

## **Do We Know That The 1990s were the warmest decade of the Millennium?**

**Stephen McIntyre**

**March 9, 2009**

**New York**

### **2009 International Conference on Climate Change**

[SLIDE] Do we *know* that whether the 1990s were the warmest decade of the millennium as we so often hear?. In my opinion, no. However, the opposite is also the case: neither do we *know* that the Medieval Period was warmer than the present.

This is because the so-called “proxies” in current use are inconsistent; and, at present, this inconsistency is an insurmountable roadblock to answering the question. Minor changes of even proxy versions from nearby sites can yield opposite medieval-modern differentials. Problems with bristlecone tree rings may be familiar to some of you, but there are equally intractable problems with tree rings in Siberia.

While I find these issues interesting both statistically and analytically, I readily concede, as some have argued, that the issue may well not “matter” for the “big problem” – Jor-El and the survival of the species sort of thing. Fair enough. But then policy makers and the public surely have the right to be annoyed at IPCC and others for so prominently featuring what they now say to be an irrelevancy in their public expositions of the AGW problem.

[SLIDE] Contrary to later claims at realclimate and similar sites, let there be no doubt that the Mann Hockey Stick was not an incidental appendage to the Third Assessment Report, something that was isolated by skeptics as a weak stray from the flock. It was used over and over and could almost be said to be the logo of the IPCC 2001 report. The picture here shows it behind John Houghton as the backdrop for the press conference announcing the Third Assessment Report.

[SLIDE] The Mann Hockey Stick gave rise to the sound bite: “the 1990s were the warmest decade and 1998 the warmest year”, a sound bite that was adopted by the Canadian government as their lead argument in raising public awareness. Indeed, that’s how I first heard of AGW as an issue.

[SLIDE] The Hockey Stick was also prominently displayed in Inconvenient Truth, where Al Gore spliced the Mann hockey stick with the CRU instrumental temperature record and mistakenly called the spliced series “Dr Thompson’s ice core thermometer”, in effect citing the Mann hockey stick as “independent” proof of itself. Gore took a passing swipe at the so-called the “fierce attack” of “skeptics” against the Stick – thus, I guess, giving Ross and I a backhanded citation.

[SLIDE] The original Mann hockey stick is retained in the 2007 IPCC Fourth Assessment, now as one strand in what I’ve termed a “spaghetti graph” – a graph in

which the various reconstructions agree on very little except that the late 20<sup>th</sup> century is slightly warmer than the Medieval Period.

Referring to this spaghetti graph, Mann observed some time ago that critics are not facing merely a Hockey Stick but an entire Hockey Team, a term that I cheerfully adopted at Climate Audit to describe Michael Mann, Gavin Schmidt and their associates.

In 2008, Mann re-entered the fray with a new MBH reconstruction, using an even more obscure statistical methodology than the original paper.

## **STATISTICAL ANALYSIS**

[SLIDE] Paleoclimate reconstructions are, in statistical parlance known outside the paleoclimate community, a form of “multivariate calibration”, a point that Ross and I raised in our recent PNAS comment on Mann et al 2008 published. Rather than applying known statistical methods to appraise his reconstructions, Mann and associates have developed their own ad hoc methodologies, the properties of which are poorly understood and which, as we observed on an earlier occasion with Mann’s modified principal components method in the original Hockey Stick paper, may have unforeseen defects..

In our 2009 comment on Mann et al 2008, we applied the consistency test of Brown and Sundberg, a prominent multivariate calibration article, to the proxies of Mann et al 2008, finding that the proxies were so inconsistent that it was impossible to establish a finite confidence interval prior to the 19<sup>th</sup> century. I anticipate that other studies would have similar results. In reply, Mann provided no support for his methodology in statistical literature, instead asserting that their method was “conventional”, citing two climate articles by his sometime coauthors..

[SLIDE] In a survey of statistical problems in ecology in the early 1990s, Jan de Leeuw, a prominent applied statistician, made the obvious requirement of a valid model that it be “stable” to “small and uninteresting perturbations”<sup>1</sup> - specifically mentioning stability to data selection and to minor variations in technique. In fact, this sort of test, typically using the word “robustness” is often made in paleoclimate. The problem is that “robustness” claims in the literature are often, shall we say, artful and sometimes even untrue, so that the exact status is often unclear with careful dissection of the analysis.

