

Stephen McIntyre and Ross McKittrick
October 29, 2003

Professor Mann's response to our paper is posted at <http://www.davidappell.com/archives/00000377.htm>

This does not appear to be the 'full response' so we are reluctant to engage it until we have a more complete idea of the rebuttal Professor Mann will present. However, since many people have asked for our feedback on what is currently there, we are proceeding as we can.

The story at Mr. Appell's page reads as follows:

The McIntyre and McKittrick ("M&M") *Energy and Environment* [paper](#) criticizing the work of Mann, Bradley, and Hughes is wrong, says Mann.

In short, here's what happened: M&M asked an associate of Mann to supply them with the Mann et. al. proxy data in an Excel spreadsheet, even though the raw data is [available here](#). An error was made in preparing this Excel file, in which the early series were successively overprinted by later and later series, and this is the data M&M used. Mann says:

"...the authors results are entirely spurious. The mistake made insures that the estimates, in particular prior to 1600-1700, are meaningless."

This leads, Mann says, to "the use of series that are artificial combinations of early [e.g. 15th-16th century] and late [e.g. 19th-20th] information accidentally spliced together" with "no relation" to the proxy data used by Mann et. al. in their 1998 (and subsequent) work.

Mann adds:

The spreadsheet file they used was a complete distortion of the actual Mann et. al. proxy data set, and was essentially useless, particularly in the earlier centuries. The authors had access to the full data, which has been available on a public ftp site for nearly two years. When they noticed, as described in their paper, some signs of problems with the Excel spreadsheet version of the data, one might think that they would have bothered to check the data available on our public ftp site.

Reply:

We still await a proper response. The above does not begin to address our critique, it ignores a crucial step in our analysis and it gets some key facts wrong.

Mann says that M&M asked an associate of Mann to supply them with the Mann et. al. proxy data in an Excel spreadsheet, even though the raw data is [available here](#).

We did not ask for an Excel spreadsheet nor did we receive one. And we did not approach Mann's associate, Mann did. Here is the actual correspondence.

Apr. 8, 2003	<p>Dear Dr. Mann,</p> <p>I have been studying MBH98 and 99. I located datasets for the 13 series used in 99 at ftp://eclogite.geo.umass.edu/pub/mann/ONLINE-PREPRINTS/Millennium/DATA/PROXIES/ (the convenience of the ftp: location being excellent) and was interested in locating similar information on the 112 proxies referred to in MBH98, as well as listing (the listing at http://www.ngdc.noaa.gov/paleo/ei/data_supp.html is for 390 datasets, and I gather/presume that many of these listed datasets have been condensed into PCs, as mentioned in the paper itself. Thank you for your attention.</p> <p>Yours truly,</p> <p>Stephen McIntyre, Toronto, Canada</p>
Apr. 9, 2003	<p>Dear Mr. McIntyre,</p> <p>These data are available on an anonymous ftp site we have set up. I've forgotten the exact location, but I've asked my Colleague Dr. Scott Rutherford if he can provide you with that information.</p> <p>best regards,</p> <p>Mike Mann</p>
Apr. 11, 2003	<p>Dear Dr. Rutherford,</p> <p>Prof. Mann indicated that you would be able to locate the ftp for the MBH proxies and I would appreciate it if you could forward me the information. Thank you for your attention.</p> <p>Yours truly, Steve McIntyre</p>
Apr. 11, 2003	<p>Steve,</p> <p>The proxies aren't actually all in one ftp site (at least not to my knowledge). I can get them together if you give me a few days. Do you want the raw 300+ proxies or the 112 that were used in the MBH98 reconstruction?</p> <p>Scott</p>
Apr. 11, 2003	<p>[Scott]</p> <p>I'd appreciate it. The 112 would be the pertinent ones; is there a Readme showing the reduction of the 300+ to the 112?</p> <p>Thanks, Steve McIntyre</p> <p>PS If I do any subsequent organizing of the data which would make it more accessible to the next user, if you'd like, I'll re-transmit that to you.</p>
Apr. 22, 2003	<p>[Scott]</p> <p>Any luck with this?</p> <p>Steve McIntyre</p>
Apr. 22, 2003	<p>Steve,</p> <p>I was just going to send you an e-mail. Honest. I haven't forgotten about you. Bear with me a little longer. Your request is just about at the top of the stack.</p> <p>Regards,</p> <p>Scott</p>
~Apr. 23, 2003	<p>Scott Rutherford sends file "pcproxy.txt".</p>

Note that Dr. Rutherford says the proxies aren't all in one location, which conflicts with Professor Mann's claim that they were all readily available at the Virginia FTP site.

