

Dummies guide to the latest "Hockey Stick" controversy

by Gavin Schmidt and Caspar Amman: [Online version](#)

Due to popular demand, we have put together a 'dummies guide' which tries to describe what the actual issues are in the latest controversy, in language even our parents might understand. A [pdf version](#) is also available. More technical descriptions of the issues can be seen [here](#) and [here](#).

This guide is in two parts, the first deals with the background to the technical issues raised by [McIntyre and McKittrick \(2005\)](#) (MM05), while the second part discusses the application of this to the original [Mann, Bradley and Hughes \(1998\)](#) (MBH98) reconstruction. The wider climate science context is discussed [here](#), and the relationship to other recent reconstructions (the 'Hockey Team') can be seen [here](#).

NB. All the data that were used in MBH98 are freely available for download at <ftp://holocene.evsc.virginia.edu/pub/sdr/temp/nature/MANNETAL98/> (and in the supplementary data at [Nature](#)), along with thorough descriptions of the algorithms.

Part I: Technical issues:

1) What is principal component analysis (PCA)?

This is a mathematical technique that is used (among other things) to summarize the data found in a large number of noisy records so that the essential aspects can more easily be seen. The most common patterns in the data are captured in a number of 'principal components' which describe some percentage of the variation in the original records. Usually only a limited number of components ('PC's) have any statistical significance, and these can be used instead of the larger data set to give basically the same description.

2) What do these individual components represent?

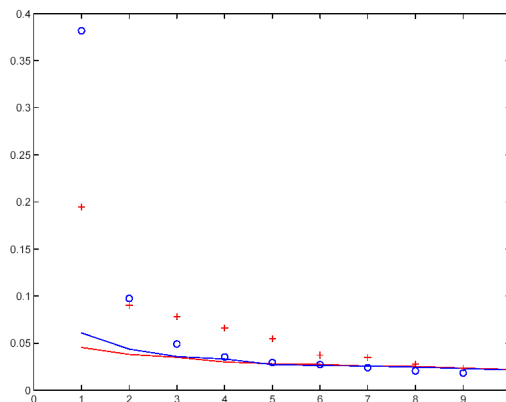
Often the first few components represent something recognisable and physical meaningful (at least in climate data applications). If a large part of the data set has a trend, then the mean trend may show up as one of the most important PCs. Similarly, if there is a seasonal cycle in the data, that will generally be represented by a PC. However, remember

that PCs are just mathematical constructs. By themselves they say nothing about the physics of the situation. Thus, in many circumstances, physically meaningful timeseries are 'distributed' over a number of PCs, each of which individually does not appear to mean much. Different methodologies or conventions can make a big difference in which pattern comes up tops. If the aim of the PCA analysis is to determine the most important pattern, then it is important to know how robust that pattern is to the methodology. However, if the idea is to more simply summarize the larger data set, the individual ordering of the PCs is less important, and it is more crucial to make sure that as many significant PCs are included as possible.

3) How do you know whether a PC has significant information?

This determination is usually based on a 'Monte Carlo' simulation (so-called because of the random nature of the calculations). For instance, if you take 1000 sets of random data (that have the same statistical properties as the data set in question), and you perform the PCA analysis 1000 times, there will be 1000 examples of the first PC. Each of these will explain a different amount of the variation (or variance) in the original data. When ranked in order of explained variance, the tenth one down then defines the 99% confidence level: i.e. if your real PC explains more of the variance than 99% of the random PCs, then you can say that this is significant at the 99% level. This can be done for each PC in turn. (This technique was introduced by Preisendorfer et al. (1981), and is called the Preisendorfer N-rule).

The figure to the right gives two examples of this. Here each PC is plotted against the amount of fractional variance it explains. The blue line is the result from the random data, while the blue dots are the PC results for the real data. It is clear that at least the first two are significantly separated from the random noise line. In the other case, there are 5 (maybe 6) red crosses that appear to be distinguishable from the red line random noise. Note also that the first ('most important') PC does not always explain the same amount of the original data.



4) What do different conventions for PC analysis represent?

Some different conventions exist regarding how the original data should be normalized. For instance, the data can be normalized to have an average of zero over the whole record, or over a selected sub-interval. The variance of the data is associated with departures from the whatever mean was selected. So the pattern of data that shows the biggest departure from the mean will dominate the calculated PCs. If there is an *a priori* reason to be interested in departures from a particular mean, then this is a way to make sure that those patterns move up in the PC ordering. Changing conventions means that the explained variance of each PC can be different, the ordering can be different, and the number of significant PCs can be different.

5) How can you tell whether you have included enough PCs?

This is rather easy to tell. If your answer depends on the number of PCs included, then you haven't included enough. Put another way, if the answer you get is the same as if you had used all the data without doing any PC analysis at all, then you are probably ok. However, the reason why the PC summaries are used in the first place in paleo-reconstructions is that using the full proxy set often runs into the danger of 'overfitting' during the calibration period (the time period when the proxy data are trained to match the instrumental record). This can lead to a decrease in predictive skill outside of that window, which is the actual target of the reconstruction. So in summary, PC selection is a trade off: on one hand, the goal is to capture as much variability of the data as represented by the different PCs as possible (particularly if the explained variance is small), while on the other hand, you don't want to include PCs that are not really contributing any more significant information.

