

## 2

THE MANN ET AL.  
NORTHERN HEMISPHERE  
“HOCKEY STICK” CLIMATE INDEX

## A TALE OF DUE DILIGENCE

*Ross McKittrick*

This chapter tells the story of the detective work of Stephen McIntyre (and, to a lesser extent, myself) regarding the famous “hockey-stick” climate history graph of Mann, Bradley, and Hughes (1998), better known as MBH98.<sup>1</sup> After studying in detail how the hockey-stick graph was done, we found mistakes in the data and methods that went unnoticed for years, even as the graph was used by governments worldwide to drive major policy decisions. The story behind the hockey stick provides a cautionary tale about the need to recognize the limited function of journal peer review and the dangers of proceeding with major policy decisions without applying a further level of due diligence equivalent to an audit or an engineering study. It also shows that the Intergovernmental Panel on Climate Change failed to carry out elementary due diligence on its most famous promotional graphic, despite widespread perceptions that it had.

In 1998, Michael Mann, Raymond Bradley, and Malcolm Hughes (hereafter MB&H) published a paper in *Nature* called “Global-scale temperature patterns and climate forcings over the past six centuries,” commonly called MBH98. In this study, they proposed some seemingly novel ways of calculating the average Northern Hemisphere temperature back to AD1400. In 1999, MB&H published a follow-up paper, which is commonly called MBH99, extending MBH98 back four hundred years to AD1000.<sup>2</sup> MBH99 did not recalculate post-1400 values; it simply extended the previous results to an earlier period. This is Mann’s famous hockey-stick curve (figure 2.1).

This graph achieved notoriety courtesy of the Intergovernmental Panel on Climate Change (IPCC), appearing in figures 2-20 and 2-21 in

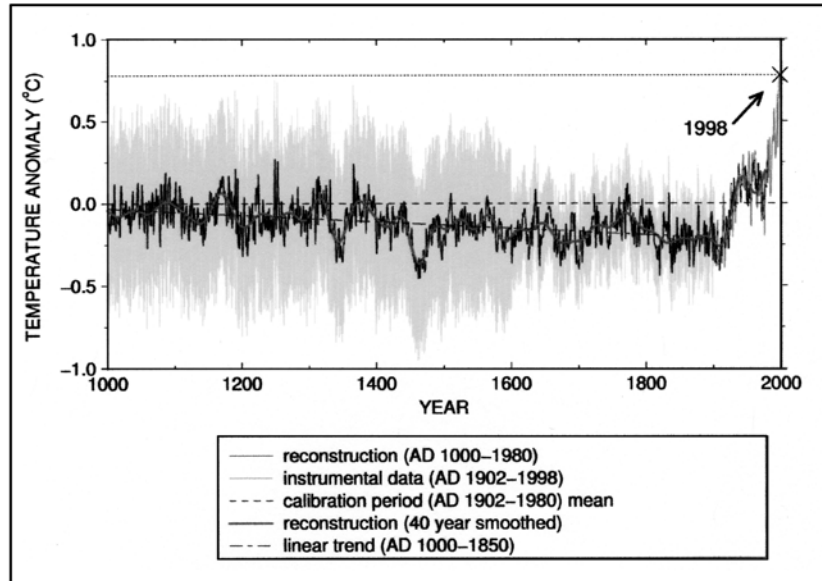


Figure 2.1.

chapter 2 of the 2001 Working Group 1 Assessment Report, figure 1b in the Working Group 1 Summary for Policymakers, figure 5 in the Technical Summary, and figures 2-3 and 9-1B in the Synthesis Report. The IPCC Summary for Policymakers (3) used this figure to claim that it is likely “that the 1990s has been the warmest decade and 1998 the warmest year of the millennium” for the Northern Hemisphere. The hockey-stick graph has been reprinted countless times and used by governments around the world as the official, canonical climate history of the world.

MBH98 was conspicuous for its obscurity. The appendix to that paper, which spells out the statistical procedures, is very hard to read. It is written in grandiose yet disorganized prose and omits the mathematical equations that would allow expert readers to attain an unambiguous understanding of what was done. For all the subsequent usage of the results of that paper, it strikes me as conspicuous that the methods of the paper have not been widely applied. Even Mann himself did not use its method in his recent publications.<sup>3</sup> My suspicion is that most readers cannot make heads or tails of Mann’s methodology. Nevertheless, the hockey-stick graph was an instant hit. Apparently no one ever checked it, and yet it was used by institutions, governments, and the media to promote the Kyoto Protocol.

## THE AUDIT: APRIL TO SEPTEMBER 2003

Hockey-stick graphs are notorious promotional tools for people in business and finance: dot-com promotions typically have hockey-stick shapes to describe future revenue and sales projections. After the dot-com boom, however, many businesspeople cringe when they see a hockey-stick graph.

In late 2002, the Canadian government was trying to sell the Kyoto Protocol to the public and was continually citing Mann's hockey stick as one of its central arguments for supporting the costly investments needed to implement Kyoto. At the time, Stephen McIntyre was a fifty-five-year-old Canadian businessman involved in financing speculative mineral exploration. In that business, McIntyre knew the importance of effective promotional graphics in raising money, but he also knew enough to be wary of them. He had dealt regularly with geologists, one of whom had pointed out to him that the climate change in recent times was far less than that observed in geological history. When Stephen looked at Mann's hockey stick, the graph immediately struck him as a promotional graphic of the type that he was used to seeing in business. At the time, the speculative mining business was slow, and he thought it would be interesting to see how this particular promotion was designed.

Note that this impulse was no more than the intuitive suspicions of someone with long experience in the speculative mineral exploration business. That was the start of the whole adventure. By no means did McIntyre begin his inquiry as a hatchet job on Mann and his coauthors. On the contrary, it never occurred to him that anything would result from his efforts other than figuring out a few things of personal interest. Nor was this an exercise in pouncing on small errors. The resulting critique centers on the lack of transparency in the original work, which is what made it so difficult to find the errors therein, and the overhasty promotion of the graph by governments and others who were making no attempt to check the calculations. That the errors we found were material to the results only highlights the need for greater transparency at the journal level, and more rigorous due diligence by users of the research, points to which I will return in the concluding section.

One thing Stephen McIntyre learned from the mining business was that it is important to look at the drill core—you need to see the raw data, not just the promoter's version. In the famous Bre-X fraud of 1998, the promoters said they had a "special" analyzing technique that required the entire core—so that visiting analysts were unable to inspect the original core. In this case, Stephen concluded that the equivalent to the "drill core"

would be the individual proxy series. He decided to plot up these data to see what they looked like.

Although the long IPCC graph came from MBH99, the methods were developed in MBH98, so the results could not be analyzed without looking at the earlier study. After searching the relevant FTP and websites, Steve was able to locate the data for fourteen series in MBH99, but not for MBH98. So on April 8, 2003, on a whim, he wrote to Michael Mann and asked him where the data could be obtained, not expecting any special effort on his behalf, merely presuming that the data had been neatly buttoned up for previous due diligence by the IPCC and others. Mann replied that it was on an FTP site they had set up. Mann added, "I've forgotten the exact location, but I've asked my colleague Dr. Scott Rutherford if he can provide you with that information." A few days later Steve queried Rutherford, who responded that the data "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." Eleven days later, Steve received the FTP location of a text file, `pcproxy.txt`, which, after downloading, was found to contain 112 columns of data, matching the number of proxies identified in the *Nature* paper.

The episode immediately caught his attention. He had assumed that the IPCC and others had carried out due diligence prior to relying on this graphic and that such due diligence would have necessarily required examination of the data—drawing from his own experience with audits and business due diligence. If the data had not been so organized, was it possible that no one had ever checked the data? It was a bizarre possibility, but the collective failure of due diligence in the Bre-X (and, for that matter, Enron) collapses were just as strange. People in the mineral exploration business are used to strange events. And someone accustomed to public markets knows about the "madness of crowds."

