

WASHINGTON ROUNDTABLE  
ON SCIENCE & PUBLIC POLICY

**The IPCC,  
the "Hockey Stick" Curve,  
and the Illusion of Experience**

By Stephen McIntyre & Ross McKittrick

GEORGE C.  
**Marshall**  
I N S T I T U T E

Washington, D.C.

# The George C. Marshall Institute

The George C. Marshall Institute, a nonprofit research group founded in 1984, is dedicated to fostering and preserving the integrity of science in the policy process. The Institute conducts technical assessments of scientific developments with a major impact on public policy and communicates the results of its analyses to the press, Congress and the public in clear, readily understandable language. The Institute differs from other think tanks in its exclusive focus on areas of scientific importance, as well as a Board whose composition reflects a high level of scientific credibility and technical expertise. Its emphasis is public policy and national security issues primarily involving the physical sciences, in particular the areas of missile defense and global climate change.

## The Washington Roundtable on Science and Public Policy

The Washington Roundtable on Science and Public policy is a program of the George C. Marshall Institute. The Roundtable examines scientific questions that have a significant impact on public policy and seeks to enhance the quality of the debate on the growing number of policy decisions that look to science for their resolution.

The opinions expressed during Roundtable discussions do not necessarily represent those of the Marshall Institute or its Board of Directors. Additional copies of this transcript may be ordered by sending \$7.00 postage paid to:

The George Marshall Institute  
1625 K Street, NW Suite 1050  
Washington, D.C. 20006  
Phone: 202/296-9655  
Fax: 202/296-9714  
E-mail: [info @marshall.org](mailto:info@marshall.org)  
Website: [www.marshall.org](http://www.marshall.org)

The IPCC, the "Hockey Stick" Curve,  
and the Illusion of Experience:  
Reevaluation of Data Raises  
Significant Questions

by Stephen McIntyre and Ross McKittrick

The George Marshall Institute  
Washington, D.C.

**Stephen McIntyre** has worked in mineral exploration for 30 years, much of that time as an officer or director of several public mineral exploration companies. He has also been a policy analyst at both the governments of Ontario and of Canada.

**Ross McKittrick** is an Associate Professor in the Economics Department at the University of Guelph, Ontario, and a Senior Fellow of the Fraser Institute in Vancouver, B.C. He specializes in the application of economic analysis to environmental policy design and climate change.

***The IPCC, the "Hockey Stick" Curve, and the Illusion of Experience: Reevaluation of Data Raises Significant Questions***

Stephen McIntyre and Ross McKittrick  
November 18, 2003

*Jeff Kueter:* Good afternoon everyone. Thank you all for coming. I am Jeff Kueter, the Executive Director of the George Marshall Institute, and we are happy to co-host this event with our friends at the Cooler Heads Coalition. Our continuing interest in the science of climate change is well known and this is another in the series of events that we have been doing over the years to bring together people who are interested in climate change policy and the science behind it. Our aim is to bring in people from our network who understand the actual science of these issues to explain complicated matters in ways we can all understand and help us to be more knowledgeable as we move forward in these debates. This particular debate which we are going to hear about today has become particularly acrimonious, as those of you who have followed it are well aware. And that results in scientific process and I hope that we find, as we move forward, particularly in today's discussion which looks really at the nitty-gritty of the data, that that's the way these discussions need to evolve, because the science is what the science is and we all need to recognize that.

*Myron Ebell:* Thank you Jeff. My name is Myron Ebell. I work with the Competitive Enterprise Institute and it is my privilege to chair the Cooler Heads Coalition. The George Marshall Institute is a member of the Cooler Heads Coalition and we are pleased to co-host this event. The chairman of the National Consumers' Coalition is Fran Smith; the Cooler Heads Coalition is a subgroup of it. Jeff, I don't believe you recognized your Chairman, Dr. Robert Jastrow, who is here today, and your President, William O'Keefe. By the way, we have the authors of another paper criticizing the results of the "hockey stick" here today, Willie Soon and Sallie Baliunas.

As you probably all know, we have done a lot of these. I think this is one of the most interesting ones, because I think we are just at the be-

---

\* The views expressed by the authors are solely those of the authors and may not represent those of any institution with which they are affiliated.

ginning of what I think will be a major controversy. We have both the authors of this paper<sup>1</sup> here today. Steve McIntyre has a long career in business, particularly in the mineral exploration business, and he has degrees from the University of Toronto and Oxford University. He has a huge amount of experience with not only handling and analyzing data, but also suspecting data and suspecting the conclusions that come from large amounts of data. He will explain how he got involved in the Michael Mann paper and I'll let him do that, but I think if you look at his background, you will see that he is almost the perfect person to look at it. Our other speaker is a welcome returnee from the Far North, Ross McKittrick, whom we've had several times and who has enlightened us on several different issues. I should mention the book which he co-authored with Christopher Essex called *Taken by Storm*, which came out last year. We had a briefing about it over on Senate side in the spring. Also on the CEI website you'll find his paper on what's wrong with cap-and-trades for regulating carbon dioxide emissions. Ross has also published widely in the economic literature and he holds a Ph.D. from the University of British Columbia. Please join me in welcoming Ross and Steve.

*McKittrick*: Thank you. Thank you for coming out. The question in our title is not one that we have an answer to. Instead, we are here to address the answer that was posed by the Intergovernmental Panel on Climate Change in their *Third Assessment Report*. They took the view that the 1990s were unusually warm compared to the millennium as a whole, based on a couple of academic papers, and one, published in 1998 in *Nature* by Mann, Bradley and Hughes, one published the year after in *Geophysical Research Letters* by the same group of authors, which was an extension of the *Nature* paper. It yielded a curve that everybody is probably familiar with by now called the "hockey stick curve." (Figure 1)

It summarized Northern Hemisphere climate in terms of a temperature index that trails down slightly at a negative rate until you get to about 1900 and then it begins this dramatic series of jumps up to 1998. So this was called the "hockey stick curve." If you have looked at any of the IPCC documents recently, you can't miss it because they use it in many places. It is very prominent in the *Summary for Policymakers*, it is shown twice in the *Assessment Report* in Chapter 2, it is shown twice in the *Synthesis Report*, and it leads to their conclusions: temperatures in the latter half of

---

<sup>1</sup> "Corrections to the Mann *et al* (1998) Proxy Data Base and Northern Hemisphere Average Temperature Series" *Energy and Environment* 14(6) 751-772.

the twentieth century were unprecedented; 1990s were the warmest decade, 1998 the warmest year. They do add the word “likely” in front of those phrases. But this is an academic study and it was integral to these conclusions from the Intergovernmental Panel on Climate Change.

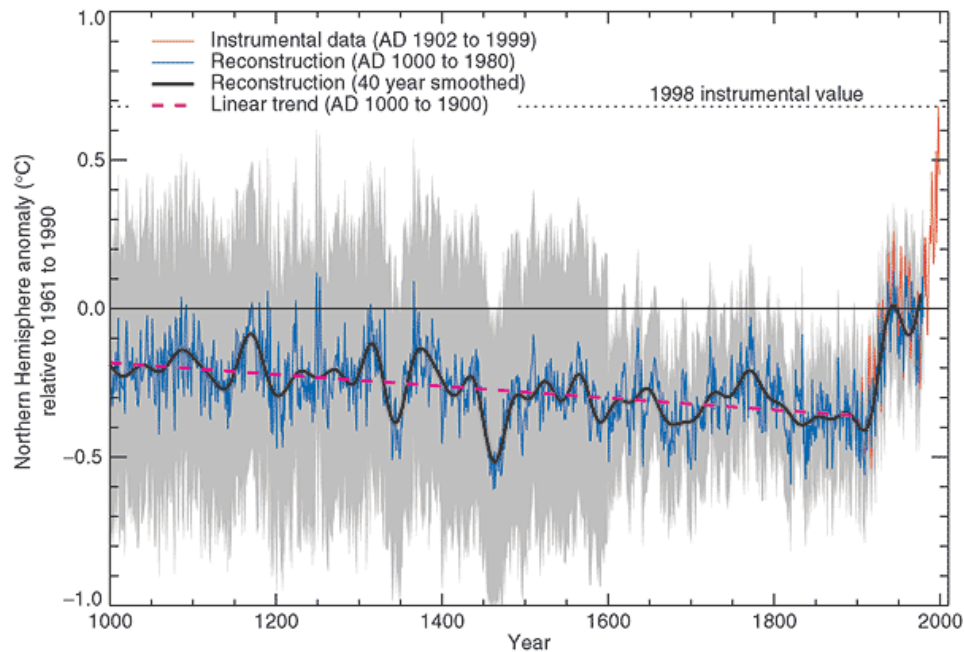


Figure 1

As an academic study, there are a number of ordinary questions that academics routinely ask when looking at these kinds of things. It is an empirical paper, so we can ask

- What data were used?
- How were the numbers crunched?
- How sensitive are the results to different ways of crunching them?
- Are there any mistakes in the data?

