus, even with a centered PC calculation, they argue that they can still "get" a hockey stick shaped reconstruction. They characterize our use of 2 PCs in the AD1400 North American network (the same number as used in MBH98) as "incorrect truncation" and "a failure to apply standard selection rules to determine the number of PC series than should be retained". From this, they argued that our criticisms of their methods do not "matter".*

We refer to this argument in passing in our E&E article (page 75, 2nd bullet) , where we specifically note that one gets MBH-type results when the PC4 is present, but MM-type results when the PC4 is absent. In their original response to MM03 http://stephenschneider.stanford.edu/Publications/PDF_Papers/MannEtAl2004.pdf, Mann et al. had argued that the North American PC1 contained the "dominant component of variance" in the North American dataset. However, we have shown that this so-called "dominance" is an artifact of the de-centering. Whereas the PC1 in the incorrect calculations accounted for over 38% of the variance in the North American data, under centered calculations, the PC4 accounts for only 8 % of the variance and is hardly "dominant". In fact, in our E&E article, we show that the MBH98 PC1 (now PC4) merely reflects the contribution of one species - the controversial bristlecone pine data in the western USA.

In MBH98 temperature calculations, their North American PC1 ends up dominating the results (see Figure 3 of our E&E article for a demonstration of this effect). Even if the pattern enters into the calculations from the PC4 position (as it would under a centered calculation), Mann et al. do not reduce the influence of this series and this relatively minor background pattern still dominates the final temperature calculations. It has virtually the same effect on NH temperature contributions from its position as a PC4, as it would have had as a PC1. Domination of calculations by a PC4 instead of a PC1 can hardly be said not to "matter".

This expedient also fails to deal with the defects of bristlecone pine growth as a so-called "proxy" for temperature, discussed in pages 81-86 of our E&E article. There is a definite 20th century pulse in bristlecone pine growth. However, there are explicit statements in specialist literature (surveyed there) that this pulse is not due to temperature and MBH co-author Hughes has stated that the pulse is a "mystery". Our E&E article discusses issues pertaining to bristlecone pine growth, showing that it is unacceptable that world climate history should be held to depend upon a PC4 made up of such controversial data.

We agree that the 8% contribution to variance in the centered North American PC4 is larger than the benchmark contribution to variance in a Preisendorfer-type calculation. But we do not agree that the pattern should be interpreted as the unique imprint of world climate history, overriding all other available data.

2. *Mann et al argue that they can "get" a hockey stick shaped reconstruction without using PCs at all, by directly using all 95 available proxies in the AD1400 network , thereby purporting to show that the issue is "spurious". (see http://www.realclimate.org/index.php?p=8 under Figure 2.)*

In this expedient, instead of increasing the number of PC series from 2 to 5, they propose to use 95 series in their regression module for the AD1400 step instead of 22 series used in MBH98 (in the 22 series, there are 3 PC series, which summarize 76 North American tree ring sites). There are 20 bristlecone pine sites, which are all used directly in this expedient, as opposed to merely being represented in one PC series - the PC1 (their system) or the PC4 (correct calculations). If one PC series can imprint the entire NH calculation, it will come as no surprise that 20 individual bristlecone pine series dominate the final results in the new calculation. Once again, the hockey stick shape is simply an imprint of bristlecone pine growth. If the calculation is done without bristlecone pines, no hockey stick results.

Moreover the geographical distribution is now completely implausible. MBH98 originally justified the use of PC methods as a means of achieving somewhat even geographical distribution of proxies, despite the over-representation of American tree ring data in the raw count. In this new expedient, instead of 7 of 22 series in the AD1400 network being American tree ring series, 80 of 95 proxy series are now American tree ring series.

It seems a little late in the day to be proposing a completely new (and non peer-reviewed) methodology to salvage results published in 1998. If this system were presented ab initio, we doubt that it would have been treated very seriously. As stated, it is simply another device for allowing the bristlecone pine data to drive the results. Moreover, since we have shown the original MBH98 failed statistical verification tests, as explained in our *GRL* article, we doubt this new method would pass either.

3. *Mann argues that they can "get" a hockey stick shaped reconstruction using a completely different method in Rutherford et al. [2005] and, thus, our critique doesn't "matter". http://www.realclimate.org/index.php?p=11; http://www.realclimate.org/index.php?p=10*

This refers to a forthcoming paper <http://www.realclimate.org/RuthetalJClim2004.pdf> in the *Journal of Climate*. None of the promised supporting calculations or supporting information for Rutherford et al. [2005] have been posted to date at fox.rwu.edu/~rutherfo/supplements/jclim2003a (Jan. 26, 2005). In any case this will in all likelihood simply be one more device for letting the bristlecone series dominate the results.