So instead of simply trying to plot up the proxy data and see what it looked like, Steve decided to do something much more ambitious. He decided to try to verify MBH98 by replicating its calculations. In later months I was sometimes asked for a brief description of my coauthor, and inevitably the people doing the asking were puzzled about the answer: a middle-aged mining businessman. What was a mining executive with no specialized training doing walking into a high-level scientific debate with international policy implications? However the situation was remarkably propitious. As far as credentials go, Steve had been a prize-winning student in math and statistics at the University of Toronto and had won a Ph.D. scholarship offer from MIT. But for a decision to go into business he could easily have had a distinguished academic career. Even though he hadn't

24 *Ross McKittrick*

done any math for nearly thirty-five years, he soon found the math in MBH98 wasn't very difficult once you got past the convoluted description. Steve figured it would be an interesting exercise and was willing to put the time into it. What a lucky break for science.

### MANN MINUTIAE AND MORE MANN MINUTIAE

After collecting the source data, Steve began looking at the individual proxies, the way that a geologist would look at drill core, and began making a series of postings on the internet group "climatesceptics" under the heading "Mann minutiae," followed by "More Mann minutiae," telling people about his work. His comments caught the interest of, among others, Dave Douglass of the University of Rochester and Bob Carter of James Cook University in Australia, who encouraged him in the specific analysis of the individual proxies, saying that this work was worthwhile and no one else was doing it.

This he began to do in earnest in the summer of 2003. Watchers of the climate skeptics list would see periodic postings from him, but none of us knew anything about him except that he seemed to be doing a lot of fascinating work on MBH98. In late July, Steve contacted me to say that he lived in Toronto (the University of Guelph is nearby) and suggested we get together to talk about this project. We finally did so in September, meeting at almost the exact moment when Hurricane Isabel hit Toronto.

Sonja Boehmer-Christiansen, the editor of *Energy and Environment*, had been interested in Steve's postings and asked him to consider submitting a paper. He agreed, flattered that his notes had occasioned any interest.

### PRINCIPAL COMPONENTS

With that encouragement, he decided to try to replicate the principal components calculations in MBH98. "Principal components" analysis is a statistical method of reducing the dimension of a large data set. A principal component series (PC) is essentially a weighted average of the original group of series, with the weights chosen in a special way to maximize "explained variance." A data set has an associated sequence of PCs (denoted PC1, PC2, etc.), each one accounting for successively less and less of the variance of the underlying data. The algorithm is included in modern sta-

tistical languages and requires a single line in R (which is the language that Steve was using) to execute.

MBH98 used PC analyses to simplify both temperature and proxy data. For temperatures, they simplify 1082 grid cell series to sixteen principal component series. For proxies, MBH98 describes 112 proxies, of which 71 are individual records and 31 are PCs computed from six "networks" of individual proxies, containing more than three hundred original series. The networks are from geographical regions with labels like "NOAMER" (North America) and "SWM" (Southwest–Mexico). The maximum number of retained PCs for each region varies from three to nine. Details of these calculations were never published by Mann, but the sites for five of six regions were listed on a Supplementary Information (SI) page at *Nature*. Aside from upgrading his computer savvy and refreshing his math skills, Steve had a formidable task in merely collecting, collating, and verifying the original data from the World Data Center for Paleoclimatology for the listed sites, which often had slightly varying formats.

He first tried to replicate Mann's temperature PC calculations using the most recent IPCC data. PC calculations only work when there are no missing data. In this case, he found that there were large amounts of missing data and the PC calculations wouldn't go through. In early September 2003, he e-mailed Mann to ask how he had dealt with missing data. He received no reply.

After laboriously collecting and verifying the tree-ring data, he calculated the PCs for five networks and compared the results to those which Mann had provided. They didn't match at all. Naturally he assumed he had erred in his own calculations and began to cross-check each step of the calculations.

He had used a standard algorithm in a high-level programming language (R), so he couldn't see that an error could have occurred at this stage. He calculated the explained variances from his PCs and those using the PC series from Mann. If a PC has been computed properly, the first PC should explain the most variance in the underlying data matrix. But for some networks, Steve obtained explained variance levels as low as 6 percent from the Mann data—which was impossible. After a while, he wondered whether he had somehow collated the data incorrectly when he read the data into R. This led him to examine the data set visually, whereupon he saw that many of the PC series that he used began in rows where the year ending is \*99 or \*49, while Mann's usual practice was to start series in years ending in \*00 or \*50. On a hunch, he shifted one of the PC networks up a year, and the

explained variance shot upward. Further inspection confirmed that this was a problem in the original file, not in his own collations. He emailed Scott Rutherford about the problem, but Rutherford said that the data were before his time and that Steve would have to contact Mann himself.

Steve also checked what happened at the end of the series. It turned out that the 1980 values for many PC series were identical to seven decimal places. It appeared that there had been gross collation errors in MBH98. Ironically, we would later learn that this problem was in the particular file (pcproxy.txt) Steve had obtained back in April, but likely did not affect MBH98 itself. Still, it was this error that indicated there were significant flaws in the database, leading us to undertake the work that uncovered the errors that did drive MBH98 results.

By this time, Steve was confident in his own PC series for each of the five regions where sites had been identified. In each case, his calculations resulted in very different series from those that Mann had obtained. The explained variance for his own calculations easily surpassed the explained variance from the PC series provided to us. We couldn't yet figure out what was wrong with the MBH98 calculations, but it was evident that something was.

While Steve was looking at the 1980 values, he noticed that the closing values for other PC series were simply "fills": extrapolations from an earlier year. Steve had just noticed these points when we met. Prior to our lunch, Steve had emailed a long paper he had drafted discussing his work up to that point. It was interesting, but the discussion was unfocused and the conclusions unclear. I agreed to work with him to help sharpen the focus and edit the paper down to a publishable form. In our collaboration since then, Steve has been the one doing all the hard stuff with data and analysis, while my role has been to ask questions, read his work, edit/rewrite, and help plan the research strategy. It's been a good collaboration.

At lunch, to avoid any possible misunderstandings, we agreed that it was time to send the entire data set to Mann and ask him for confirmation that this was the data actually used in MBH98. Steve did this. In reply, Mann said that he was too busy to respond to this or any further inquiries and referred us to a publication by Zorita et al., who, he said, had had no trouble in replicating his methods. In the aftermath of our later E&E publication, we took some flak for not giving MB&H more opportunity ahead of time to address our concerns before our paper was published. But the fact is that Mann had categorically terminated our correspondence without answering reasonable questions about his methodology and the provenance

of the data set. With a clear message that Mann did not want Steve to contact him any further, we could hardly keep sending our analysis to him; I doubt he would have responded if we had. Indeed later, Mann continued to refuse to provide information on his data and methods.

Meanwhile, I asked Steve to send me the data. I carried out my own visual inspection of the data, finding new examples of filled-in values including one pair of series where the two different columns contained identical data for nearly thirty years. I suggested that we focus first simply on the implications of the data problems, leaving methodological discussions for another day.

That discovery persuaded Steve to tear apart the whole data set and look at every series from scratch. He began trying to identify original source data for every series in MBH98 and comparing the original data to the version used in MBH98, finding that the versions used in MBH98 were frequently different from publicly archived versions. The differences were generally due to the apparent use of obsolete versions in MBH98.

In other cases, he found that MBH had truncated series without any explanation and that some of the series listed in the Supplementary Information were not actually used. While this was going on, Steve had also been patiently trying to decode the obscure methodological descriptions of MBH98, turning them into linear algebra (which he remembered from his youth) and trying to see if he could replicate MBH98 results. Using the data provided to him, and following the published methods in MBH98 as closely as possible, Steve could get a hockey stick-shaped graph that was close to the original results, but not an exact replication. However, the emulation was clearly good enough to be used to study the effect of the various data problems that we had identified.

The \$64,000 question was whether the data problems affected any results. Steve plugged in the new data set, with freshly calculated PCs and up-to-date data where updated versions had been identified, with no expectation of what would result. The result was the graph that we soon published, showing that early fifteenth-century values exceeded twentieth-century values, contradicting the MBH98 conclusion of twentieth-century uniqueness. Rumors of this finding began to spread quickly.

