

# Caspar and the Jesus Paper \*

August 22, 2008

At some time or another, most people will have seen the hockey stick – the iconic graph which purports to show that after centuries of stable temperatures, the second half of the twentieth century saw a sudden and unprecedented warming of the globe. This was caused, we were told, by mankind burning fossil fuels and releasing carbon dioxide into the atmosphere. For a while, the hockey stick was everywhere – unimpeachable evidence that mankind was damaging the planet – an impact that would require drastic measures to reverse. The stick’s most famous outing however was just a couple of years ago when it made a headlining appearance in Al Gore’s drama-documentary, *An Inconvenient Truth*. The revelation of the long, thin graph with its dramatic temperature rise in the last few decades, and the audience gasps that accompanied it, is something of a key moment for many environmentalists.

Shortly after its publication, the hockey stick and its main author, Michael Mann, came under attack from Steve McIntyre, a retired statistician from Canada. In a series of scientific papers and later on his blog, Climate Audit, McIntyre took issue with the novel statistical procedures used by the hockey stick’s authors. He was able to demonstrate that the way they had extracted the temperature signal from the tree ring records was biased so as to choose hockey-stick shaped graphs in preference to other shapes, and criticised Mann for not publishing the cross validation  $R^2$ , a statistical measure of how well the temperature reconstruction correlated with actual temperature records. He also showed that the appearance of the graph was due solely to the use of an estimate of historic temperatures based on tree rings from bristlecone pines, a species that was known to be problematic for this kind of reconstruction.

The controversy raged for several years, involving blue riband panels, innumerable blog postings, endless name-calling and dark insinuations about motivations and conflicts of interest. In May 2005, at the height of the controversy, and on the very day that McIntyre was making a rare public ap-

---

\*A version of this paper was published at [www.bishophill.squarespace.com](http://www.bishophill.squarespace.com) on 11th August 2008. It attracted worldwide interest and a great deal of supportive comment. The version presented here has been updated slightly, mainly for corrections of spelling and punctuation, but also in one place to clarify some of the issues around RE benchmarking

pearance in Washington to discuss his findings, two Mann associates, Caspar Ammann and Eugene Wahl, issued a press release in which they claimed that they had submitted two manuscripts for publication that together showed that they had replicated the hockey stick exactly, confirmed its statistical underpinnings and demonstrated that McIntyre's criticisms were baseless. This was trumpeted as independent confirmation of the hockey stick. A few eyebrows were raised at the dubious practice of using a press release to announce scientific findings. Some also noted that on the rare occasions that this kind of announcement is made, it tends to be about papers that have been published, or at least accepted for publication. To make such a dramatic announcement about the *submission* of a paper was unusual in the extreme.

The first of these papers ("the GRL paper") was submitted to *Geophysical Research Letters*, the journal of the American Geophysical Union. It took the form of a rebuttal of a McIntyre paper that had attacked the hockey stick and had been published in the same journal. From the first, the McIntyre paper had been controversial. Apart from Ammann and Wahl's paper, there were three other papers taking issue with it. However, it turned out that some of these attempted rebuttals were less well formed than others. In fairly short order, Ammann and Wahl's paper was rejected, many of its criticisms either relating to other McIntyre papers than the one at hand, or relying on the CC paper for their arguments. Since the CC paper was unpublished, it was effectively impossible for McIntyre to defend himself against these criticisms. Shortly after Ammann and Wahl's paper was rejected, another of the rebuttals, that of a physicist called David Ritson, was also shot down by the journal's editors.

Meanwhile the second, longer paper ("the CC paper") had started its long road to publication at the journal *Climatic Change*. This article purported to be a replication of the hockey stick and confirmation of its scientific correctness. However, in a surprising turn of events, the journal's editor, prominent global warming catastrophist Steven Schneider, mischievously asked none other than Steve McIntyre to be one of the paper's anonymous peer reviewers.

We have seen above that one of the chief criticisms of the hockey stick was the fact that its author, Michael Mann, had withheld the validation statistics so that it was impossible for anyone to gauge the reliability of the reconstruction. These validation statistics were to be key to the subsequent story. At the time of their press release Wahl and Ammann had made public the computer code that they'd used in their papers. By the time their paper was submitted to *Climatic Change*, McIntyre had reconciled their work with his own so that he understood every difference. And he therefore now knew that Wahl and Ammann's work suffered from exactly the same problem as the hockey stick itself: the  $R^2$  number was so low as to suggest that the hockey stick had no meaning at all, although another

statistic, the reduction of error statistic (or RE) was relatively high. It was only this latter figure that had been mentioned in the paper. In other words, far from confirming the scientific integrity of the hockey stick, Wahl and Ammann's work confirmed McIntyre's criticisms of it! McIntyre's first action as a peer reviewer was therefore to request from Wahl and Ammann the verification statistics for their replication of the stick. Confirmation that the  $R^2$  was close to zero would strike a serious blow at Wahl and Ammann's work.