Part II: Application to the MBH98 'Hockey Stick'

1) Where is PCA used in the MBH methodology?

When incorporating many tree ring networks into the multi-proxy framework, it is easier to use a few leading PCs rather than 70 or so individual tree ring chronologies from a particular region. The trees are often very closely located and so it makes sense to summarize the general information they all contain in relation to the large-scale patterns of variability. The relevant signal for the climate reconstruction is the signal that the trees have in common, not each individual series. In MBH98, the North American tree ring series

were treated like this. There are a number of other places in the overall methodology where some form of PCA was used, but they are not relevant to this particular controversy.

2) What is the point of contention in MM05?

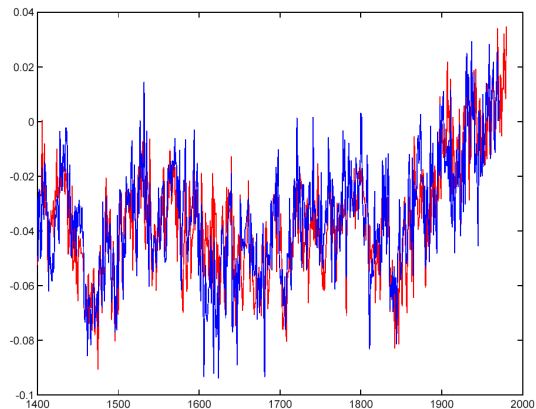
MM05 contend that the particular PC convention used in MBH98 in dealing with the N. American tree rings selects for the 'hockey stick' shape and that the final reconstruction result is simply an artifact of this convention.

3) What convention was used in MBH98?

MBH98 were particularly interested in whether the tree ring data showed significant differences from the 20th century calibration period, and therefore normalized the data so that the mean over this period was zero. As discussed above, this will emphasize records that have the biggest differences from that period (either positive or negative). Since the underlying data have a 'hockey stick'-like shape, it is therefore not surprising that the most important PC found using this convention resembles the 'hockey stick'. There are actually two significant PCs found using this convention, and both were incorporated into the full reconstruction.

4) Does using a different convention change the answer?

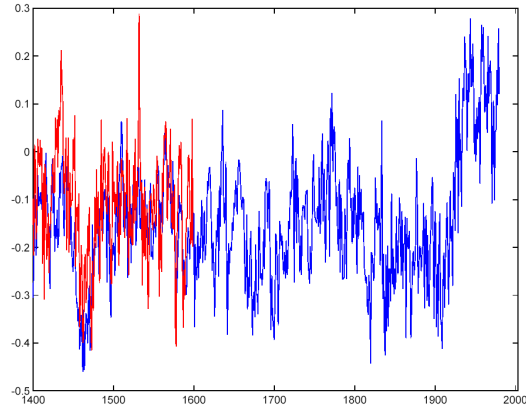
As discussed above, a different convention (MM05 suggest one that has zero mean over the whole record) will change the ordering, significance and number of important PCs. In this case, the number of significant PCs increases to 5 (maybe 6) from 2 originally. This is the difference between the blue points (MBH98 convention) and the red crosses (MM05 convention) in the first figure. Also PC1 in the MBH98 convention moves down to PC4 in the MM05 convention. This is illustrated in the figure on the right, the red curve is the original PC1 and the blue curve is MM05 PC4 (adjusted to have same variance and mean).



But as we stated above, the underlying data has a hockey stick structure, and so in either case the 'hockey stick'-like PC explains a significant part of the variance. Therefore, using

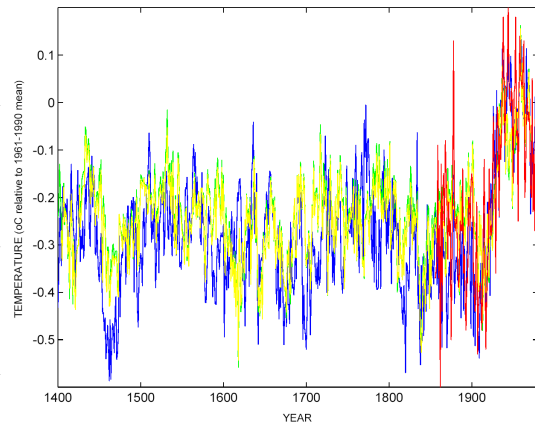
the MM05 convention, more PCs need to be included to capture the significant information contained in the tree ring network.

This figure shows the difference in the final result whether you use the original convention and 2 PCs (blue) and the MM05 convention with 5 PCs (red). The MM05-based reconstruction is slightly less skillful when judged over the 19th century validation period but is otherwise very similar. In fact any calibration convention will lead to approximately the same answer as long as the PC decomposition is done properly and one determines how many PCs are needed to retain the primary information in the original data.