The conclusion of our paper was that "the extent of errors and defects in the MBH98 data means that the indexes computed from it are unreliable and cannot be used for comparisons between the current climate and that of past centuries." Notice I am not claiming we rediscovered a Medieval Warm Period. We have tried to be careful not to suggest we're proving the fifteenth century was "warm" compared with today. We simply argued that



28 *Ross McKittrick*

this was the result of applying MBH98 procedures to updated and properly collated data.

We wrote up these findings and, knowing they were provocative, we sent the draft to many readers, which resulted in a lot of constructive criticism. We also engaged a professional statistician who consults to paleoclimatology labs to write a report for us evaluating our paper. When we were satisfied that it was our best shot, we sent it to *Environment and Energy*. The referees asked for some changes, which we made, and the paper was accepted quickly. Anticipating the impact it would have, Bill Hughes, the publisher of *Environment and Energy* kindly waived the copyrights and allowed the paper to be posted on the MultiScience website for free distribution at the end of October, in a paper now known as “MM03.”<sup>4</sup>

In our article, in addition to describing our findings, we also tried to document clearly our methodological decisions on matters where MBH98 procedures remained unreported, and provided Supplementary Information on a website ([www.climate2003.com](http://www.climate2003.com)) containing the computer code used for our calculations.

One such decision related to the vexing problem of missing data in PC calculations, which was not described in MBH98 and which Mann had failed to elucidate upon inquiry. In Steve’s program, he calculated tree-ring PC series over the maximum period for which all sites in the network were available. In some cases, that led to longer series than used in MBH98 and in some cases it led to shorter series, which we noted in our article. At this stage, we knew that there was obviously something wrong with the MBH98 PC series on multiple counts: the odd start dates, the identical 1980 values, and the low explained variance. We did not know precisely what was wrong with the MBH98 PC series. The difference in series lengths seemed merely one more strange defect. On the other hand, our own calculations were done with fresh data and could be verified as having a higher explained variance on the underlying data than those used by MBH98, so we were on solid ground with using these PCs, even though we did not yet know why the MBH98 proxy PCs were coming out wrong.

In the aftermath of publication, the resulting publicity forced Mann to disclose much previously unavailable information, including the location of a vast amount of FTP data unavailable to MM03. The new information has allowed us to nail down exactly how Mann’s hockey-stick curve was constructed. The answer does indeed lie in the PC series and our *E&E* conclusions have been vindicated on grounds we had not anticipated at the time.

## MM03 AND RESPONSES

MM03 attracted interest right away. Comments and news articles appeared in United States, Canada, Australia, Holland, Norway, Germany, Argentina, and several other countries. We attracted a great deal of attention in specialized blogs and chat lines. Since the U.S. Senate was debating the McCain–Lieberman bill, there was considerable interest in our work in Washington. Steve and I did a briefing on Capitol Hill on November 17, 2003. A few partisans like Mike McCracken and Stephen Schneider editorialized against us, but by and large the “mainstream” response was to acknowledge that we had made a legitimate critique of the hockey-stick results, that if we were right it would have serious repercussions in the global warming debate, and that there would have to be some detailed discussions in the months ahead to settle the matter fully.

Michael Mann’s first reaction to our paper was in the form of some comments provided to a U.S. website, [www.davidappell.com](http://www.davidappell.com). For one thing, Mann insisted that we had analyzed the wrong data set. The file we looked at (he said) was a special-purpose collation put together in Microsoft Excel in response to Steve’s inquiry back in April, as a “courtesy,” since Steve was supposedly unwilling to get the data from Mann’s FTP site. Supposedly some mistakes were inserted into the Excel file, and the resulting data set bore no resemblance to the one used for MBH98. He opined that we ought to have gone to his FTP site, where the real data are, and that had we done so we would have discovered the data we were using were flawed and we would then not have produced the results we did. Mann also gave out a new and different URL for the location of the MBH98 data, a URL that was not referred to at Mann’s own website or in any other public source to that date.<sup>5</sup>

Mann’s statements were untrue. Steve had not originally asked for a data file but for an FTP site. Mann had told him he had forgotten where it was and referred the request to Rutherford. Rutherford said the data were scattered over several sites, not just one. Eventually Rutherford gave us a URL at Mann’s FTP site, which pointed to a plain text file (`pcproxy.txt`), not an Excel file. After the controversy broke, we checked the date of creation of this file and found that it was the summer before, so it obviously wasn’t a special collation generated for Steve. A few days after the publication of MM03, Mann apparently erased the file, removing the evidence of its age, but we had verified the date just before it was deleted. Now Mann was saying that the “right” data were at a different URL, which, despite several previous inquiries, he had never disclosed and we could not have accessed given the information we had to that date.

30 *Ross McKittrick*

As for his suggestion that we had not noticed the errors, that was simply bizarre. Obviously we knew there were flaws in the data file we received—we spent twenty pages listing them! After finding the errors, Steve didn't keep using the flawed data; he rebuilt the whole data set from the original sources. Our reanalysis was based on the *corrected* data set, not the flawed data at the URL we had been given.

After we learned about the new URL, we compared it with the data from the old URL. Although the data at the old URL were collated, the data at the new URL were uncollated. We were quickly able to confirm that eighty-one proxy series, where there were no PC calculations, were identical in the two versions, and any comments in MM03 about these series were unaffected by the version dispute. We were also able to see what had happened in the collation at the old URL (pcproxy.txt). In the uncollated data, there were different versions of the PC series for each region, stored in subdirectories with names like BACKTO\_1750, BACKTO\_1600 and so on. Steve went through these folders series by series and was able to see how the collated data provided to us had been generated by having one column for each PC, with the values in the later subdirectory overwriting earlier values (subject to the collation error which we had noticed). On November 11 we posted our comments on the provenance of the data on our website.<sup>6</sup> Of course, that did not provide evidence one way or the other on what version was actually used in MBH98.

### THE THREE KEY INDICATORS

A more substantive response by Mann and coauthors came a few weeks later in the form of a short paper by MB&H vigorously rejecting our conclusions.<sup>7</sup> They argued that our fifteenth-century results arose because we “selectively censored” early (pre-1500) segments of three “key indicators”: a tree-ring width proxy from the site at Twisted Tree Heartrot Hill (TTHH) in northern Canada; the first principal component (PC1) of earlywood and latewood ring widths from a roster of ten sites in southwestern United States and Mexico (SW-M) studied by Stahle et al. and the PC1 of ring widths and some densities from seventy-plus North American sites (NOAMER) partly overlapping the SWM network.<sup>8</sup> They presented a simulation showing that the early-fifteenth-century portion of their NH temperature index would, like ours, exceed the late twentieth century without these series in the fifteenth century. They argued that we had improperly deleted a substantial amount of early data, that our tem-

perature index had a weaker statistical fit than theirs and accordingly, our results were worthless.

It was immediately obvious to us that the TTHH series was a red herring. The obsolete version used in MBH98 ended in 1976 and included a sixteenth-century portion based on a single tree; however, this version was not archived at WDCP. The archived site chronology goes up to 1992, but only commences in 1529, when three trees become available, and this is the version we used. However, even the TTHH series as used by MBH98 did not begin until 1459, so it is irrelevant to the pre-1450 interval in any case.

The issues with the two other indicators had arisen because of a problem in MBH98 disclosure. MBH98 said that they used 112 proxy series and "conventional" PC algorithms. Recall the question about how to handle missing data: Steve had decided to deal with it by calculating PCs over the maximum possible period. This led to different lengths for PC series than in `pcproxy.txt`, which we annotated in MM03 as one of the puzzles over the PC series in MBH98. Now Mann et al. said that they had used not conventional PC analysis, but a "stepwise" analysis in which they changed the rosters in each region in older periods, requiring 159 series instead of 112. The figure of 159 series had never been mentioned anywhere. It meant that there were PC series for the Stahle/SWM region and the NOAMER region in the AD1400 step of MBH98 calculations, which we did not have available in our calculations. This unavailability did not arise from "selective censoring," but from calculating PC series over the maximum period in which there were no missing data—which was our best interpretation of the obscure methodological descriptions in MBH98.