Mann says:

An error was made in preparing this Excel file, in which the early series were successively overprinted by later and later series, and this is the data M&M used.

No, this is not the data we used. We rebuilt the proxy data set from scratch using the sources he listed online. We will return to this below.

And note the passive tense: "An error was made." Professor Mann is accusing his associate of a very clumsy set of mistakes. But then again we did not ask for an Excel file, and what we received was a plain text file. We do not know what Excel file Mann is referring to. If someone prepared an Excel version of his data base we never received it.

Mann says:

"...the authors results are entirely spurious. The mistake made insures that the estimates, in particular prior to 1600-1700, are meaningless."

This leads, Mann says, to "the use of series that are artificial combinations of early [e.g. 15th-16th century] and late [e.g. 19th-20th] information accidentally spliced together" with "no relation" to the proxy data used by Mann et. al. in their 1998 (and subsequent) work.

Professor Mann is asserting that the data file which was sent by his associate contains accidental splices of early and late data. While there is evidence of many errors in the Mann data, this does not seem to be one of them. We have made the data file available at our websites in the original form. Perhaps Mann can identify the series numbers in which his associate made these mistakes.

We also cross-checked many Mann datasets against correct source data from the World Data Center for Paleoclimatology. While we noticed many differences between the datasets, there were also many similarities. We have provided graphs comparing the two versions on a series-by-series basis and there is no evidence in these examples that Mann's associate made the error that Mann accuses him of. Further, in our own compilation, wherever possible, we substituted data from the World Data Center for Paleoclimatology for Mann's data.

The story goes on:

Mann adds:

The spreadsheet file they used was a complete distortion of the actual Mann et. al. proxy data set, and was essentially useless, particularly in the earlier centuries. The authors had access to the full data, which has been available on a public ftp site for nearly two years. When they noticed, as described in their paper, some signs of problems with the Excel spreadsheet version of the data, one might think that they would have bothered to check the data available on our public ftp site.

Contrary to what he implies we did check with Michael Mann and Scott Rutherford about some of the problems in the data. In fact we emailed the whole file back to Professor Mann and asked him to verify we had the right data. Here is the correspondence.