5) What happens if you just use all the data and skip the whole PCA step?

This is a key point. If the PCs being used were inadequate in characterizing the underlying data, then the answer you get using all of the data will be significantly different. If, on the other hand, enough PCs were used, the answer should be essentially unchanged. This is shown in this figure. The reconstruction using all the data is in yellow (the green line is the same thing but with the 'St-Anne River' tree ring chronology taken out). The blue line is the original reconstruction, and as you can see the correspondence between them is high.

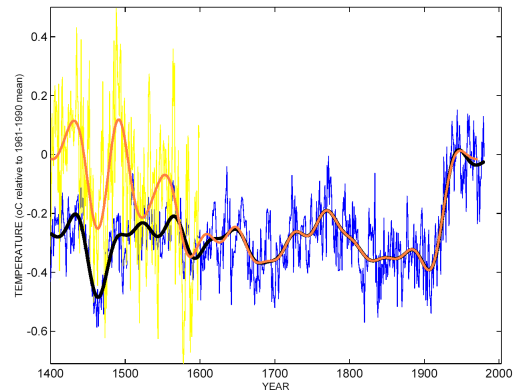


The validation is slightly worse, illustrating the trade-off mentioned above i.e. when using all of the data, over-fitting during the calibration period (due to the increase number of degrees of freedom) leads to a slight loss of predictability in the validation step.

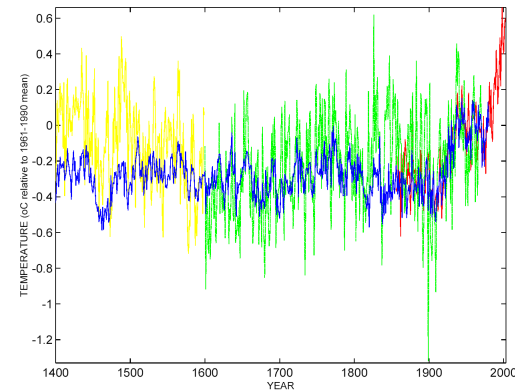
6) So how do MM05 conclude that this small detail changes the answer?

MM05 claim that the reconstruction using only the first 2 PCs with their convention is significantly different to MBH98. Since PC 3,4 and 5 (at least) are also significant they are leaving out good data. It is mathematically wrong to retain the same number of PCs if the convention of standardization is changed.

In this case, it causes a loss of information that is very easily demonstrated. Firstly, by showing that any such results do not resemble the results from using all data, and by checking the validation of the reconstruction for the 19th century. The MM version of the reconstruction can be matched by simply removing the N. American tree ring data along with the 'St Anne River' Northern treeline series from the reconstruction (shown in yellow). Compare this curve with the ones shown above.



As you might expect, throwing out data also worsens the validation statistics, as can be seen by eye when comparing the reconstructions over the 19th century validation interval. Compare the green line in the figure below to the instrumental data in red. To their credit, MM05 acknowledge that their alternate 15th century reconstruction has no skill.



7) Basically then the MM05 criticism is simply about whether selected N. American tree rings should have been included, not that there was a mathematical flaw?

Yes. Their argument since the beginning has essentially *not* been about methodological issues at all, but about 'source data' issues. Particular concerns with the "bristlecone pine"

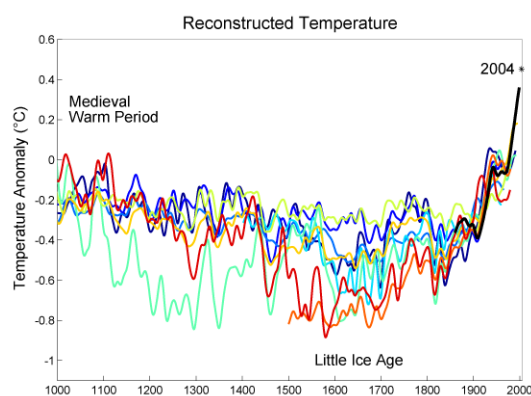
data were addressed in the followup paper MBH99 but the fact remains that including these data improves the statistical validation over the 19th Century period and they therefore should be included.

8) So does this all matter?

No. If you use the MM05 convention and include all the significant PCs, you get the same answer. If you don't use any PCA at all, you get the same answer. If you use a completely different methodology (i.e. Rutherford et al, 2005), you get basically the same answer. Only if you remove significant portions of the data do you get a different (and worse) answer.

9) Was MBH98 the final word on the climate of last millennium?

Not at all. There has been significant progress on many aspects of climate reconstructions since MBH98. Firstly, there are more and better quality proxy data available. There are new methodologies such as described in [Rutherford et al \(2005\)](#) or [Moberg et al \(2005\)](#) that address recognised problems with incomplete data series and the challenge of incorporating lower resolution data into the mix. Progress is likely to continue on all these fronts. As of now, all of the 'Hockey Team' reconstructions (shown left) agree that the late 20th century is anomalous in the context of last millennium, and possibly the last two millennia.



Used under GFDL license from [Wikipedia](#).