Replicating the stepwisemethod required a lot of information about which there was not a whisper in MBH98. First of all, what exactly were the 159 series? Mann et al. said that they had made fresh PC calculations for each region for each step, but that was clearly untrue, as this would lead to many more than 159 series. Also, at Mann's FTP site, there were directories for some periods, but not for others. Steve tried lots of different combinations, but it was impossible to come up with an exact match, balancing the right number of indicators in each period and still totaling 159 series. We asked Mann for details, but he refused to enlighten us, saying that we should be able to figure it out from his FTP site. Yet there were hundreds of PC series listed there, many of which were obviously not used. Eventually, Steve decided that the figure of 159 series was probably not correct (that still appears to be the case). But for the purpose of redoing our calculations back to 1400, Mann had stated that he used one PC from the Stahle/SWM network and two PCs from the NOAMER network, so we

started with that. Steve made his best estimate of the numbers for the other regions by step. There turned out to be other errors in the MBH98 information that made these estimates a little off, but these discrepancies didn't matter for the main point in controversy, which now centered on the early fifteenth century.

In the end, none of this mattered for the SWM PC1. It turned out that its role could, as the lawyers say, be decided on alternate grounds. The SWM network is based on Stahle et al. (1998; see footnote viii). For each site in the SWM network, MBH98 used two data series: earlywood and latewood widths, although Stahle et al. did not use latewood widths. Of the sites listed in the MBH98 Supplementary Information (SI) on the *Nature* website, only two of them (four series) are available before 1450. But Steve found that Michael Mann's FTP site listed three sites (six series) extending back prior to 1450. This seemed to suggest a mysterious third site was used: but then two of the sites (four series) turned out to have identical values for the first 120 years for earlywood widths and the first 125 years for latewood widths, each differing thereafter. So they might have been spliced versions of different sites or different editions of the same site. Either way at least two series were clearly ineligible pre-1450, leaving only two potentially eligible sites. In other regions, MBH98 did not extend a PC1 back through an interval if only two sites were still available, and consistent application of this criterion would exclude the availability of the SWM PC1 in the pre-1450 period. Moreover, one of the two remaining sites is Spruce Canyon Colo., which is also in the NOAMER roster and should therefore have been dropped from the SWM group. The data for the remaining SWM site, Cerro Barajas, as used in MBH98, includes physically impossible negative values in the early portion of the series, which are not present in the version archived at WDCP. And Stahle et al. themselves did not apply their network prior to 1706. So on many grounds, we had reason to exclude the pre-1450 portion of this network. But there was an even more important reason. Even though Mann et al. had cited it as a "key indicator," our calculations showed that its presence or absence in the fifteenth century didn't matter. In our subsequent *Nature* correspondence, when we pointed out all these quality problems with this data, Mann et al. were only too happy to agree that it didn't matter for their early-fifteenth-century results.

The whole story thus seemed to turn on the North American PC1. We recalculated all our PC series following our estimates of stepwise inclusions, including the North American PC1 and PC2 (and the SWM PC1, which didn't matter) back to 1400. Using our recollated source data, we then recalculated the Northern Hemisphere temperature index. Our results

still looked the same as MM03. So the difference was more than the recently revealed "stepwise" procedure for principal component calculations, as Mann was claiming, and it was linked somehow to the NOAMER PC1. Since the climate history of the world seemed to turn on this one indicator, we figured we'd find out everything we could about it. I'll return to this story in a moment.

In the meantime, Mann had also criticized other aspects of our emulation of his team's method. We had no interest in irrelevant disputes and figured we could settle the major methodological differences if we could just inspect the source code used in MBH98. Since Mann had redone his calculations in writing his response, we knew he had the source code at hand. We asked for it, but Mann refused to send it, or any more information for that matter. So Steve contacted the U.S. National Science Foundation (NSF), which had paid for the MBH98 research. Knowing there are rules requiring disclosure of NSF-funded data, we figured they might intervene. A program officer looked at our complaint and contacted Mann. Mann wrote back explaining what he had divulged to date and insisted he had fulfilled his NSF obligations. The file was eventually sent to the NSF general counsel, who ruled that there was no legal obligation for further disclosure: they regarded the code as Mann's personal property.

Unfortunately, we have since found this poor disclosure of data and methods is not an isolated situation in paleoclimatology. Other studies have an even worse record. Steve has contacted numerous paleoclimatologists in search of their data and has a thick file of excuses, dismissals, and brush-offs, along with a few honorable exceptions. Nor is the situation unique to paleoclimatology. Two economists recently took a 1999 edition of the *American Economic Review* and tried to replicate the empirical papers, only to find most authors unwilling or unable to share their data and command files in a usable format. So this year the AER adopted a strict policy that empirical papers will no longer be published unless the authors supply their data and computational files to the journal's online archive.<sup>9</sup> More journals, even or especially paleoclimatological journals, should adopt a strict disclosure rule like *AER*'s. Advances in software and internet communication make this feasible and inexpensive.

We got caught up in another journal's debate on the subject in December, when we were alerted that a variation of the Internet response was listed on the *Climatic Change* (CC) website as "forthcoming." Steve wrote to the editor, Stephen Schneider, to protest about some of the language in it. Schneider wrote to assure us it was not, in fact, in press, but only under review, and indeed the version we saw was not the version actually being

34 *Ross McKittrick*

reviewed (there had been a mix-up at the *Climatic Change* office). He also said that he had mailed Steve an invitation to act as a referee. The new article referred to similar calculations as the ones that we had been trying to obtain information on. So in his new capacity as referee, Steve asked Schneider to obtain the supporting calculations and source code, so that he could carry out the requested peer review. That created consternation at *Climatic Change*. Schneider said that no one had ever made such a request in his twenty-eight years of editing of *Climatic Change* and refused to ask for that information without consulting his editorial board. The request prompted a long debate within the board (we were later told) about whether *Climatic Change* should ask its authors to release their computational files. The consensus was that it would be an excessive burden on reviewers if they were expected to review source code and supporting calculations. Steve replied that he was not suggesting that *Climatic Change* change its policies; he merely said that he was willing to examine the code in this particular case, and he did not see how that would set a precedent. Further, he could see no reason why *Climatic Change* should not request the information; after all, Mann might simply agree to provide it. Schneider still refused to ask for the source code, but agreed in February to ask Mann for the supporting calculations, that is, the results for the separate steps. In the end, Mann refused to provide the supporting calculations and Schneider asked that Steve complete his referee report anyway. Needless to say, the first comment in the referee report was that Mann et al. had failed to comply with the supposedly mandatory requirement to supply their data and supporting calculations. We haven't heard any more about this file recently.

#### THE FTP SITE

We knew that the whole MBH98 edifice now hung by the single thread of the NOAMER PC1 and that we had been unable to replicate the “key” PC1 using freshly collated data and the standard algorithm (princomp) in R. Steve now turned to Mann's recently unveiled FTP directory for MBH98, a strange, sprawling collection that had suddenly materialized in the wake of MM03. Prior to MM03, no reference to this URL had ever appeared, even in Mann's own web citations for MBH98 data. Mann's FTP directory contained data for no other paper.

The FTP site contained information on the series actually used in MBH98. Steve compared these listings with those in the original supplementary information (which he had used in his PC calculations) to see if

this explained the difference in the NOAMER PC1. Discrepancies appeared immediately. Exactly 232 North American series were listed in the original SI, while only 212 series were actually used. He then checked the other networks and found similar problems: in the South American network, 18 series were listed in the original SI and only 11 were used in the actual calculations. There was no explanation why the series had been set aside, though an e-mail between the coauthors inadvertently left on the FTP site mentioned that deletion of the series arge030, one of the discrepant series, would be "better for our purposes."<sup>10</sup>

For the critical AD1400 step of the NOAMER network, the difference was only six series. We calculated the critical NOAMER PC1, using both the stated roster and the roster actually used. This did not make a material difference to the critical PC1, so the mystery of how to reconcile the NOAMER PC1 calculations remained.

While Mann continued to refuse to provide source code in response to our requests, in the folders at Professor Mann's FTP site we found remnant Fortran programs, perhaps unintentionally left on the site that had been used to calculate the PCs. This was the only step in the entire calculation where we had the opportunity to actually inspect working source code. But we struck gold here.