<p>Sept. 9, 2003</p>	<p>Dear Prof. Mann,</p> <p>I have tried diligently to reconstruct your temperature principal components as described in MBH98, but without success and would appreciate some assistance.</p> <p>I downloaded hadcrut2.dat from CRU (July 2003 edition), truncated the data to 1902-1995 and further truncated it to the 1082 cells at gridpoints.loc and arranged as 1082 time-series with 1128 monthly readings. This step was successful as I could match your map of cell locations. I standardized each series to mean 0 and sd 1 for the period 1902-95. In MBH98, you say that you carried out “conventional” PCA, but there is so much missing data that conventional PCA failed when I tried. In particular, 4 cells had no values at all and I don’t see why they were included in your selection. Most PCA algorithms balk at missing data or exclude it. How did you deal with the extensive missing data?</p> <p>I downloaded the EOFs, PCs and eigenvector loadings from ftp://eclogite.geo.umass.edu/pub/mann/MANNETAL98/EIGENVECTORS/. I spliced the EOFs into a 16x1082 matrix and the PCs (pc01.out, etc.) into a 92x16 matrix. I made a diagonal of the first 16 values in column 2 of “tpca-eigenvals.out”, which look like eigenvalues, and carried out an expansion. I then deducted the grid-box values generated from this expansion from the Jones data as above; calculated variance for each year across available cells and made a sum, comparing this to the variance similarly calculated in the standardized Jones data. I obtained very low/much lower explained variance from this than you got. I also tried some experiments and it also doesn’t seem to me that the first 16 EOFs maximize explained variance, as they should. I would appreciate any assistance or clarification which you could give.</p> <p>Regards, Steve McIntyre</p> <p>Note: No reply was received.</p>
<p>Sep. 10, 2003</p>	<p>Dear Scott, In the file pproxy.txt which you sent me a while ago, a disproportionate number of series start in the years *99 or *49, for example, series 73-83, 86-92 and 106-107. Is this intentional? Is this also the case in the underlying data? Thanks, Steve</p>
<p>Sep. 11, 2003</p>	<p>Steve,</p> <p>These would be the principal components (PCs) of dense tree ring networks. As trees drop out of the network the PCs are recalculated every 50 or 100 years. So I think what is going on is that once the tree network is too sparse, Mike stopped calculating the PCs at the last step where there was a decent network. (This was before my time, but I'm pretty sure that's story.)</p> <p>Scott</p>
<p>Sep. 25, 2003</p>	<p>Dear Prof Mann</p> <p>Here is the pproxy.txt file sent to me last April by Scott Rutherford at your direction. It contains some missing data after 1971. Your 1998 paper does not describe how missing data in this period is treated and I wanted to verify that it is the correct file. How did you handle missing data in this period? In earlier periods, it looks like you changed the roster of proxies in each of the periods described in the Supplementary Information using only proxies available throughout the entire period. I have obtained quite close replication of the rpc1 in the 20th century by calculating coefficients for the proxies and then calculating the rpc's using the minimization procedures described in MBH98 and the selection of PCs in the Supplementary Information. The reconstruction is less close in earlier periods. I also don't understand the reasoning for reducing the roster of eigenvectors in earlier periods. The description in MBH98 was necessarily very terse and is still very terse in the Supplementary Information; is there any more detailed description of the reconstruction methodology to help me resolve this? Thank you for your attention.</p> <p>Yours truly,</p> <p>Steve McIntyre,</p> <p>Toronto, Canada</p>

Sep. 25,
2003

Dear Mr. McIntyre,

A few of the series terminate prior to the nominal 1980 termination date of the calibration period (the earliest such instance, as you note, is 1971). In such cases, the data were continued to the 1980 boundary by persistence of the final available value. These details in fact, were provided in the supplementary information that accompanied the Nature article. That information is available here (see first paragraph):

<ftp://eclogite.geo.umass.edu/pub/mann/ONLINE-PREPRINTS/MultiProxy/data-supp.html>

and here:

http://www.ngdc.noaa.gov/paleo/ei/data_supp.html

The results, incidentally, are insensitive to this step; essentially the same reconstruction is achieved if a calibration period terminating in 1970 (prior to the termination of any of the proxy series) was used instead.

Owing to numerous demands on my time, I will not be able to respond to further inquiries. Other researchers have successfully implemented our methodology based on the information provided in our articles [see e.g. Zorita, E., F. Gonzalez-Rouco, and S. Legutke, Testing the Mann et al. (1998) approach to paleoclimate reconstructions in the context of a 1000-yr control simulation with the ECHO-G Coupled Climate Model, *J. Climate*, 16, 1378-1390, 2003.]. I trust, therefore, that you will find (as in this case) that all necessary details are provided in the papers we have published or the supplementary information links provided by those papers.

Best of luck with your work.

Sincerely,

Michael E. Mann

Note that we asked Mann to verify that we had the correct file, and he said nothing in reply about checking it against an on-line source. Moreover he cut off further inquiries. He can hardly object now that we did not inquire further.

Professor Mann's reply only raises more questions. Why did the data file have to be assembled from scratch? Did he not have a copy for his own work? Has no one ever asked for it before? Is he accusing his associate, Scott Rutherford, of inserting all the fills? And if what we received was "a complete distortion," and bears "no relation" to the data set he used, how were we able to replicate his original results so closely?