These are all everyday, ordinary questions that academics ask of each other's work. Now in talking about this as an academic exercise, you need to understand that there are two distinct stages of scientific review. There is the peer-review process which is a pre-publication review. It provides sometimes minimal, sometimes more extensive review, but it is a first stage quality control and it happens prior to publication. It is providing ad-

vice to the editor about whether the paper should be published, but it is not providing a definitive once-and-for-all answer about whether the results are right, only whether these results should be put out in published form. The second stage of the review is the more extensive one. That happens after publication where the work is examined, challenged and in a sort of a core practice of science, where others try to replicate the published results. The second stage is often the most important part of the review.

The paper that Stephen McIntyre and I published is an example of a Stage 2 exercise, an exercise in examining, challenging and replicating a published paper. In the paper, we analyze the data set that had been represented to us as the data behind the Mann, Bradley and Hughes 1998 *Nature* paper. In the process of analyzing it, as Steve will explain, we found many apparent errors in the data. We then rebuilt the data set from scratch using corrected and updated sources and attempted a replication of their results using the methodology that they described in their paper. We got different results than they did.

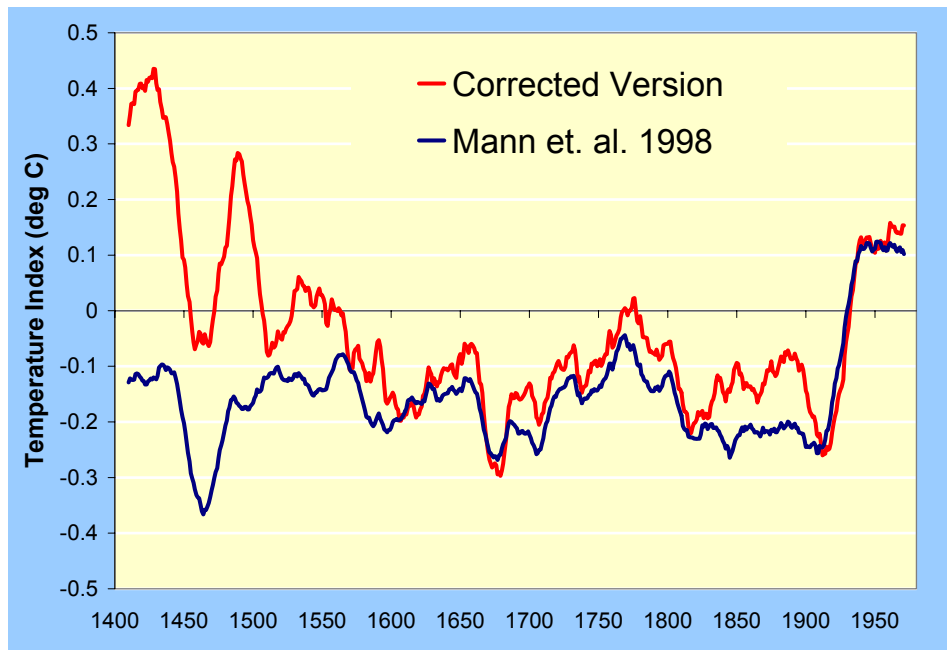


Figure 2

Figure 2 is a comparison graph. The blue line is a smoothed version of the Mann-Bradley-Hughes “hockey stick curve” back to 1400. The



red line is a smoothed version of what we derived. You can see there are minor differences back to the 1500s and then they diverge quite dramatically after that. The important point that we emphasize in our conclusion, and we will try to emphasize it again today, is that we are not arguing that the red line is the correct climate history of the Northern hemisphere. That's a statement that we are not qualified or inclined to make. What we do say, though, is that this is an authentic version of what the Mann, Bradley and Hughes data set creates, and on that basis, the conclusion cannot be asserted that the late 20<sup>th</sup> century is unusual compared to the previous 600 years.

So we have come up with a challenge to their results. This happens every day in academia. People challenge each other's results, they put out different interpretations of data, and so forth. And there is an ordinary sequence you expect to follow: you publish a challenge, you then have to reconcile any dispute about what is the appropriate data, you then have to isolate differences in analytical methods or theoretical background or in any of the data handling procedures. Once those two things are dealt with, then everyone is in a position to figure out if the original results need to be amended in some way, and then that's the end of the story. As I say and emphasize, this is everyday stuff in the world of science, in the world of economics, any kind of academic arena.

But this is no ordinary paper. There is a large political structure that has been built on the "hockey stick." The IPCC depended on it heavily; the Kyoto Protocol arguably was influenced, perhaps even strongly influenced by this set of results; there are countless government reports and countless government websites that show this graph. I was told just last week by a friend of mine who works for the federal government in Canada, the government of Canada even has a museum exhibit on the climate of the past thousand years in which schoolchildren are shown the "hockey stick curve" as a central feature of that exhibit. So this is no ordinary paper. Does that mean that challenging these results is a political act? No, absolutely not. What we are doing is an ordinary part of the scientific process. If some other people build a great political structure on a study which had not gone through its full second-stage review process, that's their problem. It's not something that we are here to address one way or another. And on that score, I was interested to read an email that we got from a scientist shortly after our paper came out that said:

*“I was one of the myriad of reviewers of the IPCC 2000 prior to its publication. One of the major concerns I expressed was the high level of credence to the Mann, et al temperature history testing without it having been seriously subjected to testing. I strongly recommended that this had some dangerous implications, should the reliance on that research prove premature.”*

So it's not like they weren't warned. But I am saying this to set aside all this secondary structure that has been built on the paper. That's other people's concerns; we are here to talk about the paper itself.

The lesson that comes from that is that there are two stages to the review process. Journal peer-review is important, but it should not be over-sold. The Stage 2 part is the ultimate check of a result, but it is a slower process; it takes time. With that, I am now going to hand you over to Stephen, who will walk you through some of the practical details.

*McIntyre:* Thank you very much for coming. My name is Steve McIntyre. I'd like to express my appreciation to Marshall Institute and CEI for paying my expenses down here. This question is in deference to David Appell who has kindly come down here to hear this presentation. I have spent most of my career in the mineral exploration business. I studied mathematics and statistics at university and I have a lot of experience in handling data and in the requirements of public disclosure. One of the things that struck me last year when the Kyoto Treaty became a big political issue in Canada, when I read the disclosure documents by the U.N., was what seemed to me to be highly promotional presentations, and highly promotional graphics – and this is from somebody who spends his career in financing speculative mineral explorations. In terms of having something where I have an expert opinion, I have pretty good qualifications to recognize promotions.