Steve went through the Fortran programs line by line to see what they had done. Before doing a singular value decomposition (an algebraic factorization that yields PCs), they had "standardized" the series by subtracting the mean over the 1902–1980 subsegment, then dividing by the standard deviation over the same subsegment, then dividing again by the "detrended" standard deviation. In calculations using standard software, any standardization is done using a mean and standard deviation computed over the full length of the series, (say) 1400–1980 for the period in controversy, but in the MBH98 program they used the post-1901 portion of the data to compute the mean and standard deviation. This was not how the procedure had been described in MBH98. There they described subtracting the 1902–1980 mean and dividing by the 1902–1980 standard deviation prior to the regression module (i.e., later in the computation sequence). But they did not describe this procedure as having also taken place prior to the PC calculation. Since PCs are sensitive to changes in the way the data are standardized this would undoubtedly have raised the eyebrows of reviewers and readers, had they been told. It appeared that this had been done inadvertently, since an experienced user of PC methods such as Mann would have known that this would have an impact on PC calculations and that he would therefore have an obligation to disclose it. This apparently inadvertent error



36 *Ross McKittrick*

was at the heart of MBH98 and this, together with a little “trick” explained below, yielded Mann’s famous hockey stick.

Steve also looked through the FTP site to see if there was any information on the results of the individual calculations for each step (especially the AD1400 step) which were used to support MBH’s claims of high statistical accuracy, but he could not find any.

We remained frustrated with the Mann’s obduracy in refusing to identify the 159 series or to disclose the results of his steps. Accordingly, in November 2003, we sent a “materials complaint” to *Nature*, under their policy requiring authors to disclose their data and methods.

The complaint included an item expressly noting the inaccuracy in MBH98 regarding the data transformation prior to the PC calculation, which we regarded as the issue most affecting the early fifteenth-century results. But we also listed other items that affected the integrity of the publication record, including the discrepancies between the series listed as being used and the series actually used. Sometimes, series were used twice in MBH98—with no notice to the reader or clear explanation. In the email mentioned above, one of the authors listed series that at a “wild guess” seemed to duplicate one another and recommended their removal, but, when we checked, his recommendations had not been carried out. While the NOAMER PC1 calculation accounted for most of the difference between MM03 (and now MM04) and MBH98, one of these duplications proved to account for the rest.

Another complaint was that, in a directory containing long temperature records, two series (nos. 6 and 8) were excluded without explanation.<sup>11</sup> Series 6 shows a conspicuously declining twentieth-century trend (figure 2.2). We hated to be suspicious, but the discrepancies were piling up. In total, we listed ten items where the disclosure in MBH98 appeared to be inaccurate; *Nature* forwarded this listing of ten items to MB&H for a response.

#### HOW TO MAKE A HOCKEY STICK

In January 2004, in addition to our materials complaint, we submitted an article to *Nature* that showed how the hockey stick was manufactured. We showed how the undisclosed programming error that Steve had discovered on the FTP site—subtracting the 1902–1980 mean (instead of the mean of the period of the principal component calculation (e.g., 1400–1980)—worked to pick out hockey stick-shaped series (if they were available in the network) and load them into the PC1.

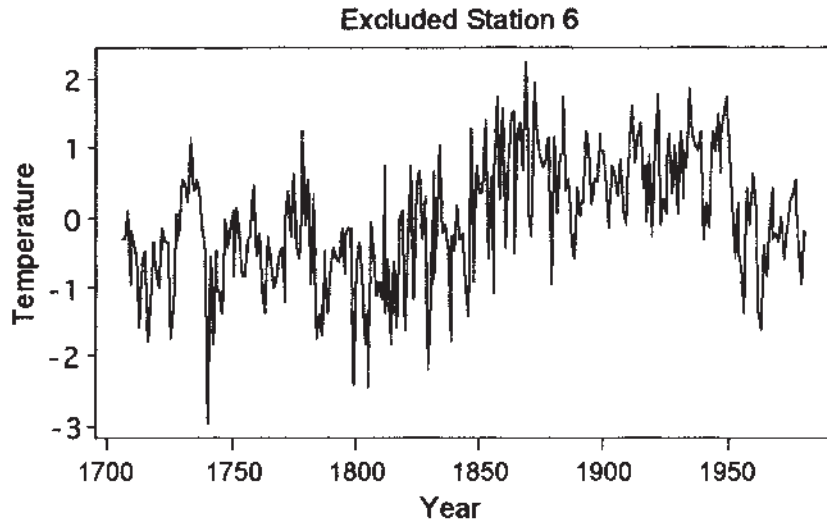


Figure 2.2.

The error would not make much difference to networks that did not contain series with hockey-stick shapes (e.g., the Stahle SWM group). But the North American network did contain hockey-stick shaped series and the error had a big effect on the NOAMER PC1. For hockey-stick shaped series, the 1902–1980 mean is higher than the (say) 1400–1980 mean, sometimes significantly so. Subtraction of the 1902–1980 mean therefore inflated the variance in hockey stick-shaped series relative to what would happen from subtracting the 1400–1980 mean. Since PC algorithms load extra weight in the PC1 on series that have higher variance, the error resulted in picking out hockey stick-shaped series and overweighting them in the PC1. This was, presumably, a simple programming error, but it had a huge effect.

We showed this in a couple of different ways. First we showed the extreme differences in series weightings in MBH98 calculations. In the NOAMER roster for the AD1400 step, the most heavily weighted site is Sheep Mountain, Calif. (ca534). Sheep Mountain has a hockey-stick shape and Mann's algorithm gives it a whopping 390 times the weight in the PC1 of the least weighted series, Mayberry Slough, Ark. (ar052). The different shapes are shown in figure 2.3.

We'll call the Fortran program Mann1. Once we saw how the program worked, we tried an experiment to see if it could generate a hockey stick from random numbers. To generate the data we took the seventy

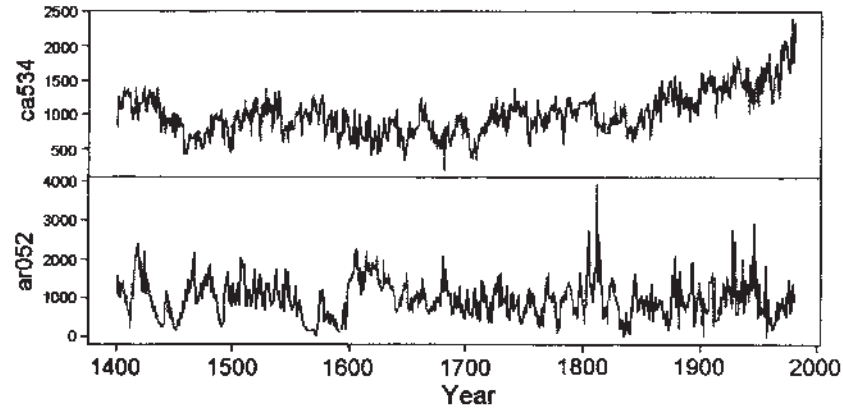


Figure 2.3.

NOAMER sites available back to 1400 and fitted a lag-1 autoregression model to each. The coefficients ( $\beta^1, \dots, \beta^{70}$ ) were all of magnitude less than one. Then we generated seventy random vectors  $a^1, \dots, a^{70}$  of length 1081, using the AR1 formula  $a_t^i = \beta^i a_{t-1}^i + e_t^i$ ,  $i = 1, \dots, 70$ . Each series was initialized at zero and run using standard normal ( $N(0,1)$ ) errors  $e_t^i$ . The first five hundred values were then dropped from each series, yielding seventy vectors of stationary red noise, each of length 581.