Two calculations pertinent to the present discussion are reported in Rutherford et al [2005]. In one, Rutherford et al. appear to have simply used the original MBH98 dataset, with the very PC series in dispute. The description (see section "multiproxy/PC dataset") refers to 112 indicators in the AD1820 network and 22 indicators in the AD1400 network. These are the same figures as MBH98 and require the use of PC series in the new calculations. While no details on the dataset as used have yet

appeared, there is an overwhelming probability that the PC series in the new calculations are the same as in the old calculations, and as such the new calculations are completely irrelevant to the issue of the effect of the PC methodology on final results.

The difference in calculations between MBH98 and Rutherford et al. [2005] appears to occur after the construction of the proxy networks. MBH98 appears to have done its calibration of proxies and estimation of past temperatures through linear regression. (Mann et al. have refused to provide source code for these calculations and, as discussed elsewhere, neither Nature nor the U.S. National Science Foundation have required the disclosure of this source code.) Rutherford et al. appears to have replaced the regression module with a new method ("RegEM"); in this case, source code is promised, but was not actually archived as of Jan. 26, 2005. Mann, Bradley and Hughes are co-authors of Rutherford et al.; in passing, we think that it would be more helpful if they archived the code for MBH98 before they worry about code for Rutherford et al. [2005]

A second calculation is described in section 4(B), where Rutherford et al. report that they get a hockey stick shaped result using all available proxy records in the calculation. This is precisely the same situation described in (2) above and all the responses apply here as well.

4. *Mann argues that other independent multiproxy studies get similar hockey stick shaped results (http://www.realclimate.org/index.php?p=11 see #1; http://www.realclimate.org/index.php?p=7, realclimate#Yet http://www.realclimate.org/index.php?p=98).*

We discuss this issue briefly on page 91 of our E&E article.

We point out that these "independent" studies are not "independent" as most people understand the word. Look at the co-authors of the new paper by Rutherford et al [2005]: Mann, Bradley, Hughes, Jones, Briffa and Osborn. These co-authors have had a hand in nearly every multiproxy study, be it Jones et al. [1998], Briffa et al. [2001], Mann and Jones [2003], Jones and Mann [2004], Briffa and Osborn [1999] or Mann, Bradley and Hughes [1998, 1999]. Earlier studies include Bradley and Jones [1993] and Hughes, Bradley and Diaz [1994].

In the E&E article, we cite a comment by Briffa himself that some proxies are common to nearly all these studies. The effect of these recurrent proxies needs to be closely analyzed: for example, what is the impact of bristlecone pine series on each of the other studies, either directly or through Mann's PC1? Additionally, the process of proxy selection is seldom discussed and yet there is a very real possibility of subconscious data mining. Null testing of the form carried out in our MBH98 red noise simulations needs to be done for every such study.

Unfortunately, many of these multiproxy studies lack adequate archival records, even though they were used by the IPCC. Crowley has said that he has "mis-placed" the original data for Crowley and Lowery [2000] (although he was able to find transformed and smoothed versions) and he was unable to recall where he got the digital information for the bristlecone pine versions used in his study. Jones was unable to provide the weights for the series in Mann and Jones [2003] as only Mann had that information. Briffa has not archived the 387 sites used in Briffa et al [2001] and has refused to provide the information. We believe that each of these studies needs to be examined individually by truly independent reviewers before any reliance can be placed on them as a whole. To our knowledge, no one has ever done so.

Most importantly, MBH98 must stand on its own merits. If it fails statistical significance tests or robustness tests, other multiproxy studies cannot save it (although the IPCC might now choose to try to rely on these other studies rather than MBH98.)

5. *Mann says that their reconstruction has been proven to have statistical skill using the 'RE' statistic, the "preferred" statistic for climatologists, while our reconstruction "fails statistical verification exercises, rendering it statistically meaningless and unworthy of discussion in the legitimate scientific literature."(http://www.realclimate.org/index.php?p=11; http://www.realclimate.org/index.php?p=8)*

While climatologists may "prefer" the RE statistic, virtually all paleoclimatological literature recommends the use of several verification statistics (e.g. Cook, Briffa and Jones [1994], an article which includes two co-authors of Rutherford et al. [1995]). Eduardo Zorita, a climatologist interviewed by NWT and frequently cited by Mann, said that the R2 statistic should certainly have been examined as well. These recommendations would seem to have particular weight when the study is being used for major policy decisions and when the study uses a nonstandard methodology that mines for hockey stick shaped series. Because the RE statistic has no known distribution, significance benchmarks must be computed by numerical simulation. Therefore, looking at related test statistics (such as R2), for which there are known distributions and published tables for significance levels, is a simple way to cross-check significance, in case the RE benchmarks were computed incorrectly or were spurious for any other reason. In this case, a simple check of the R2 statistic would have revealed major problems.