The first reaction of somebody in the mineral business presented with a set of data is to plot out some of the graphs. One of the things that struck me was how little change there was in many proxies. I looked first at were some of the proxies from Mann's 1999 paper. Figure 3 is the Greenland oxygen-18 series. Nothing much happens in the 20<sup>th</sup> century. So I started to wonder, which series are really driving Mann's results? If a lot of the series are not showing much action, there must be some series in there that are driving it. Which ones were they?

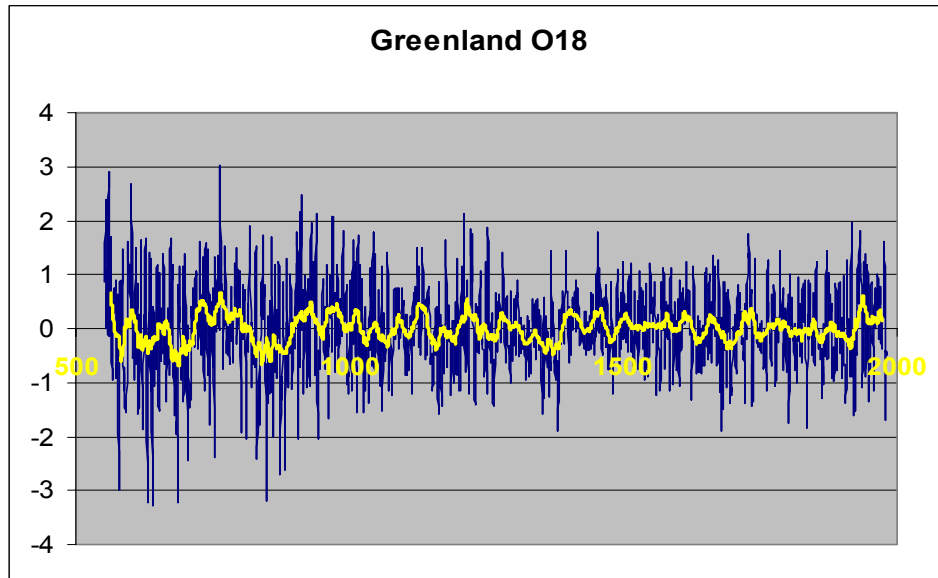


Figure 3

I thought about this for a while and familiarized myself with the issues. In April of last year I wasn't able to find any of the proxy data on an FTP site for the 1998 paper, though I found data for the 1999 paper. I emailed Professor Mann and asked him for an FTP location for the 1998 data, which I was going to plot up and see what it looked like. I had no intention other than to see – I suspected that something was driving the result, I just didn't know what. I got an odd response, an odd, very fumbling response, that they didn't seem to know where the FTP site was and they had trouble locating the data. I thought, this is a big study, there are billions of dollars being spent on it. I wasn't expecting anybody to do anything special for me, I was somebody they didn't know from Canada. But I thought if they can't find this data, maybe nobody has ever looked at it. Stranger things have happened. A couple of weeks later, I eventually got the data set and there were 112 series in it. There were descriptive files for 112 series; there were 112 series described in the *Nature* article. I thought, well, I will go to work on it and see if I can find anything interesting in this data set.

I am going to use the words "principal components" today. I have been warned by my host that if I uttered these two words, that it would instantly send the audience to sleep and perhaps it will. The point that I want

to emphasize is that we do a principal component calculation in modern software in one line on a computer. You say, I want the principal components of this data set, and you've got it. There's no magic to it; it's not something you can screw up. The results of one person and another should be exactly the same. The other thing to keep in mind is that a principal component is really an index series that summarizes a lot of data. If you think of the Standard and Poor 500 as an index series that somehow represents the patterns of change in 500 stock prices, you get a sense of what a principal component is. The analogy is not perfect, but when I say principal component, if you think about that kind of an index series, you'll be thinking about the right thing.

The first thing I tried to do was to replicate the temperature principal components in the Mann paper. We don't discuss this in our paper, but it's where I started. Mann said they used conventional principal components. To do a principal component calculation, you cannot have any missing data. The temperature data I downloaded from England had buckets and buckets of missing data. In fact, four of the cells that Mann selected seemed to have no observations in them at all, so it was impossible to apply a conventional principal component algorithm and derive the answer. I was really puzzled by this; they seemed to be doing something different than what was described in the journal. I still haven't really resolved what they did, but I just note it because this problem of missing data and its application to principal components calculations will come back a little later in the paper.

Next, Mann relies heavily on tree-ring data and he calculates principal components for six regions using 300 sites. There is a listing of the sites at the *Nature* Supplementary Information. I organized that list and figured out how to download source data from the World Data Center for Palaeoclimatology, which is funded by the U.S. government. I would like to comment that this is a tremendous archive and should be supported. I had nothing but excellent service from them and it is extremely important that there be this type of public archive of data. Collating these 300 series was a pretty big job. I carried out a PC calculation. The results were completely different from Mann's. In fact, Mann's results were literally impossible; they didn't explain enough variance in these calculations. There was again something mysteriously wrong with this and I was really quite puzzled by it. I went back to look at the data to see if I had somehow goofed in collating the data. I had a sinking feeling, after doing this for a couple of weeks, that maybe I had put the data in the wrong year and as a result, eve-

everything was a little bit at cross-purposes. I checked to see what years his data started. Mostly it started in odd years, 1999 and 1949, not the even years we like to start with. I thought I must have inserted the data wrong, so then I went back to the original email where I obtained the data. Lo and behold, the same problem was there. I hadn't collated it wrong. Whatever it was, was also in the original data. So I wrote back to Scott Rutherford who provided the data, and pointed this out to him. He said that he didn't know what the problem was, as it was before his time. I wrote to Mann and sent him back the whole data set and said, Look, is this the right data set? He said he was too busy to respond to this or any other inquiry.

So we looked at the data and said, okay, if they put in everything one year too early, as it appeared, what happened at the other end?

*The 1980 values for all 9 Stahle/SWM PC series were identical. Similar problem identified in the Vaganov and NOAMER regions – 16 series altogether.*

| Year        | PC1               | PC2               | PC3               | PC4               |
|-------------|-------------------|-------------------|-------------------|-------------------|
| 1975        | -0.03525440       | 0.06191900        | 0.01469890        | -0.03386820       |
| 1976        | -0.04758900       | 0.09825240        | -0.01345320       | 0.01161880        |
| 1977        | 0.02738590        | -0.11581500       | 0.02995960        | 0.01370230        |
| 1978        | 0.09249040        | -0.00125138       | 0.08667150        | 0.07659540        |
| 1979        | -0.01054950       | -0.17253000       | -0.00999568       | -0.04078750       |
| <b>1980</b> | <b>0.02303040</b> | <b>0.02303040</b> | <b>0.02303040</b> | <b>0.02303040</b> |

| PC5               | PC6               | PC7               | PC8               | PC9               |
|-------------------|-------------------|-------------------|-------------------|-------------------|
| 0.06205270        | -0.02129230       | 0.00062418        | 0.04612720        | -0.01503450       |
| 0.01822490        | 0.03648180        | 0.04604640        | -0.04273910       | 0.00526230        |
| 0.03782570        | 0.00327476        | 0.07170230        | 0.03729640        | -0.10195200       |
| 0.02200060        | 0.04614070        | 0.03223540        | 0.02464170        | 0.02726110        |
| 0.09144420        | -0.00608904       | -0.00508424       | -0.03537360       | -0.08408310       |
| <b>0.02303040</b> | <b>0.02303040</b> | <b>0.02303040</b> | <b>0.02303040</b> | <b>0.02303040</b> |

Figure 4

As you see in Figure 4, the 1980 values for a lot of these series were identical to seven decimal places, which is obviously impossible. So we looked at this and thought there is some monumental screw-up in this data set; this looks wrong; it is just impossible for these years to be like that. Particularly when you have got a lot of leverage in the last year of the se-

ries, if all of a sudden you have got a lot of wrong data at the end of the series.