The (conventional) first principal component from these seventy series, after smoothing, showed the expected stationary sawtooth pattern (figure 2.4, top). Mann1 yielded the hockey stick-shaped PC1 shown in figure 2.4, bottom panel. The reason for the hockey-stick shape is that some of the underlying series randomly trail up or down at the end of their length,

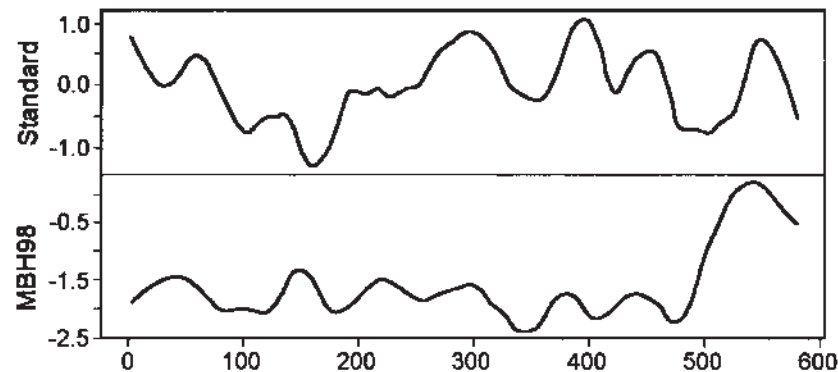


Figure 2.4.

and these are selected for high weighting by the MBH98 method. Repeated experiments consistently returned hockey-stick shapes in the PC1.

Mann1 had a big effect on the AD1400 step of the North American network and we were able to show this led to the MBH98 version of the NOAMER network PC1, whereas a correct calculation led to our results.

One other error enabled a complete reconciliation of the differences between MM03-MM04 results and MBH98. Many series in MBH98 are duplicated within the data base. One of these, the Gaspé "northern tree-line" series is used as a separate proxy (treeline 11) and in the NOAMER PC collation as cana036.<sup>12</sup> The data begin in 1404. When used as treeline 11, MBH98 gave the start date as 1400 and filled the empty first four cells by extrapolation (see MM03 table 5). This was the only extrapolation of a start date in the entire MBH98 corpus and we were consequently curious about it, especially now that the early-fifteenth-century results were at the heart of the controversy. The misrepresented start date enabled them to avoid disclosure of the unique extrapolation; the extrapolation enabled them to include this series in the AD1400 step, rather than withholding it until the AD1450 step. We calculated the NH temperature index without the extrapolation in the duplicate treeline 11 version and found that this seemingly innocuous four-year fill had a major effect on early fifteenth-century results (up to 0.4 degrees C in some years) and fully reconciled those differences between MBH98 and MM03-MM04, which were left after correcting the PC calculations.

We found other problems with the Gaspé series. Steve found that the first fifty years of the chronology fail standard minimum signal criteria.<sup>13</sup> The underlying dataset commences in 1404, but is based on only one tree up to 1421 and only two trees up to 1447. Dendrochronologists do not use site data where only one or two (or zero!) trees are available for generating a chronology. In fact the source authors don't use the series before AD1600 (see Jacoby and d'Arrigo 1989; D'Arrigo and Jacoby 1992).<sup>14</sup>

We now could carry out a comparison similar to the one in MM03, but based on a much closer replication of the MBH98 data and methodology, this time using the NOAMER PC1 and PC2 back to AD1400 (thereby answering the previous criticism from Mann et al.), but using correctly calculated PCs. The results are in figure 2.5, which was the conclusion to our submission to *Nature*. We submitted the article in January 2004, along with a cover letter explaining that it dealt with issues separate from our Materials Complaint, where we dealt with defective disclosure in MBH98, while the article dealt with impact of methodological errors (the methodologies also happening to be also undisclosed).

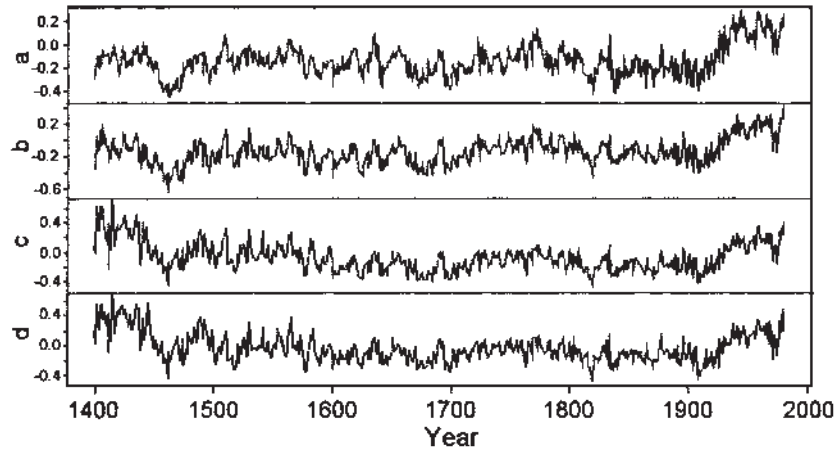


Figure 2.5.

Figure 2.5's top panel shows the MBH98 graph, and the second panel shows Steve's best emulation, which has a correlation of 0.89. Since the replication also shows a unique late twentieth century and the basic hockey-stick form we were not concerned about the small remaining differences in result, and without full methodological disclosure by Mann there was nothing we could do about it anyway.

The third panel (c) shows the result using the same method as for (b), but applying correct PCs and removing the extrapolation of the duplicate Gaspé version. Those changes alone suffice to refute the conclusions in MBH98. The fourth panel (d) adds in all the data corrections as outlined in MM03 and the Corrigendum of Mann et al.<sup>15</sup> Obviously the climate of the later twentieth century is unexceptional compared to the fifteenth century; to the extent this index summarizes the state of the climate.

At the end of February 2004 we received notice from *Nature* that, after consideration of the response by Mann et al. to our complaint, MB&H would be instructed to publish a Corrigendum to MBH98 and be required to provide a new website with a listing of data and methodology sufficiently clear to permit replication of their results. We were specifically assured that the corrigendum would not engage in the controversy over the materiality of the errors but would simply list and correct them.

MB&H had also responded to our submitted article, and the papers were sent to two referees. Neither was convinced by the MB&H response and both supported publication. In March, we were asked to submit a revised version, responding to comments by referees and Mann et al. We were

especially intrigued by an MB&H response point that our presentation exaggerated the effect of Sheep Mountain, since (they said) fourteen other series contributed heavily to the PC1. We wanted to know what the other fourteen series were, and this comment inadvertently proved to be a Rosetta Stone for the final decoding of MBH98.

Almost all the NOAMER series selected for overweighting were of a single type and from a single researcher, Donald Graybill. The series were high-altitude bristlecone pine tree-ring chronologies, many of which had been studied by Graybill and Sherwood Idso as possible examples of CO<sub>2</sub> fertilization of tree growth, following a similar study by Lamarche et al. on Sheep Mountain.<sup>16</sup> The sites were selected for "cambial dieback," that is, the bark had died around most of the circumference of the tree. Graybill and Idso reported anomalously high twentieth-century growth for trees with cambial dieback, as compared with "full bark" trees at the same site. They reported that the anomalous twentieth-century growth was unrelated to the temperature data from nearby weather stations. In the case of Sheep Mountain, a weather station had operated within 10 km for nearly thirty years. Mann, Bradley and Hughes themselves wrote in MBH99: "A number of the highest elevation chronologies in the western United States do appear, however, to have exhibited long-term growth increases that are more dramatic than can be explained by instrumental temperature trends in these regions." Later, coauthor Hughes in Hughes and Funkhouser (2003) would state that these elevated growth rates were a "mystery."

Yet these sites were selected by Mann1 for such high weighting as to nullify the contributions of all other series in the NOAMER collection put together. None of the other proxies in MBH98 look like hockey sticks, but their influence was wiped out. The bristlecone and related series accounted for more than 99 percent of the weighting in PC1, which in turn was said to account for 37 percent of the variance in the North America network. In the subsequent regression calculation, the North America PC1 imparted its shape to the whole Northern Hemisphere temperature index, giving it the distinctive hockey-stick shape. This was startling enough, but there was an even bigger surprise to come.

At Mann's FTP site, there is a folder in the NOAMER directory called BACKTO\_1400-CENSORED. You can imagine how intrigued we had been by this folder. It contained PCs, but it did not say what series were in it or what its purpose was. In another directory, Mann had a listing of 212 uncensored series and 192 "censored" series. Now that our attention had been drawn to the Graybill series, we checked to see if there was any connection to the twenty excluded sites. Again Steve struck gold. All fourteen

of the Graybill sites were among the excluded series. The other six sites were also Graybill sites. We found that all twenty series were included among the seventy series in the uncensored BACKTO\_1400 file, while there were only fifty series in the CENSORED file.

