HOCKEY STICKS, PRINCIPAL COMPONENTS AND SPURIOUS SIGNIFICANCE

Stephen McIntyre
512-120 Adelaide St. West,
Toronto, Ontario Canada M5H 1T1
smcintyre25@yahoo.ca

Ross McKitrick
Department of Economics,
University of Guelph,
Guelph Ontario Canada N1G 2W1.

Accepted Abstract: American Geophysical Union Meetings, December 2004, San Francisco.

Multi-proxy studies have been the primary means of transmitting paleoclimatic findings to public policy. For policy use, such studies should be replicable in the sense of King (1995). The best-known and most widely applied multi-proxy study is Mann, Bradley and Hughes (1998) ("MBH98") and its 1999 extension, which claimed to have exceptional "robustness" and "skill".

We attempted to replicate MBH98 results and found, among other problems, that MBH98 methodology included two important unreported steps:
(1) Subtraction of the 1902-1980 mean prior to principal components (PC) calculations (rather than, say, the 1400-1980 mean in the AD1400 step);
(2) Extrapolation of a duplicate version of the Gaspé tree ring series.

We show that, due to high early 15th century values, their results are not robust for the following cases:

a) Presence or absence of the extrapolation of 4 years at the beginning of the Gaspé tree ring series;
b) subtraction of the 1400-1980 mean rather than subtraction of the 1902-1980 mean, while using the same number of retained PC series in each step as MBH98;
c) the presence or absence of the North American PC4, while subtracting the 1400-1980 mean and using 5 PCs in the AD1400 step;
d) presence or absence of a small subset of high-altitude tree ring sites, mostly "strip bark" bristlecone pines, mostly collected by one researcher, Donald Graybill.

The subtraction of the 1902-1980 mean dramatically inflates the role of the bristlecone pine sites, which then impart a distinctive hockey stick shape to the MBH98 PC1 and then to the NH temperature reconstruction.

MBH98 claimed "skill" through apparently significant Reduction of Error (RE) statistics, reporting 0.51 in the AD1400 step, as compared to a reported 99 percent significance level of 0, which they calculated through simulations using red noise with low AR1 coefficients (0.2). We benchmarked a more realistic significance level by applying MBH98 PC methods to 10,000 sets of 70 red noise series modeled through fractional difference models to have the same red noise persistence as the critical North American AD1400 tree ring network. These calculations regularly resulted in "hockey-stick" shaped PC1s with sharp inflections at the start of the 20th century. We then modeled the resulting 10,000 PC1s against NH temperature and found that the 99 percent RE significance level was 0.59. By this benchmark, the reported RE statistic (0.51) in MBH98 for the AD1400 step lacks statistical significance.

Most dendroclimatic reconstructions also provide statistics other than an RE statistic, including R2 and Coefficient of Efficiency, but MBH98 does not and the authors have refused to provide supporting data from which the statistics can be calculated. In our emulations of their calculations, we have been unable to replicate anything close to the reported RE results other than through re-tuning, a procedure not described in MBH98. With a re-tuning step, for the critical AD1400 step, we have obtained an RE of 0.46, but with an R2 of only 0.02 and a CE of minus 0.26, all of which lack statistical significance.

This case study illustrates for extreme caution in basing public policy on articles, such as MBH98, whose claims cannot be verified.