At this time, I had lunch with Ross, with whom I had corresponded from time to time. Ross lives quite near Toronto. We had lunch where I was reviewing some of my thoughts with him and seeking some advice. For believers in omens, we had lunch at the exact hour that Hurricane Isabel hit Toronto, so we were almost “taken by storm.” Ross was intrigued by some of these questions. At that point, I had other issues in mind, there were methodological issues that were bothering me. I’d say that I certainly hadn’t sorted out what the key issues were in all of this, but I was certainly feeling pretty uncomfortable with the data that I was seeing. Ross looked at the data and found there were two different series which had identical values for twenty years. For some reason, the values in one series had been copied into another series. We looked and found that up to thirty series had 1980 values that were either plugged or had these kinds of copy errors. So in terms of relying on these closing years, in any sense, a big portion of the data was pretty meaningless. When we noticed this, we thought, well, look, there are really some problems with this data set. We will try to look at this top to bottom. We will look at every single series in this, try to get original data, and see what turns up. The first thing we found was that there was a lot of obsolete data, that when we got the source data from the World Data Center, that the newer editions had quite different looking data than the old series.

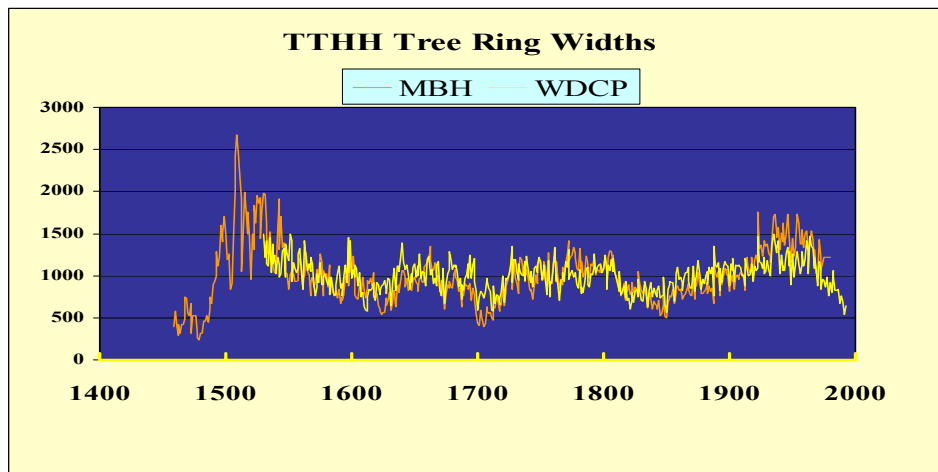


Figure 5

Figure 5 is one example and this is a series that will actually turn up a little later. The yellow shows the new version of the series, the orange shows the original version. The author of this series, Jacoby, withdrew the early data in his final version; I don't know the exact reason. Usually if they are unable to replicate sites, then they withdraw some of the data. The yellow data is the final version that was archived in 1998 or 1999.

We are not addressing the issue of whether an obsolete version was used at the time of the paper or whether the data was already obsolete at the time of the paper, though we know in some cases that was the case. Our concern is, if you redid the whole thing with up-to-date data, what's the result? In this case, it is pretty easy; we used the 1999 data rather than the 1998 data.

I will just mention in passing an issue that we don't deal with, but is relevant to anybody from a policy point of view who is relying on proxy-based information; you notice that this proxy falls off in the 1980s. So to the extent that people are saying that this proxy is in some sense an index for temperature, it should show warming in the 1980s. This proxy should be sensitive to that particular warming. If not, people have to look pretty hard at whether they are in fact proxies for temperature. That exercise is not carried out in either of Mann's papers. From my point of view, there needs to be really a pretty full-scale, engineering-quality study to follow up, probably something that's 400 pages long, to actually nail down the validity of these proxies, to look at every one, to redo the original data. In fairness to Mann, he was doing a paper in 1998. He wasn't expecting that this paper would become a centerpiece of global climate studies. I don't blame Mann for not doing his study at the engineering level of detail at that level, but somebody needs to do it now.

*Question:* Do tree rings give an accurate picture of climate history?

*McIntyre:* It is asserted that the tree rings are a proxy for temperature. I am just pointing out that if it is, then you'd expect a different result in the 1980s. I am trying to stay away from evaluating the validity of proxies. I just raised that as an open issue that other specialists should deal with. I am not trying to opine myself on whether this is or is not a valid proxy.

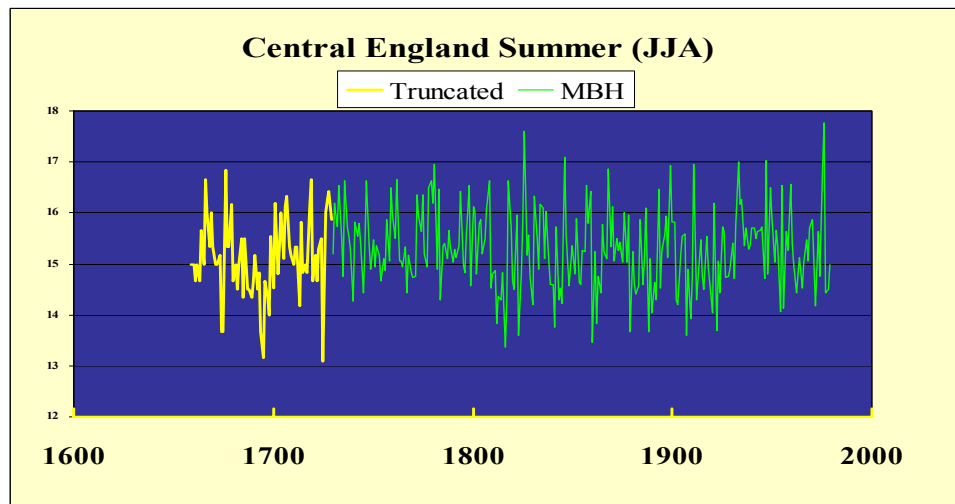


Figure 6

The next thing we noted, and this was really very strange, was that seventy-five years of data had been chopped off of the central England series. (Figure 6) The green is the data that's in Mann's study and the yellow is the data that was chopped off. The latter includes the late 17<sup>th</sup> century, the Little Ice Age period in England. The same thing was done where twenty-five years were chopped off of the Central England series. We pointed these out in our paper. Subsequently when we were directed to Mann's FTP site, we found that the exactly correct data, annualized, not truncated data, existed on Mann's FTP site. So there are duplicate versions of these series, but the truncated one is the one that was used in his paper. This is really quite a startling situation.

In total, these are the kinds of problems we found: truncated sources, arbitrary plugging of data, use of obsolete data, geographical mislabeling. Here is one that is rather fun: There is a data series that was inserted for a grid box for precipitation near Boston and the data actually came from Paris, France. This was just a crazy goof.