So now we now knew what the calculation in the CENSORED file was—it redid the North American PC calculations in the critical AD1400 step after excluding the controversial Graybill sites, about which even Mann et al. were evidently worried.

This provided a test of our argument that the faulty PC algorithm was driving the MBH98 results. By removing these outlier series, the PC algorithm no longer had hockey sticks to overload on, and therefore should revert to a conventional shape. The comparisons appear in figure 2.6. The top panel shows the NOAMER PC1 as computed by MBH98 using their erroneous standardization. The resemblance to the Sheep Mountain site (figure 2.3a) is evident. The second panel shows the simple mean of the seventy proxies in question. This is quite similar to the third panel, which shows the first PC of the same seventy series correctly computed, that is, after standardizing over the full period. The bottom panel in figure 2.6 shows the PC1 from the CENSORED file, i.e. as computed by MB&H themselves after dropping the twenty Graybill sites. It matches the correctly computed PC1 (MM04) almost perfectly. Evidently we weren't the first ones to discover the role these bristlecone series were playing! Mann had also wondered about them, and redid his calculations without them. Given

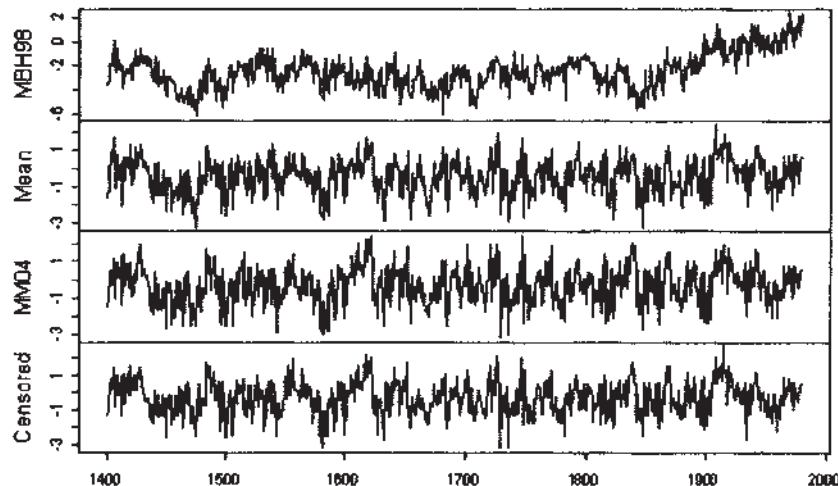


Figure 2.6.

his subsequent furious excoriation and accusations of "selective censoring," we were amused to find results similar to ours lurking on his FTP site in a folder called "CENSORED."

I must note the hilarious irony in the fact that data published (in part) by Sherwood Idso, the famous global warming skeptic, would, as the result of someone's programming error, reappear many years later as the NOAMER PC1, now billed as the "key indicator" and principal evidence for the IPCC's position on global warming! We submitted our response incorporating these comments on March 18.

Meanwhile, on March 16, we received a page proof of the corrigendum. It left out the most important inaccuracy of MBH98—the incorrect disclosure of the (incorrect) PC methodology, as well as many other items. It even contained a few new errors of its own.

The corrigendum attributed the discrepancy in the series listings to the application of quality control rules additional to those already mentioned in Mann et al. (2000), which included the following criteria: a tree-ring chronology would not be used unless the mean correlation of individual tree-ring records with the site chronology was at least 0.5; the chronology commenced by 1626; and it was composed of at least eight tree-ring segments by 1680.<sup>17</sup> Steve had previously gone through the MBH98 data and found thirty-nine series that did not commence by 1626, twenty-two sites that did not have eight trees by 1680, and 171 sites that had less than 0.5 mean correlation of the individual trees with the site chronology. In one of the cases (cana153) the tree-ring segments had so little correlation with the site chronology that Steve emailed the originating author (Roseanne D'Arrigo) to ask why. She discovered that the wrong data had been posted at WDCP and immediately asked them to remove the listing. That surely would have been noticed earlier if Mann had indeed applied his quality control rules to all the data.

We wrote to *Nature* to object all of these matters and they temporarily pulled the Corrigendum out of production. About ten days later, however, *Nature* dismissed our concerns about the inaccuracies in the purported explanation of the discrepancies, on the grounds that they were "irrelevant" to our original materials complaint. *Nature* also said that "space limitations" precluded a listing of all the errors, but they felt confident that the new SI would provide a complete record of the data actually used. On the same day that our objections to the draft corrigendum were dismissed because of "space limitations," we were asked to cut our article down to eight hundred words to fit the format of a Communication Arising, which we did on April 9, 2004.



The corrigendum was eventually published on July 1, 2004, as was a new online data archive for MBH98. The printed version contained the preposterous claim that none of the errors had any effect on the results, a claim that had been inserted after production of the page proofs we were shown, and which was obviously inconsistent with material then under review at *Nature*.

Almost four months later (August 4) we were told our article would not be published. *Nature* said the principal reason was that the material could not be adequately explained in the (now) five hundred words they said would be available for our paper.

We did not have an opportunity to respond to the second round of comments from Mann and his coauthors. One counterargument seemed to impress one of the referees. Mann et al. acknowledged the data transformation prior to computing the PCs, but claimed that, even using corrected PC calculations, they could get MBH98-like results if they expanded the list of retained PCs in the AD1400 step for the NOAMER roster to five, instead of only two as in MBH98. They pointed out that the bristlecone pines get heavily weighted in the PC4. For their regression module, it doesn't matter whether the series appear as a PC1 or PC4; either way the bristlecone pines still lever the final NH temperature index into a hockey stick. Without the NOAMER PC4 (which explains less than 8 percent of the variance of the NOAMER network), they would end up with the same results as we got; by including it, they get results that resemble MBH98. So their key conclusion hinges on ginning up a reason to include a minor PC representing the growth of trees that numerous experts have said are not good proxies for temperature.

Ultimately the issue is robustness. Mann et al. made grandiose claims about the "robustness" of their methods—even claiming in Mann et al. (2000) that their method was robust to the exclusion of all tree-ring information. Yet it is not even robust to the presence or absence of the bristlecone pines or to the NOAMER PC4.

In addition to loading hockey-stick shapes into the PC1, the programming error also made the bristlecone "signal" in the PC1 appear to be far more dominant than it really was. It attributed 37 percent of the explained variance in the entire network to the (incorrect) PC1, thus suggesting the Graybill-Idso sites were the dominant pattern for the whole continent, while the correctly calculated PC4, to which these sites actually get weighted, accounts for less than 8 percent of the variance in the NOAMER network. Thus, what was erroneously argued to be the "dominant" signal in the entire Northern Hemisphere climate turns out to have

been a local phenomenon specific to a group of high-altitude bristlecone pines, whose influence was inflated due to a programming error. Mann's hockey stick hinges (literally) on this. And on that flimsy foundation the Intergovernmental Panel on Climate Change based the conclusions of its third assessment report.

The MBH98 data, if analyzed using MBH98 methods, but without erroneous principal component calculations and without the duplicate Gaspé data versions (or even without the undisclosed extrapolation in the duplicate Gaspé series), does not support the conclusion that the twentieth-century climate is unusually warm and does not enable any far-reaching conclusions about where the 1990s rank in millennial temperature.

#### CONCLUSION: PEER REVIEW AND THE POLICYMAKING PROCESS

Canada's Chief Climate Science advisor, Henry Hengeveld, recently dismissed our work in the *Canadian Meteorological and Oceanographic Society Bulletin* by arguing that we are not experts and therefore not competent to identify errors in MBH98. If so, just think what some real "experts" could have discovered had they seriously scrutinized MBH98. But of course the real "experts" did not bother. Despite its spellbinding role in international science and public policy, other than our efforts, there has been no due diligence on Mann's hockey stick at all.