Wahl and Ammann's response was to refuse any access to the verification numbers, a clear flouting of the journal's rules. As a justification of this extraordinary action, they claimed that they had shown that McIntyre's criticisms had been rebutted in their forthcoming GRL paper, despite the fact that the paper had been rejected by the journal some days earlier. At the start of July, with his review of the CC paper complete, McIntyre took the opportunity to probe this point, by asking the journal to find out the anticipated publication date of the GRL paper. Wahl and Ammann were forced to admit the rejection, but they declared that it was unjustified and that they would seek publication elsewhere.

With the replication of the hockey stick in tatters, reasonable people might have expected some sort of pause in the political momentum. Seasoned observers of the climate scene, however, will be unsurprised to hear that global warming eminences grises like Sir John Houghton and Michael Mann continued to cite the Wahl and Ammann papers despite the CC paper being in publishing limbo and the GRL paper being apparently dead and buried. The Wahl and Ammann press release was not withdrawn either.

Events soon took another surprising turn., It was announced that the editor in chief of *Geophysical Research Letters*, Jay Famiglietti, had taken over the file for the McIntyre paper and its responses. This was justified, he claimed, because of the high number of responses – four – that the McIntyre paper had received. That two of those responses had been rejected and were no longer in play was not mentioned. The reason for the change quickly became apparent though when, at the end of September, the rejected response from David Ritson turned out not only to have been re-submitted but had also been accepted for publication. This was another clear breach of the journal's rules, which required that an article's author should be able to comment on responses before they were accepted. Famiglietti however refused to make any on-the-record comments about why he behaved as he did.

If McIntyre had any suspicions about the implications of Famiglietti's malfeasance, he must have been quite certain when, shortly afterwards, hockey stick author Michael Mann commented on his RealClimate blog that both the CC and the GRL papers were going to be accepted shortly. Sure enough, in the last week of September, the GRL paper was resubmitted and revisions were made to the CC paper. Both papers were back in play again.

As 2005 neared its end, two important events loomed large. The first was the year end deadline for submission of papers for the IPCC's Fourth Assessment Report on the state of the climate, and realisation soon dawned on McIntyre and the observers of the goings-on at GRL: the IPCC *needed* to have the Wahl and Ammann papers in the report so that they could continue to use the hockey stick, with its frightening and unprecedented uptick in temperatures. Mountains were going to be moved to keep the papers in play.

The other important happening was the fall meeting of the American Geophysical Union, which would be attended by many of the big names in paleoclimate and at which both McIntyre and Ammann would be making presentations. McIntyre's plan was to use the question and answer session after Ammann's presentation to once again press for the  $R^2$  number for the hockey stick, a figure that had never been released, despite it being constantly requested over the previous years by McIntyre, journals, politicians and journalists. Sure enough, when confronted, Ammann once again prevaricated.

After the session, McIntyre attempted to clear the air by inviting Ammann to lunch. In the circumstances, this seems to have been a relatively amicable affair, but McIntyre's suggestion that he and Ammann write a joint paper outlining where they agreed and where they differed was not taken up. When McIntyre later formalised this offer in an email, Ammann failed even to acknowledge it.

While the AGU was meeting in San Francisco, *Climatic Change* had provisionally accepted Wahl and Ammann's CC paper, any objections which might have been raised by McIntyre swept aside by simple means of not inviting him to review the second draft. The resubmitted version of the paper turned out to be almost identical to the old one, except that a new section on the statistical treatments had been added, presumably as a condition of acceptance. And here there was an upside because, buried deep within the paper, Ammann and Wahl had quietly revealed their verification  $R^2$  figures, which were, just as McIntyre had predicted, close to zero for most of the reconstruction, strongly suggesting that the hockey stick had little predictive power. Their decision to reveal these key data is necessarily obscure, but may well have been prompted by McIntyre's decision to file a complaint of academic misconduct about Ammann with his employers, UCAR. Although the complaint was rejected, it may well have put sufficient pressure on Ammann and the journal to show the numbers that everyone wanted to see.