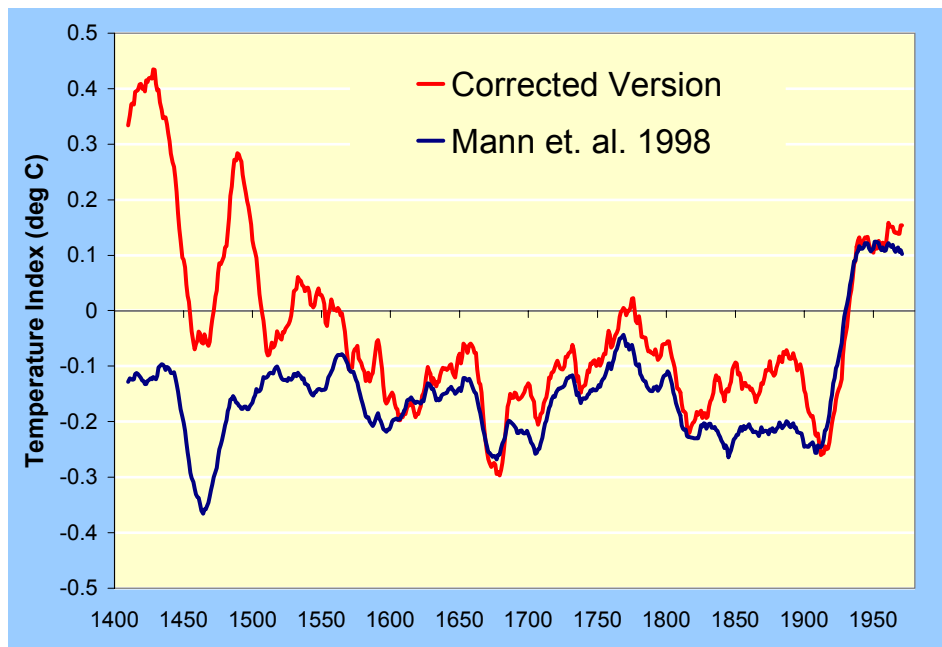


Figure 7

We then put this data all back into a new proxy data set and redid the calculations using publicly disclosed methods. We tried to get some direction and some additional information on the reconstruction methodology from Mann, without any success. We carried the reconstruction out and ended up with this result: a pretty high degree of replication in the later part of the series, but in the early part, obviously there are big differences. (Figure 7) As Ross mentioned, we have tried to emphasize that we are not saying that the 15<sup>th</sup> century was exceptionally warm. We are just saying that if you play the ball where it lies, use Mann's methodology, and use the updated data, that's what you get. So if you are saying that there is something particularly unique about the 20<sup>th</sup> century, based on this, you can't say it. It's a type of *reductio ad absurdum* argument.

*Question:* Dr. McKittrick, didn't you show slides something like that?

*McIntyre:* We showed the exact same thing

*Question:* The same thing. Your first slide is based on the tree-ring data. Is this tree-ring data?

*McKitrick*: This is the result from using the fully updated and corrected version of the data sets.

*McIntyre*: It's the same paper. They are not different. We are not in competition on this.

*Question*: What is the last year on this data?

*McKitrick*: We took it up to 1980, which is the last year of the proxy reconstruction that Mann did.

*Question*: What was the event in 1450 that caused the tremendous drop in temperature?

*McIntyre*: Maybe Pat Michaels or Fred Singer can tell you. I am just trying to comment on data issues.

That's the end of the first chapter. We published the paper. It has attracted some interest. The first and I guess the most active reporter on this is David Appell, who is right here. He has been a keen follower of this story. He has not been a supporter of ours by any means, but he has paid attention to us. The story that David wrote from talking to Mann was that we had requested an Excel spreadsheet; that Mann had directed us to his FTP site, but we insisted on an Excel spreadsheet; that they in their infinite kindness prepared this, but the associate who did it accidentally made some mistakes in collating the data and that we had failed to notice that there were errors in this collation. As a result, all our results were spurious and that the right data was at his FTP site. We looked at his FTP site; we were actually directed there by a reference from David's site. Lo and behold, the identical file that was sent to us was already listed on his FTP site, dated a year earlier. As well, there was a Matlab version of the identical data on the same day. So however this file was created or whatever errors or non-errors were in it, it was obviously at least a year old and it wasn't prepared as a special-purpose file for us, it was prepared much earlier. Then in a very interesting turn of events, these two files were then deleted from Mann's FTP site. Given that we had very public derogatory statements made against us for using incorrect data, this is surprising. We were alert and went to the FTP site on October 29 and copied it all, so we got copies of the data, but if we had been a day or two later, this evidence would have been removed.

As to the suggestion that we had failed to notice the errors in this file that had been sent to us, it seemed to us a very odd response, since we had spent twenty pages in minute details talking about errors in this file and all the errors that we had supposedly not noticed, we had described in great detail. In fact, we had gone to the extent of collating 300 series from scratch in order to obtain new principal components calculations. So we firewalled ourselves from these particular data errors. I was actually a little surprised at the resonance of the suggestion that we had got the wrong data. While there was, I guess, a slight smug satisfaction for those people thinking we had used the wrong data, it was a criticism that didn't bother me because I knew we hadn't, and so it wasn't a criticism that in any sense stung.

I want to emphasize that the collation errors only affect thirty-one principal component series. We looked at eighty-one series, where there had been no principal component calculations. We traced these series back into the uncollated data in Mann's FTP site and found that all the same problems that we had outlined still existed in this FTP site. So we know these criticisms carry forward. Mann has said that he didn't make the collation errors in the 1998 paper and I think that that is actually possible. I think it is possible that one of his Ph.D. students did a study a couple of years ago and that he sent us a data set that resulted from that study a couple years ago. I don't know that, but I am not excluding that. The way you can eliminate speculation on this is pretty easy: you just simply produce the correct collation or you produce a computer program showing that you are reading in the series right. Quite frankly, in his shoes, I'd do that in a heartbeat. But instead he has refused to provide this information. I think it puts Mann in a bad light. It is a pointless kind of exercise, because he will have to produce his series and data at some point. The other problem is this: he just pointed us to his FTP site and that contains over 430 principal component series and we were invited to pick seventy-eight that were actually used in the study, with no description. Again, he has got to identify them; it is not simply enough to say, well, try to guess the right series.

The next response, and this was a very interesting response, they published a paper saying that we incorrectly omitted three key indicators and they did a recalculation showing that if they omitted these three indicators, that they would get a 15<sup>th</sup> century result that would look almost exactly like ours. (Figure 8)

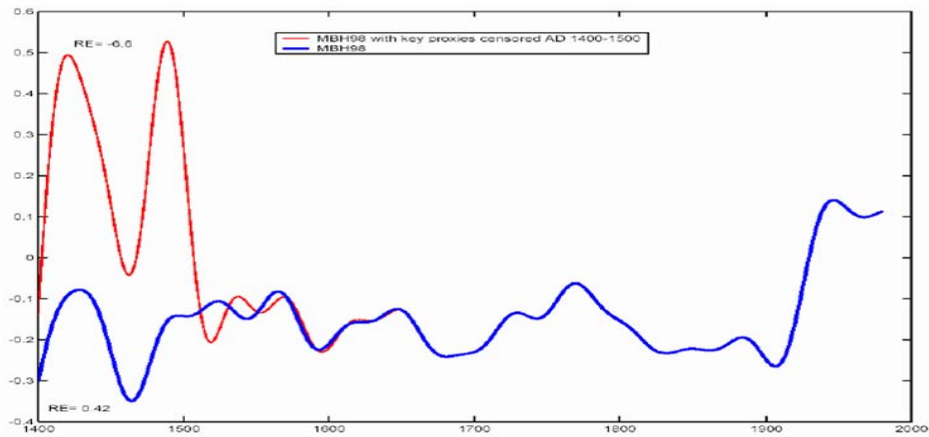


FIGURE 1. COMPARISON OF MBH98 RECONSTRUCTION (BLUE) WITH RECONSTRUCTION RESULTING FROM THE ELIMINATION OF KEY PROXY DATA SETS (1)-(3) OVER THE AD 1400-1500 INTERVAL. THIS YIELDS ESSENTIALLY THE SAME RESULT OBTAINED BY MM BY ELIMINATING A SIGNIFICANT FRACTION OF THE MBH98 DATA AVAILABLE FOR THAT PERIOD (BOTH SERIES HAVE BEEN SMOOTHED WITH A 40 YEAR LOWPASS FILTER).

Figure 8

Again, this caused a certain amount of satisfaction in the early returns. Our size-up on this was really quite different. First of all, they show that the entire reconstruction really depends on three key indicators. So we are talking about 112 proxies, but whatever these three indicators are, you get very different results and you get results entirely like ours. The other thing that I found very satisfying is that it showed that even though we were doing a reconstruction based on poor public disclosure, that we had replicated the major ingredients of his methodology because we had two graphs that looked pretty much the same, depending on the presence or absence of these three indicators. Again, from a policy point of view, you would say what on earth are these three indicators that we are deciding to spend billions and billions of dollars on? Again, I would like to have a 500-page report on these three indicators.