[SLIDE] This sort of sensitivity analysis was an important aspect of our articles and many Climate Audit posts. One section of our 2005 E&E article was entitled the “Effect of slight variations on 15<sup>th</sup> century temperature results”, where we reported that slight

---

<sup>1</sup> We usually do not want a small and uninteresting perturbation of our data to have a large effect on the results of our technique. Classical statistics has always studied stability by using standard errors or confidence intervals. Gifi thinks this is much too narrow and other forms of stability are important as well...

variations in retained principal components and variations in the presence/absence of bristlecones had a noticeable impact on results, resulting in early 15<sup>th</sup> values of a Mann-type reconstruction exceeding late 20<sup>th</sup> century values. Since we did not believe that a slight tweak of Mannian proxies and methods would magically yield a valid reconstruction, we did not present this sensitivity analysis as an alternative temperature history, but as a demonstration that questionable methodological and data selection decisions in the original article had a significant impact on results..

[SLIDE] The results are unstable because, in the MBH AD1400 network, only the North American tree ring PC1 dominated by bristlecones and the equally problematic Gaspé cedars have a hockey stick shape. If these series receive less weight, then there is nothing in the other 20 series in the network that generates a hockey stick shape.

[SLIDE] Surprisingly, given everything that you might read, there is not actually any dispute between the parties on a sufficiently well-specified calculation.

On the top left, is a figure from Mann et al's reply to our 2003 article, in which they also obtained very high early 15<sup>th</sup> century values by excluding bristlecones – that's not how they described the calculation; they described us as having “thrown out” essential data – data that we later determined to be the bristlecones. The top right shows a figure from Bürger et al (2006), confirming and extending our findings that some seemingly innocuous methodological variations yielded 15<sup>th</sup> century values higher than 20<sup>th</sup> values. On the bottom left, our 2005 results are re-plotted, together with two scenarios by Mann's frequent coauthors, Wahl and Ammann, in which they varied the methodology and number of retained principal components in the middle and the presence/absence of bristlecones on the right. The Wahl and Ammann results and ours reconcile to 7 9s accuracy, though you'd never know it from the literature. All four groups show that the presence/absence of bristlecones or methodological variations that substantially change their weights alter the 15<sup>th</sup> century-20<sup>th</sup> century differential.

[SLIDE] A panel of the U.S. National Academy of Sciences considered some of these issues in 2006 and reported that the Mann reconstruction wasn't robust to the presence/absence of bristlecones, no other interpretation for the code words “proxy records from individual regions” being possible. Ironically, they credited Wahl and Ammann for this observation rather than us.

[SLIDE] Despite the fact that four different groups, two of which included Mann and his close associates, had confirmed that these variations resulted in 15<sup>th</sup> century values exceeding 20<sup>th</sup> century values, the IPCC Fourth Assessment Report's last word was that the impact of these problems was negligible, only about 0.05°C, far below the amount necessary to affect the relationship of early results to 20<sup>th</sup> Century results. As IPCC reviewers, Ross and I protested this characterization of the matter articulately and in detail, but our comments were disregarded.

[SLIDE] The IPCC claim was based on an assertion by Wahl and Ammann that, if they added enough principal components back into the mix, they could once again get a hockey stick shape. By adding more principal components, the bristlecones were re-introduced as a lower order principal component series.

Wahl and Ammann presented a very strange argument for including bristlecones regardless of whether there was a relationship to local temperature. They conceded that the Mann reconstruction had no validity without using bristlecones. They argued that this very failure somehow demonstrated that the bristlecones contained climatic information at the “level of eigenvector patterns in global surface temperatures”. In the plain light of day, the language is absurd, but that didn't stop IPCC from relying on it as the last word on the matter.

[SLIDE] In doing so, the IPCC also ignored specific findings of another 2006 panel, this one led by Edward Wegman, chairman of the NAS Committee on Theoretical and Applied Statistics. Wegman said that Wahl and Ammann's tactic of adding back enough principal components to get a Hockey Stick had no “statistical integrity”. As an IPCC reviewer, I asked that the Wegmen Report be considered and cited; IPCC refused.

Returning to the original question of stability to data selection and methodology, regardless of any spin from Mann and his associates, I submit that it is impossible to contemplate a smaller and more “uninteresting” methodological variation than whether 2 or 4 covariance principal components should be used in a reconstruction. A valid reconstruction simply should not turn on such a point – an observation that we made clearly in 2005. I also submit that it is hard to contemplate a more appropriate test of data selection stability than the presence/absence of bristlecones. The Mann reconstruction fails on both counts, and, in fact, failed to even live up to its own robustness warranties<sup>2</sup>.