The CC paper's provision acceptance date was December 12th, just a few days before the AR4 deadline. Strangely, the version that was accepted seems to have been dated 24th Feb 2006, so according to its rules the IPCC shouldn't have been able to consider it. And what is more, it appears that the new sections discussing the statistical verifications were only added in

this post year-end version. As McIntyre put it:

“So under its own rules, is IPCC allowed to refer to Ammann and Wahl [2006]? Of course not. Will they? We all know the answer to that. When they refer to Ammann and Wahl [2006], will they also refer to its confirmation of our claims about MBH verification  $R^2$  statistics. Of course not. That information was not available to them in December. But wait a minute, if Ammann and Wahl was in press in December, wouldn't that information have been available to them? Silly me.”

In other words, the version of the paper which had gone forward to the IPCC didn't include the adverse verification statistics, but the version accepted by the journal did. The IPCC got their rebuttal of McIntyre and the journal got a fig leaf of respectability to cover up its duplicity.

By March, the CC paper had been fully accepted, but there was to be another hiccup that would threaten its existence. After all the shenanigans at GRL, with the replacement of the editor and the resubmission of letters, the journal decided once again to reject Wahl and Ammann's attempt to rebut McIntyre's work. Ostensibly this was because the arguments were “already out there”, but the truth was surely that there were so many holes in the statistical arguments as to make their publishing an embarrassment to the journal.

This new rejection was a problem for the CC paper, as I will explain below. When using an  $R^2$  verification, researchers can refer to tables of benchmarks to gauge the significance of their results. Now that the fact that the hockey stick and Ammann and Wahl's replication of it were public, Ammann was arguing that the correct measure of significance was in fact the alternative RE statistic. His problem was that for RE statistics, there are no tables of benchmarks for the researcher to refer to – he has to establish a benchmark of his own by other means. And Ammann had done this in the GRL paper which had just been rejected. Without the GRL paper, he couldn't even argue that his results in CC were statistically significant.

There is a rule of thumb for RE statistics: this says that positive RE numbers have some significance while negative ones do not. Unfortunately for Ammann, this rule applies only to linear regressions; as the hockey stick was clearly not linear, it couldn't apply. The original hockey stick authors had claimed that they had created a benchmark through other means, and that the figure was still zero. Now, while they had been silent on the issue in their original GRL submission, Ammann and Wahl announced in their resubmission that they had performed benchmarking calculations and that had confirmed that the significance level for the RE should remain at zero.

However, now that the resubmission had been rejected by GRL, the establishment of this benchmark was cancelled out, and the statistical arguments in the CC paper which relied on it could no longer be maintained.

And then silence. A year later, the CC paper was nowhere to be seen, despite having been accepted for publication. It was stuck in a kind of publishing limbo once again. This left the IPCC and Climatic Change with a problem. McIntyre observed:

“I’m intrigued as to what the final Wahl and Ammann version will look like. They have an intriguing choice: the inclusion of a reference to this article in [the Fourth Assessment Report] was premised on their article being in press which would prohibit them from re-working their article to deal with the GRL rejection. But the article needs to be re-worked since it will look pretty silly to describe their GRL article as “under review” over 18 months after it has been rejected.”

In the background, however, much had been happening. Suddenly, in September 2007, and with the IPCC report published, the CC paper suddenly appeared in print, preceded in the same journal by another paper by the same authors. What had happened was that Wahl and Ammann were quietly allowed to rewrite their rejected GRL paper and submit it to *Climatic Change* instead. All reference to the rejected GRL paper in the CC paper could be replaced by reference to the new paper, (which I will call the Jesus paper, in light of its extraordinary resurrection and for lack of any less confusing name). With identical authorship, and a maze of cross-references between them, the two CC papers were carefully designed to make understanding how their arguments relied on each other as difficult as possible.

The beauty of this approach was that it allowed for retention of the original acceptance date for the CC paper, and hence its inclusion in the IPCC process. It did leave them with the embarrassing problem that a paper that was allegedly accepted in March 2006 relied upon another paper that even the journal itself said was only received until August (and in reality, even later than that). Readers should note that this matters because unless the paper was accepted by the journal by the deadline, it should not have been accepted by IPCC for inclusion in the Fourth Assessment Report. But the IPCC *needed* the CC paper and despite the inconsistency being pointed out to them, the IPCC they waved the objections aside as irrelevant.