So we have looked at them. One thing I just want to say is that we didn't omit anything. I will explain: certainly these indicators became unavailable. I mentioned to you that principal components don't work with missing data. In some of these site rosters, some of the sites were missing data in the 15<sup>th</sup> century, so the indicators simply became unavailable. So it wasn't that we omitted anything, it's just that using a principal component algorithm, that's what happened.

The three series are actually quite interesting. (Figure 9) One of the three series that we are accused of omitting is the series that I showed you above [TTHH tree ring widths], where we used a non-obsolete version of the series. The obsolete version went back seventy-five years earlier and had very low values in the 15<sup>th</sup> century. Actually in terms of somebody who asked about what accounted for the low in the middle of the 15<sup>th</sup> century in their version, probably this series contributes an awful lot to it. As I mentioned before, we didn't subtract this data from the series; the original researchers subtracted it. So whatever their reasons were for subtracting it, we consistently relied on the most up-to-date version. We make no apology for using this indicator.

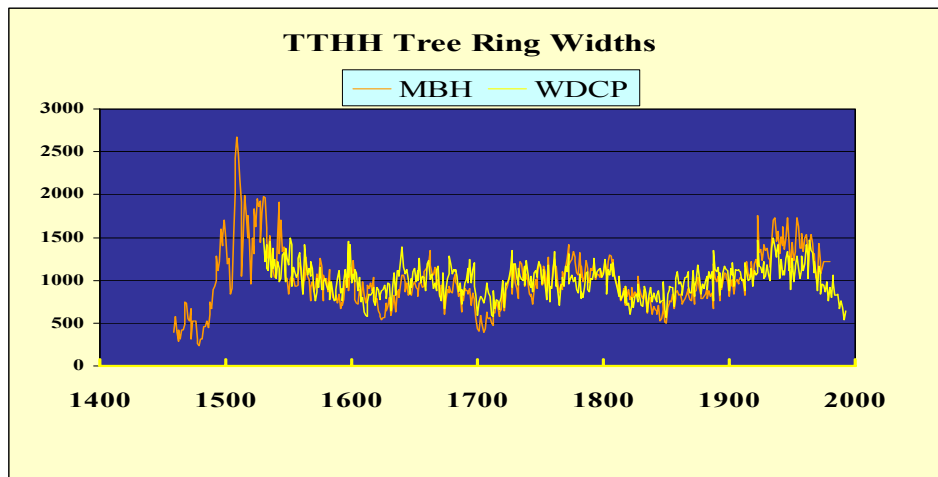


Figure 9

The second indicator is a principal component for the Southwest-Mexico region. We haven't reported on this formally, but what we found there is that there were many very elementary data quality issues in that. By the time that this series was taken back to the 15<sup>th</sup> century, there were only three sites in the series. There was a difference between the disclosure documents and the FTP documents and on the FTP site, he used one site twice with slightly different versions. The site that he used twice, interestingly enough, was a site at Spruce Canyon, Colorado. It was not a site that was listed in the original Stahle study. Stahle had no sites from Colorado or New Mexico. Exactly what this site is doing in this region is mysterious, so that two of the three sites that were used in this key indicator, upon which Kyoto rises or falls, were sites that didn't belong in the original region and were slightly duplicate versions of one another. On a more fun-

damental basis, he has a North American region which has sites from Alaska to Georgia, and in the middle of this, there is this little region in Texas and Oklahoma which is carved out as a separate region, completely mysteriously. Interestingly enough, the Spruce Canyon, Colorado series also occurs in the North American region, so it not only is duplicated twice in the Stahle series, it also occurs in the other region. This is not what we felt was a high-quality indicator and again, we don't make any apologies for not including it in our 15<sup>th</sup> century data set.

The third key indicator was his North American principal component. What Mann did to make it available was to change the roster of sites in the 15<sup>th</sup> century to the available sites. This procedure of changing rosters was not disclosed in the original publication. I think it is a material disclosure because better statisticians than us might very well have wondered about the validity of this procedure. But we are not taking up that particular cudgel here. We adopted Mann's procedure and said, okay, we will re-include that indicator back into the mix. We found that there was a discrepancy between the sites disclosed in his *Nature* disclosure and the sites actually used at his FTP site. We used the disclosed sites, recalculated it, and got an answer that was pretty much the same as where we started. So the difference seems to lie in the differences in these rosters, but this one indicator calculated with the disclosed (as opposed to the actually used) data actually doesn't overturn anything.

I will wind this up now. I just want to point out there are really two quite different kinds of issues here. One is just the problems in the data itself and the other is the assessment of the impact of the problems. The response to our paper so far has mostly been criticizing our assessment of the impact of the data errors. Nobody at this stage has made a denial of the existence of the use of obsolete data, no denial of the truncation of data. So whether we have completely replicated Mann's reconstruction methods, we have certainly tried to do so; based on public disclosure, I think we have done so. The fact that our results so closely match Mann's in the presence or absence of those three indicators gives me some confidence that we have captured the key features of it. The next step, or an important step, would then be for them to disclose the actual computer programs that they used to select the sites and to carry out the calculations again. Given the kind of controversy this already has gotten, I would certainly do that in a heartbeat. There is no reason not to. I guess we have previewed here that there is also an underlying issue of how these series were selected. This is a theme that we are going to address in some other

work because it is actually a pretty important issue. Now that we have looked at the FTP site, we have got many questions about the selection of series.

The question that was asked was whether the 1990s were the warmest decade in the past 1,000 years. Our answer before was that Mann's methodology applied to corrected and updated data does not enable them to say that. We don't make any assertions ourselves as to whether it was or it wasn't. Also we want to say that having received two rounds of responses, we stand entirely behind everything we have said. None of the responses have touched any important issues and in fact, if anything, we believe that they have confirmed the principal points of our analysis.

I would like to add that we have put up on our websites every computer program that we have used to make these calculations; we have put up where we have made fresh collations of these 300 tree ring series; we have put up the data files with their collations. We have tried to be as transparent as possible in our disclosure, so if we have made an error somewhere, it is easy for someone to spot and that everything is as transparent as possible.

\* \* \*

#### *Questions and answers.*

*Question:* Pat Michaels, University of Virginia. I think what you're really uncovering here is a larger and pervasive problem in science, which is the peer-review process seems to be missing important and obvious issues, perhaps failing because of the sociology of global warming science. I would like to just take a minute to explain to the audience and see if I can get their comments on it. What the methodology was that was used here because it's not clear to everyone: a series of trigger mechanisms were trained on data ending in 1980. Those triggering mechanisms explain about half the variation in temperature from when the training set begins in the 19<sup>th</sup> century, ending in 1980. When you take the principal component, formed like the index of them, that explains roughly about 50% of the proxies. So you are down to 50% times 50% of the variation in the temperature. Now after 1980, the temperature record goes up, the surface record, everyone knows this; it goes up beyond 1980. Because so little of the behavior of

the training record remains in the proxy, that guarantees mathematically that the period from 1980 to the end of the record will be the warmest in the analysis. Why was this not picked up in the peer review process?

*McKittrick*: Beats me. Obviously we don't have any insights into the kind of questions *Nature* asked in the review stage.

*Question*: Do you agree with my mathematics?

*McKittrick*: I certainly agree with the point about the way this graph is put together, by taking temperature data and splicing it to a larger data set. It uses, as you call it, "training," or just generating a statistical mapping, so that it can then use the proxy data back here and feed it into a calculation that will spit out representative temperature data. The explained proportion of the temperature data is not 50%. Once you move back to the 1800s, the explained portion with the available proxies declines much more rapidly. As to the question of peer review, I will turn it over to Steve.