Quite separately from these issues, in our 2005 articles, we reported many questions from specialist literature about whether bristlecones were valid proxies in the first place. The NAS panel stated that strip bark proxies, another code word for bristlecones, should be avoided - a recommendation that was promptly ignored by both IPCC and the paleoclimate community, which, if anything, redoubled its use of bristlecones and even Mann's discredited PC1.

## **A Dozen Independent Studies**

---

<sup>2</sup>MBH99 obscurely alludes to the impact of bristlecones on the verification statistics for AD1000 results (neglecting to mention their effect on the presence/absence of long-term trend), previously asserted in MBH98 to be robust to the presence/absence of all dendroclimatic indicators (and, a fortiori, to bristlecones). MBH99 explicitly denied that the post-1400 network of MBH98 was affected, a claim that was untrue. Mann et al 2000 made no mention of the caveat alluded to in MBH99 and asserted in categorical terms that the reconstruction was not sensitive to the presence/absence of dendroclimatic indicators.

[SLIDE] The other supposed support for the Stick has been that similar results have supposedly been obtained by more than a “dozen independent studies”. The claim was made in Mann's 2003 Senate testimony, by Wigley in a newspaper interview responding to our 2003 article, in Mann’s 2005 letter to the House Energy and Commerce Committee and remains in Wikipedia to this day.

[SLIDE] The IPCC Fourth Assessment was a little more guarded, recognizing that the data was “not entirely independent”<sup>3</sup>.

[SLIDE] However, this completely misrepresents the true situation. In fact, the dozen supposedly “independent” studies re-cycle the same data over and over. Bristlecones are used 9 of 12 studies in the spaghetti and versions of Polar Urals and Tornetrask are used in every study in the spaghetti graph. This lack of independence means that, if, for any reason, there should be a problem arise with bristlecones, Polar Urals or Tornetrask or all of them, every IPCC reconstruction is affected. And this proves to be the case.

### **Bristlecones**

[SLIDE] At face value, Mann's bristlecone ring width chronologies show little difference between the Little Ice Age and Medieval Warm Period in California. This is at odds with ecological information – it has been known since the 1970s that medieval bristlecones grew well above the present tree line. In 2006, Constance Miller and associates used ecological niche information to estimate that California was about 3.2 degrees warmer in the medieval period than at present. As a reviewer, I asked IPCC to cite this article; they refused.

[SLIDE] In 2006, the bristlecone chronologies, originally collected by Donald Graybill in the 1980s, were over 20 years old. Whether they were right or wrong, they represented the results of only one researcher. Mere prudence suggests that Graybill’s bristlecone sites should be re-sampled. Given the relative warmth of the past 2 decades, this would be an ideal out-of-sample test of their validity as unique world thermometers in warm conditions – a point that we forcefully raised in a 2005 Op Ed. There's an extra reason in this case – ironically, Graybill was trying to show that carbon dioxide fertilization influenced growth. Arch-skeptic Sherwood Idso was a coauthor of the Graybill article originally publishing the bristlecone series relied upon by Michael Mann. In 2003, Idso was astonished to learn that this data had been used as the active ingredient in the hockey stick.

Despite the urgent need for modern data, the IPCC Fourth Assessment asserted that, unfortunately, there was no recent data at the key tree ring sites. Mann had previously justified this lack of due diligence by saying that updates would have require “heavy equipment” and expensive travel to out of way sites. I’m familiar with mining exploration

---

<sup>3</sup> As with the original TAR series, these new records are not entirely independent reconstructions inasmuch as there are some predictors (most often tree ring data and particularly in the early centuries) that are common between them, but in general, they represent some expansion in the length and geographical coverage of the previously available data (Figures 6.10 and 6.11

and, at Climate Audit, I teasingly formulated what we called the Starbucks Hypothesis: that a UCAR scientist could have a latte at Starbucks in the morning, update the bristlecones in the day and still be home for dinner. More on this later.