In our GRL article, we provide compelling evidence that the early portion of Mann's own reconstruction fails an important and standard statistical verification test (R2), as well as other tests. We also show that the benchmark for RE significance was incorrectly calculated and, using the new benchmarks, even the RE statistic of the controversial early portion of MBH98 is not statistically significant.

We did not put forward "our" version of MBH98 as a climate history, but to show the effect of correct calculation of principal component series. We concur that an MBH98-type reconstruction with correctly calculated PC series ("our" version) lacks statistical significance. But more importantly, the *original MBH98 version also lacks statistical significance* (as do any of the new expedients now proposed to salvage MBH98). By Mann's own criterion that makes it "meaningless and unworthy of discussion in legitimate scientific literature."

6. *Rutherford et al. [2005] says that some of the errors which you reported were based on the use of an incorrect version of the MBH98 dataset.*

The analyses presented in the present papers are based on data located at Professor Mann's FTP site at the location ftp://holocene.evsc.virginia.edu/pub/MBH98 and the July 2004 Corrigendum by Mann et al. As a result of our initiative, Mann et al. have already published one Corrigendum, admitting that the original Nature Supplementary Information listed over 35 series which were not actually used in MBH98 calculations. If Rutherford et al. [2005] is suggesting that the datasets at these locations are *still* incorrect, then MBH98 should obviously be retracted.

In 2003, Rutherford advised us that the data used in MBH98 was located at ftp://holocene.evsc.virginia.edu/pub/sdr/pcproxy.txt. When we noticed problems, we asked Mann for

confirmation that this was the data set actually used in MBH98, but he said that he was too busy to respond to this or any other question. After publication of McIntyre and McKitrick [2003], a new URL was made public for (to our knowledge) the first time. The collation errors in ftp://holocene.evsc.virginia.edu/pub/sdr/pcproxy.txt are not present in the files at ftp://holocene.evsc.virginia.edu/pub/MBH98, but all the other errors reported in M&M [2003] are repeated in the files at the new URL – not least the incorrect calculation of principal component series. Some of the errors reported in MM03 are amusing and have not been corrected even in the new Nature SI: for example, MM03 reported that a French precipitation series was incorrectly located in New England. Nearly all the MBH98 precipitation series have been incorrectly located; none of these locations have been corrected in the new Nature SI, even though Mann et al. (and Nature) are aware of the inaccuracies.

7. *Rutherford et al. [2005] says that the reconstruction of McIntyre and McKitrick [2003] is flawed because it failed to implement a stepwise PC procedure used in MBH98.*

All calculations in the present papers have implemented the stepwise procedure for PC calculation implied by the schedule of PCs shown in the *Nature* Corrigendum Supplement. The criticism is irrelevant to any calculations carried out since November 2003.

The stepwise procedure was not described in the original article and was not implemented in our 2003 article. Subsequently, there has been much inconsistency in descriptions of this procedure by Mann et al. In an Internet article, they said that they used 159 series altogether, rather than 112. The use of 159 series is not demonstrated at the Corrigendum SI, but rather 139 series. They have also said that PC series were re-calculated for each network for each calculation step. This is also incorrect.

This matter is discussed in the NWT article. In any event, it is completely irrelevant to the issues of PC method, statistical significance and bristlecone pines.

8. *Mann has said that your results are flawed, because they can emulate your high early 15th century results by censoring 70 North American tree ring series.*

We discuss this issue on page 88 of our E&E article. It is quite true that you get high early 15th century results if you exclude 70 North American tree ring series – because the bristlecone pines are thereby excluded. Indeed if you carry out calculations with 50 of the 70 North American series (only excluding the 20 controversial bristlecone pine series), you get high early 15th century results. So it's the presence or absence of the bristlecone pines that cause the effect. Mann et al. were well aware of this effect, since they had carried out this exact calculation, but did not report it.

The calculations in the present articles do NOT censor any series. The PCs are calculated using all the underlying data. However, when correct PC methods are used, the bristlecone pines are demoted to the PC4, from the PC1 and do not affect the calculations if only 2 PCs are used, as in the original study. Mann et al. call this "effective censoring" – which is obviously untrue since it is the PC weighting algorithm that assigns the role for the bristlecone series. In fact, because they themselves used a data mining method, their PC1 consisted only of 14 bristlecone pine sites, "effectively censoring" the rest of the North American network.