*McIntyre*: I want to take that. I am not as hard on peer review as most people. You couldn't expect a peer reviewer to do the kind of work that we did on this. If you required that in peer review, which is an unpaid job, it would have a chilling effect on people publishing stuff. A peer reviewer says, "I have no beef with this paper being published as it is." As I mentioned before, at the time this paper was originally published, it wasn't the centerpiece of the UN study. As somebody who has been involved in feasibility studies, I refer to the requirement to do engineering-quality work on some of these things before you start making large investment decisions on them. I think at the next stage, the IPCC stage, there should have been a much more thorough review. That's the stage where I think there was an incorrect reliance, but that's not a peer review, that's a matter of saying, the international public is viewing the IPCC as a professional organization that is carefully evaluating the data. If they were relying on a paper that had only been peer reviewed, the public thought there was much more due diligence than that. This analogy is from a business background: a peer review has less due diligence in it than an audit, so that essentially it is the equivalent of unaudited financial statements. These essentially unaudited materials have passed through a big chain of usage without any engineering-level verification. I refer from time to time in saying, if someone wants to make proxy-based histories of this stage, you need to do a 400-page report, you need to get a whole bunch of really good scientists to do it, tear apart all the proxies, do it from scratch and see what you get. I don't



blame Mann in any sense for not doing that; that wasn't what he tried to do in the first place.

*Question:* Jay Ambrose. I wonder if the two of you have faced criticisms that goes beyond ordinary scientific disputation and if you had, could you describe those?

*McKittrick:* Before this came out, we showed it to a quite a lot of colleagues in a variety of disciplines. A few of them said, "Steel yourselves; you are going to be attacked, you are going to be slammed." I didn't actually expect that we would be, and we haven't. This has obviously generated some lively discussion and I am sure there are people who would much prefer that this had never been done. My impression is that within the scientific community, the response is pretty much what I expected. They recognize that this is a serious paper raising serious questions. There are some issues that are going to have to be sorted out and everyone is going to hold their judgment in check until that process has really worked itself out.

*McIntyre:* Actually Jay and I corresponded in the past. I once sent a letter to Jay on a completely different topic and we had the nicest correspondence where we vehemently disagreed. For people with completely different political views, he gave me a very nice response.

*Question:* Fred Singer. I'd like to address the point that Pat Michaels raised. It is an important point. Could you put the IPCC "hockey stick" on, please? I want you to notice something. I was a reviewer on the IPCC report and in the first draft that I saw, the Mann curve going back to 1000 was in black. The instrumental curve based on temperature thermometers was in blue. You couldn't tell the difference, you couldn't tell them apart unless you looked very closely. They then changed it to red, but the initial one was in blue. The thing I noticed, and you can see it fairly clearly here, is that the Mann analysis stops in 1980 and then the "hockey stick" is really entirely due to the thermometer data, which as you probably know are suspect, or at least they are under attack by the people who believe, as I do, that the satellite data are more nearly correct. We can argue about that later. In other words, the surface data, the thermometer data, are in controversy.

Now I corresponded with Mann and I have this email correspondence which I am now digging back and I will publish for every one. I asked him, why did you stop in 1980? Why didn't you go forward to the

year 2000 or 1998, the date of his paper? His reply was very strange. He said there were no suitable data available, proxy data, that is. I knew this was not the case. I have found more than half a dozen proxy data between 1980 and 2000, none of which showed an increase in temperature. Some showed a decrease in temperature. I then started to pursue this subject and I am now focusing my efforts on trying to see what all the proxy data show after 1980. Steve McIntyre has been very helpful in sending me a whole bunch of data. I have not found any yet that show an increase in temperature. In other words, the proxy data disagree with the thermometer data in the last twenty years; they do not show a warming. I have published that in a number of places and I want to do a full, complete publication, if the referees in *Science* will accept it. Now the question is why did Mann not use data after 1980? His excuse is a lame one; it is just not true. The answer I think is that if he had used proxy data after 1980, he would have found them to be in disagreement with the accepted, politically correct surface data from thermometers and it would have destroyed his calibration. Also it would have destroyed the IPCC, so he preferred to stop his analysis in 1980. I think that is the real reason, but I have not got him to admit this yet. Maybe we will.

*McKitrick*: I am not sure I have a comment. We didn't really go into that. I do know that in the data set we were sent, there are some series that extend past 1980. You could easily get up to 1984 with a reasonable data set.

*Question*: His email says that there are no suitable data.

*Question*: I have a question about some of the proxies before 1500. Have you done a statistical analysis about what would happen to the entire reconstruction if you included those key series?

*McIntyre*: Let me just jump forward to the diagram. One thing that I want to stress about this picture: we did not draw this picture. Mann drew this picture. In the reply that he wrote, he suggested that we had deleted seventy-five or some large number of series. Remember, a principal component takes a large matrix and just represents it as a single index, so what we are talking about here is just one principal component. There is still a lot of pre-1500 data in our graph; otherwise we wouldn't have had any values at all back then.

*McKittrick*: Think of it as the Tree Ring 70, along the lines of the Standard and Poor 500.

*McIntyre*: There are only three series, though, that are removed out of however many, forty or fifty, that are available in the 15<sup>th</sup> century.

*Question*: So about 70% of the data were removed?

*McIntyre*: Well, first of all we didn't "remove" it. It is a matter that under the principal component calculation, it was unavailable. The seventy series that he described are summarized into one indicator so that there is only one of 112 proxy series that was affected by this calculation. As we also mentioned, we have subsequently re-analyzed it, in which we changed the site rosters, as he now discloses that he did. He never previously disclosed that he changed site rosters, but if we change the site rosters trying to follow his methodology, it doesn't make much difference. He needs three indicators in place to make that difference; two of them are clearly not usable. We re-inserted the third one and we find that the values are more like the red one with only one of the three indicators back in.

When Mann talks about seventy series, in fact the disclosed series are even more than that, we have actually included in our preliminary recalculation about 77 or 78 series instead of 70, because he has excluded several disclosed series from the ones he actually used for no apparent reason. When we calculate the Tree Ring 77, it looks a little different from his Tree Ring 70, but it doesn't affect our conclusion very much. We will be responding to that on a more formal basis, but our size-up right now is that it won't make any difference.

*Question*: David Appell. You said you talked to Ross around the time of Hurricane Isabel came up on the East Coast, so that you didn't submit your paper to *Energy and Environment* before Hurricane Isabel. I was wondering when you did send it to *Energy and Environment* and if the peer review process there was only a few weeks long, how much reliability can you have for the peer review process at *Energy and Environment*? Secondly I was wondering why you chose not to respond in *Nature* or *GRL*, given that that is where traditionally you would respond, since that is where the original paper appeared.

*McKittrick*: Since this is all posted on the website, I am sure you have the answer already, so I will just respond for everyone else's benefit. Our first strategy when Steve and I talked about this was that since the paper was published in *Nature*, our submission should be to *Nature*. The problem is that *Nature* has a 1,500-word limit for this kind of submission. We wrote it up to that word limit, showed it around to a bunch of colleagues and the response, even from people who were familiar with Mann's work, was that they just couldn't make sense of what we were doing. There just wasn't enough word space there. So the advice that we got, which I think was correct advice, was to publish it somewhere where you can spell out the whole argument at once and then follow up with a communication to *Nature* when there is something there that can be done in a crisp 1,500 word format. So that was the plan.

As for the peer review at *Energy and Environment*, well, the whole point of our paper is not to overrate peer review at a place like *Nature*. So I don't think there is any danger people are going to say that as a result of reading our paper, they are overrating the role of peer review. Peer review is, like I say, a first stage quality control process. It is advice to an editor whether to put this into play in published form. If anyone is working under the misapprehension that peer review means this stuff is infallibly correct, then I would hope they had been disabused of that long ago. Peer review just means the editor was advised that this is solid enough and deserves to be published in this journal and it would be interesting to our readers.