Steve McIntyre likes to contrast what we've witnessed with the layers of due diligence involved in even small offerings of securities to the public. A prospectus must contain audited financial statements. Auditing is carried out by specialized and highly paid professionals and, for large corporations, the audit is virtually a full time occupation. A company issuing an exploration prospectus must provide a qualifying report on its geological properties by an independent geological professional. The geologist must be truly independent. A person signing a prospectus could not use his own reports as "independent" reports. Both the auditor and the independent geologist must approve the relevant language in the prospectus and provide signed consent letters to the securities commission. The prospectus itself is reviewed by two sets of securities lawyers—one for the issuing corporation and one for the underwriter or broker acting as agent. Then the prospectus is reviewed by the securities commission, a nit-picking process. Any errors identified by or concerns of the securities commission must be dealt with, regardless of whether it "affects the results." The process is expensive and

painstaking. After all this, the officers and directors have to sign a form certifying that they have made “full, true, and plain” disclosure, which means not only certifying that everything in the prospectus is true to the best of their knowledge, but also that they have not omitted anything from the prospectus that is material. Even small public offerings have multiple layers of due diligence.

Despite the multiple layers of due diligence for prospectuses, frauds still occur. Lots of people believed in Bre-X and Enron, including people as eminent in their fields as those currently supporting IPCC, and they turned out to be wrong. In both of those cases, there were lapses in due diligence. In the case of Bre-X, the drill core was famously never available for inspection. During its main boom, Bre-X never issued a prospectus. When it listed on the Toronto Stock Exchange, it filed an ore reserves study by a well-respected and eminent engineering firm, which contained the caveat that the ore reserve calculation relied on company information and that no examination of drill core or verification were carried out. The fraud was immediately exposed when the first third-party drill core was done.

In the case of Enron, in retrospect, it seems that analysts never really knew what Enron did or how it made its supposed profits. In both cases, and in common with the dot.com boom, there was a “madness of crowds.” Those happen from time to time in public markets despite the best efforts of analysts and regulators. While these examples may seem very foreign to the academic world, how does due diligence for MBH98 bear up in comparison?

First, “peer review” for an academic journal is a much lesser form of due diligence than an audit of financial statements. The referees of our submission expressly stated that attempting to determine who was right or wrong as between MBH98 and ourselves was far beyond the scope of review that they could provide. Yet the differences are important to public policy and the resolution of these issues should be quite routine, if the original authors were required to cooperate. Auditors would definitely resolve this matter in a business situation. In our dealings with *Nature*, even where issues were explicitly identified, peer reviewers were unable to provide sufficient due diligence to resolve the matter. We are quite confident that *Nature*’s peer reviewers for the original publication did not examine the data or the programs used to produce Mann’s hockey stick or carry out any audit level due diligence.

At the IPCC level, the IPCC itself made no attempt to verify any MBH98 findings, relying only on the prior peer review by *Nature*. There is a common misunderstanding by the general public and the numerous Nobel laureates who endorsed the IPCC report that the IPCC carried out sub-

stantial due diligence of its own. That is not the case. Obviously, problems can result if people think that due diligence has taken place when it hasn't.

The failure of the IPCC to carry out such independent verification or to audit studies may be partly explained by the lack of independence between the chapter authors and the original authors. Michael Mann was lead author of the chapter relying on his own findings, a lack of independence that would never be tolerated in ordinary public offerings of securities.

Subsequent to MBH98, there has been no effort by any paleoclimatologist to specifically replicate MBH98. One paper copied the method to explore how simulated proxy data might compare to simulated temperature data in a climate model simulation, but did not attempt to reproduce the MBH98 results.<sup>18</sup> There are other multiproxy studies arriving at somewhat similar conclusions as MBH98, but most of these studies are not truly "independent." The most often cited multiproxy studies are nearly all by a small group of coauthors: Mann, Bradley, and Hughes (1998, 1999), Mann and Jones (2003), Jones and Mann (2004), Bradley and Jones (1993), Jones, Briffa et al. (1998), Briffa, Jones et al (2001) and so on, and many reuse the same basic underlying data.

In most branches of science, specific replication is required before results are accepted. Yet in paleoclimate, the idea of our merely trying to replicate MBH98's findings has been derided by many climate scientists. Apart from expressing scandal that we didn't just take the findings on authority, they argued that we should have developed our own proxies and produced our own index. But whenever someone proposes a new method, involving advanced methods and a great deal of data handling, if the results are deemed canonically significant it seems self-evident that the programs should be checked to see that there were no errors or unstated methodological variations. No one bothered.

One simple suggestion to minimize similar problems and facilitate due diligence is this: to make independent replication possible, the data and the methods must be published in unambiguous form. Rules such as those now in force at the *American Economic Review* should be universally applied in paleoclimatological publications in the future. In dealing with the backlog of poorly documented papers, institutions seeking to apply older papers should determine whether they have met full disclosure standards prior to citing the papers, and any shortcomings in the availability of complete details on data and methodology should be prima facie grounds to forbid a paper's use in public sector decision making.

Beyond that, there is an obvious need for additional due diligence prior to use of academic articles in public policy. In the private sector, no

one would build an oil refinery based an academic article. There is a process of engineering due diligence. Some of the most highly paid professionals are principally involved in verification. Yet governments will make far larger, costlier decisions based on the chimerical standard of academic peer review. Merely stating the contrast points to the need to ramp up standards in the public sector, and quickly.

## NOTES

- [QU1:** <QU1>  
**Pls. sup-** 1. M. E. Mann, R. S. Bradley, and M. K. Hughes, *Nature* 392 (1998): 779–87.  
**ply title** 2. M. E. Mann, R. S. Bradley, and M. K. Hughes, *Geophys Res Lett* 26 (1999):  
**for each** 759–62.  
**periodi-** 3. M. E. Mann and P. D. Jones, *Geophys Res Lett* 30 (2003) doi: 10.1029/  
**cal cite.]** 2003GL017814; P. D. Jones and M. E. Mann, *Rev Geophys* 42 (2004): RG2002,  
 doi:10.1029/2003RG000143. <QU2>
- [QU2:** 4. S. McIntyre and R. McKittrick, *Environment and Energy* 14 (2003): 751–71.  
**These** 5. ftp://holocene.evsc.virginia.edu/pub/MBH98/ (hereafter *FTP*).  
**notes** 6. http://www.uoguelph.ca/~rmckitri/research/trc.html  
**are un-** 7. ftp://holocene.evsc.virginia.edu/pub/mann/EandEPaperProblem.pdf  
**usual.** 8. D. W. Stahle et al., *Bull Amer Meteorol Soc* 79 (1998) :2137–52; D. W. Stahle  
**In-** and M. K. Cleaveland, *J Climate* 6 (1993): 129–39.  
**tended?]** 9. See the editorial statement in the *AER*, March 2004, 404.  
 10. ftp://holocene.evsc.virginia.edu/pub/MBH98/TREE/VAGANOV/  
 ORIG/malcolm\_29-JUL-97  
 11. ftp://holocene.evsc.virginia.edu/pub/MBH98/INSTR/TEMP  
 12. This series was included in the North American northern treeline network.  
 The Gaspé peninsula is nowhere near the northern treeline.
- [QU3:** The Gaspé peninsula is nowhere near the northern treeline.  
**Pls. sup-** 13. T. M. L. Wigley, K. R. Briffa, and P. D. Jones, *J Clim App Meteor* 23 (1984):  
**ply title** 201–13.  
**of chap-** 14. G. C. Jacoby and R. D. D’Arrigo, *Clim Change* 14 (1989): 39–59; R. D.  
**ter** D’Arrigo and G. C. Jacoby, in *Climate since A.D. 1500*, ed. R. S. Bradley and P. D.  
**cited.]** Jones. (New York: Routledge, 1992), 246–68. <QU3>
- [QU4:** 15. M. E. Mann, R. S. Bradley, and M. K. Hughes, *Nature* 2004,  
 doi:10.1038/nature02478. <QU4>  
**Pls.** 16. D. A. Graybill and S. B. Idso, *Global Biogeochem Cycles* 7 (1993): 81–95.  
**supply** 17. http://www.ngdc.noaa.gov/paleo/ei/ei\_nodendro.html  
**title and** 18. E. Zorita, F. Gonzalez-Rouco, and S. Legutke, *J Climate* 16 (2003): 1378–90.  
**date of**  
**article.]**