The argument in the CC paper leads from the text, to the appendix and then onto the Jesus paper. At places in the Jesus paper, the argument referred back to the CC paper, creating a neat, if logically flawed, circular argument. One notable feature of the CC paper and the Jesus paper was that they relegated some of their key argumentation to their Supplementary Information (SI) sections, online appendices to the published papers. In particular, the Jesus paper stated that the statistical discussions and more precisely, the establishment of RE benchmarks could be seen there. To have key arguments in the SI was most unusual and it quickly became apparent

why it had been done: the SI was *nowhere* to be seen. Even the peer reviewers appear not to have had access, and once again, Ammann refused McIntyre’s request for the data and code. His reply to this request was startling (and remember that Ammann is a public servant):

“Under such circumstances, why would I even bother answering your questions, isn’t that just lost time?”

Again, everything fell silent. For the next year nothing more was heard of the two papers. McIntyre pressed from his blog for release of the SI and the politicians were able to quietly take advantage of the political space created by the IPCC report.

Then, just a few weeks ago, and entirely unannounced, the Supplementary Information suddenly appeared on Caspar Ammann’s website, some three years after that first press release announcing the refutation of McIntyre’s work. With it, and a godsend to McIntyre, was the code used to establish the benchmark for the RE statistic. With no more than a few days work, McIntyre was able to establish exactly what had been done.

You will remember that Ammann and Wahl had claimed that they had established a benchmark of zero for a 99% significant RE score – that is to say, there is only a 1% chance that you might have got that score by chance. McIntyre had, much earlier, shown that if you ran red noise through the process, you could get RE scores of more than 0.5. (Red noise is best described as a random walk – a line which wiggles at random, but is not entirely random like white noise.) To reduce your chance of random error to 1% you actually needed to score 0.54 for RE. How Ammann had come up with zero as his benchmark was a mystery.

Now, with the code in front of him, McIntyre could see exactly what Wahl and Ammann had done. And what they had done was to calculate almost exactly the same figure as he had! The number they had arrived at was 0.52, just a whisker away from McIntyre’s own 0.54, but they had reported to the world that it was sufficient only to score a positive number – a result which only applied to simple linear scenarios! Of course, this wasn’t picked up by the peer reviewers because, as we’ve seen, they didn’t have access to the Supplementary Information, but the IPCC’s purposes had been served - the hockey stick found its way intact into the Fourth Assessment Report, unscathed by skirmishes with inconvenient statistical truths.

However, the figure of 0.52 was insufficient for Wahl and Ammann’s purposes. Their problem was that the key component of the hockey stick had a verification RE of 0.48, leaving it tantalisingly just below the calculated benchmark. They needed it to be in the top rank, and getting it there was going to be tricky. For each simulation, a thousand runs through the statistical sausage machine were performed and the RE number, the correlation

with the temperature record, was recorded. Then all the runs were sorted in order of RE value, the best runs having the highest RE and the worst the lowest. Wahl and Ammann needed to show that the hockey stick RE was right up there with the best simulations - in the top one percent. While its RE was high, it wasn't good enough. And it was no good simply removing runs which had a higher score than the hockey stick, since this would not increase its position enough – they would have been reducing the total number of runs as well as the number of runs which were scoring better than the hockey stick. To get the answer they needed, the higher scoring runs had to be made to be lower than the hockey stick, but left in the calculation.

To do this, Wahl and Ammann came up with a value which they called a calibration/verification RE ratio. As the name suggests, this was the ratio of the two RE numbers for calibration and verification. This ratio is however, entirely unknown to statistics, or to any other branch of science. But it was not plucked out of the air. The ratio and the threshold value which was set for it by Wahl and Ammann was carefully calculated. They argued that any run with a ratio less than 0.75 should be assigned a score of -9999. Since the hockey stick had a score of 0.813, 0.75 was pretty much the highest level you could go to without rejecting the hockey stick itself. However if you set your ratio threshold too low, not enough runs would be rejected and the hockey stick would no longer be 99% significant. Some of the results of this ratio were entirely perverse – it was possible for a run that had scored a reasonably good RE in the calibration (there was a good correlation between it and the actual temperatures) to be thrown out of the final assessment on the grounds that it had done *very* well in the verification – the correlation with actual temperatures was considered too good!

With this new, and pretty much entirely arbitrary hurdle in place, Wahl and Ammann were able to reject several of the runs which stood between the hockey stick and what they saw as its rightful place as the gold standard for climate reconstructions. That the statistical foundations on which they had built this paleoclimate castle were a swamp of misrepresentation, deceit and malfeasance was, to Wahl and Ammann, an irrelevance. For political and public consumption, the hockey stick still lived, ready to guide political decision-making for years to come.