*Myron Ebell*: I would like to point out that cold fusion analysis was peer reviewed, but one could replicate the results.

*Question*: So when did you submit it to *Energy and Environment*?

*McKittrick*: I will look it up later and tell you, if it matters that much to you.

*Question*: Can you give me an approximate date now?

*McKittrick*: It was between Hurricane Isabel and today.

*Question*: Bob Hershey. You pointed out this data set, or the early part of it, was later declared obsolete by the author. It seems to cover this period where there is the controversy between the two curves. I wonder if the au-

thor who had declared his data obsolete has indicated why he wanted to withdraw it.

*McIntyre*: We had no reason to inquire. There were many series that we used the later versions on. All we were trying to do at that point was see what the impact was of using up-to-date data and whether the results were stable to later data versions. In fact, I think that the use of later data is probably one of the most important things that accounts for the differences in the results. There's a lot of discussion about the collation errors and so on, and they were the things that caught my eye in the first place. But I think in fact the differences in data versions are much more substantial in driving the differences in results. When I talk about the question of data selection being a problem, I can't help but think that if they happened to use the current version of the data and got that sort of result, I can't help but suspect that they would have changed their selection of proxies so that the answer looked like more of what they wanted.

*Question*: I'd like to respond to some of the points you've made. I'd like to point out that McKittrick and McIntyre's data do not respond to the study by Mann, *et al*, patterns; this study showing the findings the patterns and temperatures and it reveals that this hockey stick pattern is shown by about ten other different independent studies by different authors. I have those figures here, showing the different models and they all agree that there is an increase in temperature and they also show the "hockey stick" pattern.

*Myron Ebell*: Could I respond to that first? It is well established in the literature for decades from Hubert H. Lamb on that there was a Medieval Warm Period and a Little Ice Age. If you can show that in the "hockey stick," then you have made a *prima facie* case; otherwise I think what you have told us does not have anything to do with their analysis of this paper. Do you have anything to add to that?

*McKittrick*: I'd just like to say that we are not offering a rival climate history. I mean, I am not particularly wedded to the red line. What we are showing is what you get if you take the data set that is specified in the *Nature* disclosure, collate it correctly using updated sources and apply his methodology. And if the result of that contradicts what other people have published, that is not our problem; that would be Mann's problem.

*McIntyre*: If this result or this methodology is not stable to updated data, then essentially Mann's result is meaningless, so it is impossible for these

other studies to confirm something that is itself meaningless. That his result looks the same is just an accident because when you do up-to-date data, you get a different look. Now we bit off a lot in this paper. I think that it is quite reasonable to address some of these other studies. I have certainly looked at some of them and I can assure you that I have got big questions about how some of these other studies were done and I feel fairly confident that if I go through them with a fine-toothed comb, that I would have something interesting to say about them. But that's another day.

*Question:* This is a basic disclosure question. What led you to take on this project? How were you funded and have you analyzed other related climate studies?

*McKittrick:* First of all, on the funding: we did not receive any money from anyone to do this. I have basically blown away my fall sabbatical doing this; it wasn't what I planned to do and the sooner it's over, the happier I will be. But we didn't get funding from anyone for doing this. We didn't ask for any and we didn't receive any. As for what got us into it, Steve has told his story: just being suspicious about the graphs. I had seen some postings that Steve made on the internet where he was working through the data and occasionally posting some notes about what he was finding. But I didn't even know he lived in Toronto until he sent me an email and said, you are not that far away, can I ask you some questions about statistics and methodology. We got together and at that point I thought it was interesting on many levels, but in particular when you have some basic problems with the underlying data, I think academics have a duty to help get that kind of information available. Not so that I can get involved in this field; it's not my field and I have no designs on getting into it, but so his colleagues can understand what the data are, how he did his results, and then they can bat it around. It just sort of fell into my lap and that's why I am doing it.

*McIntyre:* I think I more or less answered it earlier. This is *costing* me money to do. Normally I would be working on some business deals. I spent quite a bit of time on this and I found it quite interesting. Fortunately I have had some stocks go up; it has been a good market for junior exploration companies, but normally right now I would be trying to do some business. My wife asked me whether I am going to start earning money again.

*Question:* Aloysius Hogan. I have heard questioning of the statistical and methodological practices associated with a number of papers and I would like to get an opinion from you both about the level of statistical and meth-

odological analysis among normal peers. Are the people who are doing the peer review really qualified in those areas as statisticians or they are just educated laymen?

*McKittrick*: Now are you talking about the journal peer review or the IPCC review process?

*Question*: I am talking about the peer review for four or five different cases.

*McKittrick*: It is up to the editor of a journal to choose the reviewers and presumably they choose people who are competent to review this. A couple weeks ago I reviewed an economics paper for a journal. It was a study of variations in water pollution levels in India. I didn't ask to see their data and I didn't ask to see the printouts of the stats packets because it is a very simple, straightforward data collection process and I know where they got their data from and it is a straightforward regression analysis and the results look plausible and fit into the literature and there aren't actually huge implications one way or another. If they were putting forward some results that contradicted what other people had been saying, had huge policy implications and was going into a high profile journal, then I would have wanted to see their data, I would have wanted to see their computer printouts and I would have wanted to have them verify that they could analyze the data in a number of different ways and basically get the same answer back. So in part the kinds of questions the reviewer is asking is triggered by the paper itself. I have to suspect that Nature has a group of extremely competent reviewers who, if they thought to ask the questions, would have learned some of these things.

\* \* \*

## RECENT WASHINGTON ROUNDTABLES ON SCIENCE AND PUBLIC POLICY

Hans Mark – *Sea-based Defense against Ballistic Missiles* (October, 2003)

Michael Canes – *Unraveling the Puzzle: Differing Economic Estimates of Climate Policy* (September, 2003)

Robert Gallo – *The AIDS Global Health Crisis: Strategies for Policy and Science for Its Resolution* (September, 2003)

Gregory Canavan – *Missile Defense for the 21<sup>st</sup> Century* (September, 2003)

Michael Gough, Roger Bate & Henry Miller – *A Discussion of Politicizing Science* (July, 2003)

Marc Landy & Charles Rubin – *Civic Environmentalism: Developing a Research and Action Agenda* (June, 2003)

James Oberg – *Toward a Theory of Space Power: Defining Principles for U.S. Space Policy* (May, 2003)

Willie Soon – *Was the 20<sup>th</sup> Century Climate Unusual? Exploring the Lessons and Limits of Climate History* (May, 2003)

Harvey Rubin – *Interdependent Security - Approaches to Protecting an Increasingly Connected Society* (April 2003)

Sallie Baliunas and Willie Soon – *Extreme Weather Events: Examining Causes and Responses* (March 2003)

Randall Correll – *National Security Implications of the Asteroid Threat* (February 2003)

***The Marshall Institute – Science for Better Public Policy***



GEORGE C.  
**Marshall**  
I N S T I T U T E

**Board of Directors**

**Robert Jastrow, Chairman**

*Mount Wilson Institute (ret.)*

**Frederick Seitz, Chairman Emeritus**

*Rockefeller University*

**William O'Keefe, President**

*Solutions Consulting, Inc.*

**Bruce N. Ames**

*University of California, Berkeley*

**Sallie Baliunas**

*George Marshall Senior Scientist*

**Thomas L. Clancy, Jr.**

*Author*

**Will Happer**

*Princeton University*

**Willis Hawkins**

*Lockheed Martin (ret.)*

**Bernadine Healy**

*Cleveland Clinic Foundation*

**John H. Moore**

*Grove City College*

**Robert L. Sproull**

*University of Rochester (ret.)*

**Chauncey Starr**

*Electric Power Research Institute*