9. *Mann et al. state that our claims have "additionally been discredited in a recent peer-reviewed article by Rutherford et al (2004)" http://www.realclimate.org/index.php?p=8; http://www.realclimate.org/index.php?p=10; http://www.realclimate.org/index.php?p=11 #1.*

   Rutherford et al. [2005] does not consider any of the observations made here about the flawed PC method, statistical significance or the effect of bristlecone pines. In question #3, we showed that their attempts to salvage MBH98 results were themselves flawed, because one calculation simply re-used the PC series in dispute and another calculation abandoned any attempt at even geographical distribution. Both merely try to insert the bristlecone pine imprint through a back door.

   Rutherford et al., in the preprint currently available (Jan. 26, 2005) does not provide a suite of verification statistics for the controversial 15th century step. The preprint only contains RE statistics for the steps; the CE statistic is promised at Rutherford's website, but is not currently available. (Jan. 26, 2005). This article contains this extraordinary claim:

   > "To aid the reader in interpreting the verification diagnostics and to illustrate the shortcomings of R2 as a diagnostic of reconstructive skill, we provide some synthetic examples which show three possible reconstructions of a series and the RE, CE and r2 scores for each".(page 26 of preprint).

   These synthetic examples are not yet available (Jan. 26, 2005). From our own experience with 10,000 simulated PC1s, the MBH98 method for computing PC1s yields high RE values on machine-generated noise, so this is obviously a spurious effect. If the benchmarks are not calculated correctly (as in MBH98), spurious significance can easily be attributed to statistically insignificant reconstructions.

   We also believe that, if the statistical significance of MBH98 rests on a disproof of the validity of the well-known R2 statistic, original readers were entitled to a clear report of the R2 values and an exposition by MBH as to why the R2 statistic should be disqualified. It is conspicuous that they did not try to establish such a position at the time, and it is implausibly late for them to try now.

10. *Mann et al. state that "the use of non-centered PCA is well-established in the statistical literature and, in some cases is shown to give superior results to standard, centered PCA" http://www.realclimate.org/index.php?p=98. They go on to cite two studies.*

    First, we note that MBH98 stated that they used "conventional" PCA. We have elsewhere pointed out that "conventional" PCA calculations are "centered" – a position acknowledged here by Mann. Mann et al. are obviously quite free to argue the merits of non-centered PCA for this particular calculation – an argument which we believe will be unsuccessful –but they should have stated that this was what they were doing in the first place, and perhaps issue another Corrigendum in which they report that the description of their PCA method in MBH98 was inaccurate.

    The first source cited by Mann is course notes for an ecology course by Dean Urban at Duke. Urban cites Pielou [1984] as authority for use of non-centered PCA in ecology; Pielou in turn cites Noy-Meir [1973]. The ecological context for non-centered PCA in these studies was in determining patterns in counts of ecological species at different sites along a gradient. In these studies, zero did not represent an arbitrary location along a scale (such as Centigrade or Fahrenheit), but a physical count of 0. Noy-Meir [1973] pointed out that uncentered PCA could be used to isolate the effect of

individual species (rather than an overall effect). Later ecological literature has tended to be quite critical of PCA as an ordination method.

The second presentation cited by Mann is a Powerpoint presentation on the Internet by Jolliffe (a well known statistician).

> Jollife explains that non-centered PCA is appropriate when the reference means are chosen to have some *a priori* meaningful interpretation for the problem at hand. In the case of the North American ITRDB data used by MBH98, the reference means were chosen to be the 20th century calibration period climatological means. Use of non-centered PCA thus emphasized, as was desired, changes in past centuries relative to the 20th century calibration period. (http://www.realclimate.org/index.php?p=98)

In fact, Jolliffe says something quite different. Jolliffe's actual words are:
> "it seems unwise to use uncentered analyses unless the origin is meaningful. Even then, it will be uninformative if all measurements are far from the origin. Standard EOF analysis is (relatively) easy to understand –variance maximization. For other techniques it's less clear what we are optimizing and how to interpret the results. There may be reasons for using no centering or double centering but potential users need to understand and explain what they are doing."

Jolliffe's presents cautionary examples showing that uncentered PCA gives results that are sensitive to whether temperature data are measured in Centigrade rather than Fahrenheit, whereas centered PCA is not affected. Jolliffe nowhere says that an uncentered method is "the" appropriate one when the mean is "chosen" to have some special meaning, he states, in effect, that having a meaningful origin is a necessary but not sufficient ground for uncentered PCA. But he points out that uncentered PCA is not recommended "if all measurements are far from the origin", which is precisely the problem for the bristlecone pine series once the mean is de-centered, and he warns that the results are very hard to interpret. Finally, Jolliffe states clearly that any use of uncentered PCA should be clearly understood and disclosed - something that was obviously not the case in MBH98. In the circumstances of MBH98, the use of an uncentered method is absolutely inappropriate, because it simply mines for hockey stick shaped series. Even if Mann et al. felt that it was the most appropriate method, it should have had warning labels on it.

13