[SLIDE] In fact, contrary to IPCC claims, some important sites had been updated, including the most important site in the Mann reconstruction, the Sheep Mt bristlecones, which had been updated in 2002 by Linah Ababneh, then a PhD student at the University of Arizona. MBH coauthor Hughes was on her thesis committee. Her results totally failed to replicate Graybill's distinctive hockey stick – Graybill's results are shown on the left; hers on the right. This was not mentioned by IPCC. Mann et al 2008, for which Hughes was a coauthor, perpetuated the use of the obsolete Graybill chronology rather than the up-to-date Ababneh results, making no attempt whatever to reconcile the discrepancy or to justify the decision.

[SLIDE] In 2007, another Graybill bristlecone site at Mt Almagre in Colorado was updated. This is now both the most up-to-date and highest 1000 year tree ring chronology in the world. Instead of ring widths increasing in the 30 years since 1980, they declined and currently are more or less at their long-term average.

[SLIDE] The samplers also proved the controversial Starbucks Hypothesis. At 7 a.m., they had coffee at the Starbucks in Colorado Springs and by 9 a.m. were ready to sample bristlecones. You may recognize one of the samplers.

[SLIDE] Pete Holzmann, a Climate Audit reader, managed to locate some of the trees tagged by Graybill in the 1980s. This was pretty lucky as there were no maps and the coordinates on file were not very precise.

[SLIDE] From our work at Almagre, we think that there may be an amusing explanation of the Hockey Stick. Strip bark trees are trees where the bark has died around the circumference, leaving only a strip, often leading to bewildering contortions. In one test of two cores about 9 inches apart, one core had a 6 sigma growth pulse lasting over a century and then subsiding, while the adjacent core didn't. We think that the 6-sigma growth pulse may be a mechanical reaction to the bark stripping off and have nothing to do with temperature or carbon dioxide. In the usual small sample of 20 or so cores, it is obvious that even a couple of 6-sigma excursions could have a profound impact on the average. This is not a simply theoretical concern as Graybill *said* that he intentionally selected strip bark and we identified this precise problem in one of Graybill's tagged trees. From seeing incipient strip bark, Pete Holzmann thinks that strip bark may originate from broken branches, which in turn may come from heavy snowfall. Ironically, the explanation of the Graybill chronologies may be a century long strip bark pulse, with severe weather in the US West in the 1840s occasioning an unusual incidence of strip bark formation. An odd explanation indeed for the Hockey Stick.

### **Polar Urals**

[SLIDE] There are still three IPCC reconstructions that don't use bristlecones. But each

of these – Jones et al 1998; Briffa 2000 and D'Arrigo 2006 – is unstable to one data version decision – a choice between two versions of Polar Urals data.

[SLIDE] The 1998 Jones reconstruction had only 3 series in the 11<sup>th</sup> century, one of which was Briffa's then recent Polar Urals chronology published in 1995. Briffa's Polar Urals tree ring reconstruction was an opening shot against the then prevalent concept of a Medieval Warm Period, claiming that 1032 was the coldest year of the millennium.

[SLIDE] As in California, this chronology was at odds with ecological information. In 1995, Shiyatov, a Russian specialist, reported that the medieval treeline was well above current treelines and that the medieval period was a time of record growth. In 2004, Naurzbaev et al estimated that the medieval period was 1.5 to 3 deg C warmer than at present. As a reviewer, I asked the IPCC to cite this article; again they refused.

[SLIDE] It turned out that the Briffa chronology was based on only 3-4 poorly dated cores in the early 11<sup>th</sup> century - far short of usual dendro standards. In the late 1990s, new material was cross-dated 11<sup>th</sup> century, yielding a totally different result: an elevated Medieval Warm Period

[SLIDE] Briffa did not publish the new Polar Urals data (nor did anyone else) – I obtained the data only through quasi-litigation with Science. In his 2000 reconstruction, instead of using the updated Polar Urals series, Briffa substituted a new chronology from Yamal, about 90 miles away from the other site. Remarkably, unlike the updated Polar Urals series with its elevated medieval period, the new chronology had a pronounced hockey stick shape.

[SLIDE] Briffa's Tornetrask versions were used in all the spaghetti graph studies. However, in 2006, a new Tornetrask version was published without Briffa. The data had been collected some years earlier. Once again, the effect was that the medieval period was considerably enhanced relative to the modern warm period.

[SLIDE] Contrary to IPCC claims, these instabilities really do matter for the spaghetti graph studies. On the left, I've shown the spaghetti graph from the NAS panel report, and, on the right, a corresponding spaghetti graph derived by minor variations in data selection – in each case using only data sets that have been used in one of the spaghetti graph studies.