While the claim implicit in Professor Mann's defence is that he actually did work from correctly collated data file in his 1998 paper, this still fails to address the substantial problems of obsolete series, mislabeled locations, truncation of sources, extrapolations of missing data, use of JJA data where annual are available etc.

He is now pointing to a url (<ftp://holocene.evsc.virginia.edu/pub/MBH98/>) that was not given to us when we first asked. And this site only contains individual series, not a collated set. For a study that has been this influential it is not enough to claim that researchers could, in principle, find the data in an unprocessed form on a web site other than the ones known to be associated with the paper and which he declines to identify, not only to people requesting the data but even to his associate on whom he relies to manage the data (who, recall, said the data are spread across several sites and it took him 12 days to compile them). In the private sector there is a rule of "full, true and plain disclosure." This falls short, to say the least.

As to the claim that the discrepancy can be entirely blamed on our use of the wrong data, Professor Mann is forgetting that after completing the audit of what he sent *we built a new data base*. If errors were made

by Scott Rutherford of the type described by Mann, then they affect series #69-99 in the file pproxy.txt and do not affect any of the other files. In our reconstruction, we have completely recalculated series #69-80 and series #84-99 from original (and reliable) World Data Center for Paleoclimatology data. Accordingly, our recalculation only contains the alleged Scott Rutherford data for series #81-83, which extend back to 1450, 1600 and 1750 respectively and do not affect the early 15th century calculations. We are in the process of reconciling this file against the additional FTP references now provided by Mann.

But now there is a confusing extension to this story which is also posted at David Appell's web site. It goes like this (see especially the highlighted text below).

M&M: The Details

Apparently Mann had told McIntyre last April that his raw data were available on his FTP site, according to this email (shared with permission):

Date: Wed, 09 Apr 2003 05:58:29 -0400
To: Steve McIntyre
From: "Michael E. Mann"
Subject: Re: Proxies in MBH
Cc: Scott Rutherford

Dear Mr. McIntyre,

These data are available on an anonymous ftp site we have set up. I've forgotten the exact location, but I've asked my Colleague Dr. Scott Rutherford if he can provide you with that information.

best regards,

Mike Mann

At 01:47 PM 4/8/2003 -0400, Steve McIntyre wrote:

Dear Dr. Mann,

I have been studying MBH98 and 99. I located datasets for the 13 series used in 99 at <ftp://eclogite.geo.umass.edu/pub/mann/ONLINE-PREPRINTS/Millennium/DATA/PROXIES/> (the convenience of the ftp: location being excellent) and was interested in locating similar information on the 112 proxies referred to in MBH98, as well as listing (the listing at http://www.ngdc.noaa.gov/paleo/ei/data_supp.html is for 390 datasets, and I gather/presume that many of these listed datasets have been condensed into PCs, as mentioned in the paper itself. Thank you for your attention.

Yours truly,

Stephen McIntyre,
Toronto, Canada

This quickly gets into some hairy details, but **Mann says that the crux of M&M's error is their use of a Excel dataset with only 112 columns (where each column represents one set of proxy data--tree rings, ice cores, historical temperature data, etc.), when in fact the full paleoclimatic data series requires 159 to be used properly in the analysis behind the Mann, Bradley, and Hughes 1998 paper whose results they're trying to check.**

(emphasis added). Note in the above paragraph, the “crux of the error” according to Mann is our use of an Excel spreadsheet with 112 columns, “when in fact the full paleoclimatic data series requires 159 to be used properly in the analysis behind the Mann, Bradley and Hughes 1998 paper.”

This is something new altogether. Where has it ever been disclosed there were actually 159 proxies behind MBH98? Their *Nature* paper (see for example page 781, “based on the full multiproxy network of 112 indicators” etc.) and all supporting material only mention 112. The data file we were sent only contained 112 series. Scott Rutherford’s April 11 email refers to “Do you want the raw 300+ proxies or the 112 that were used in the MBH98 reconstruction?” The file we sent back to Mann only contained 112. In the Zorita paper Mann refers to above they only refer to 112 proxies too (p. 1379.)

It seems odd that 5 years after the publication of the paper it is now revealed for the first time that there were 159 proxies, not 112.