[SLIDE] By now, I'm sure that you're sick of tree ring data. But ice core delta O18 – Lonnie Thompson's hockey stick – offer little help in resolving the dilemma. The chart at right shows 6 O18 series over the past 1000 years from north to south. Greenland (top) and inland Antarctica (bottom) show little change in the period. Law Dome in Antarctica (second from the bottom) had elevated medieval delta O18. Mt Logan, Yukon (second from top) shows the opposite effect to the one generally expected. 20<sup>th</sup> century delta O18 values are much lower than 19<sup>th</sup> century values – the authors attribute this not to decreasing temperatures, but to changing regional circulation. Fair enough, but once you invoke that explanation, there is no basis for being sure that increasing O18 values at (say) Dasuopu in the Himalayas are not also a result of regional circulation changes.

[SLIDE] One also needs to consider glacier retreat evidence, which may well point towards the modern period being warmer than the medieval period. In Alberta in 1999, a retreating glacier exposed in situ stumps that were dated about 2800 BP, stumps that do not appear to have been exposed in the medieval period.

[SLIDE] At Quelccaya glacier in Peru, Lonnie Thompson has identified plant remains from the retreating glacier, some of which are dated about 4500 BP and which do not appear to have been exposed in the medieval period.

[SLIDE] In the Swiss Alps, Joerin and associates have dated wood fragments from receding glaciers, primarily prior to the medieval period. They posited that the Alps were green in Roman times, illustrating their concept with the colorful comparison to the present day shown here.

So there is definitely glacier retreat evidence that points to the modern period being warmer than the medieval period. But this evidence is a two-edged sword for proponents of unprecedentedness. The very evidence showing that the modern period is warmer than the medieval period also shows that plants and trees grew even higher in even earlier warm periods. The plant remains recovered by Lonnie Thompson from the receding Quelccaya glacier are up to 400 meters higher than their present limit.

[SLIDE] Now we come to a harder question. Does any of this matter in a big picture sense?

[SLIDE] Mann's realclimate colleague, Gavin Schmidt, has stated that the position that the Stick is the "least important" figure in understanding climate:<sup>4</sup>

If this is the "least important" figure, third parties are surely entitled to ask why IPCC, Al Gore and others used it so extensively. Wouldn't the IPCC and others have carried out their responsibilities more effectively if they'd focused on presenting the "most important" figures to a public that is starving for knowledge? And having publicized the "least important" figures so much, surely they are in no position to blame others, if the public has trouble understanding their message. Maybe they should look into the mirror and try to insure that they focus on what's important. Indeed, on the basis that the hockey stick was as irrelevant as Gavin Schmidt said, I suggested that the Fourth Assessment Report save space by deleting the entire topic and focusing on matters that were important. They thought otherwise and thus we have the spaghetti graph.

[SLIDE] Having said all this, in preparing for this conference, it occurred to me that there is one version of the Hockey Stick type diagrams that actually is relevant to policy makers and the interested public. On the left, I show a figure from the 2001 Synthesis Report (not the WG1 report), which splices the Mann hockey stick with the various IPCC scenarios. On the right, I've plotted the despised Lamb schematic on a corresponding scale together with Fourth Assessment scenarios. In my opinion, given their forecasts, IPCC could reasonably illustrate the scale of change under their forecasts as compare to

---

<sup>4</sup> [http://sciencepolicy.colorado.edu/prometheus/archives/climate\\_change/000641reflections\\_on\\_the\\_c.html](http://sciencepolicy.colorado.edu/prometheus/archives/climate_change/000641reflections_on_the_c.html). See also Stefan Rahmsdorf <http://www.realclimate.org/index.php?p=114>



something like the Medieval Warm Period, even under Lamb's view. This turns the question back to the validity of the forecasts – which has always been the big question – and away from the intricacies of obscure statistical manipulations of tree ring data.

[SLIDE] On the other hand, there is another line of opinion that the Hockey Stick is important because paleoclimate information is used in estimating climate sensitivity and, if the Stick is wrong, ironically the real situation is much worse than we think. Regardless of whether this position is right or wrong, I have a simple answer: well, if that's the case, we'd better find out if the Hockey Stick is correct and govern ourselves accordingly. And if it matters, other people besides me should take an interest in critical analyses of the methods and data used in these reconstructions. And efforts to obstruct the determination of whether the Stick is right or wrong through withholding of data and code should be condemned not just by me, but by the broader climate science community.