

Comment on “Hockey sticks, principal components, and spurious significance” by S. McIntyre and R. McKittrick

P. Huybers

Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

Received 3 May 2005; revised 9 August 2005; accepted 23 August 2005; published 21 October 2005.

Citation: Huybers, P. (2005), Comment on “Hockey sticks, principal components, and spurious significance” by S. McIntyre and R. McKittrick, *Geophys. Res. Lett.*, 32, L20705, doi:10.1029/2005GL023395.

[1] *McIntyre and McKittrick* [2005] (hereinafter referred to as MM05) point out a bias in the *Mann et al.* [1998] (hereinafter referred to as MBH98) Northern Hemisphere temperature reconstruction, one tending to enhance trends during the last century. Having reproduced the statistical results of MM05, this comment is prompted by further questions regarding appropriate implementation of principal component analysis (PCA) and the presence of discrepancies in their estimate of significance levels.

[2] MBH98 use principal component analysis (PCA) to distill the large number of tree ring records (90% of the total 415 proxy records) into a smaller number of principal components (PCs). MM05 focus on a subset of the data, the seventy North American tree ring records (NOAMER) extending back to AD1400, and show that the MBH98 normalization leads to biases in the leading principal component (PC1). It is in this same step that MM05 use a questionable normalization procedure, making it useful to describe the various normalization conventions in detail.

[3] The MBH98 normalization convention for a record, x , is $x_{\text{MBH}} = (x - \bar{x}_{1902})/\sigma'_{1902}$, where \bar{x}_{1902} and σ_{1902} are the mean and standard deviation computed between 1902 and 1980. MBH98 compute the standard deviation after detrending x , indicated as σ' , an additional step that seems questionable but turns out not to influence the results. Because proxy records span different intervals, it is impossible to both normalize records over a fixed interval and ensure that records are zero-mean over their entire duration. MBH98 presumably chose the 1902 to 1980 normalization period because almost all records span this interval, but which MM05 point out leads to a bias in the results.

[4] The reason for the bias in the MBH98 PC1 can be understood by considering that PCA maximizes the variance described by each principal component where variance is measured as the sum of the squared record, $\sigma^2 = \sum x_i^2$, and x is not necessarily zero-mean. The MBH98 normalization tends to assign large variances to records with a pre-1902 mean far from the 1902 to 1980 mean, and records with the largest variance tend to determine PC1. This bias was checked using a Monte Carlo algorithm independent of MM05's. To remove the bias in MBH98's calculations, MM05 set records to zero mean over the entire 1400 to 1980 interval, giving the normalization $x_{\text{MM}} = x - \bar{x}_{1400}$.

[5] MM05 list fifteen records as dominating the MBH98 PC1 (see MM05, Table 1). The MBH98 normalization leads to these fifteen records having roughly twice the variance of the other records, whereas the MM05 normalization effectively down-weights these same records by a factor of two (see Figure 1). What, then, is the best normalization?

[6] NOAMER records are standardized chronologies [*Cook and Kairiukstis*, 1990], reported as fractional changes from mean tree ring width or maximum ring density after correcting for the effects of increasing tree age. The variance of the chronology is a function of both environmental variability and the trees' sensitivity to the environment. Sensitivity depends on factors such as species, soil, local topography, tree age, location within a forest, and what quantity is being measured [*Fritts*, 1976]. The most striking example of varying sensitivity is that the two NOAMER chronologies indicating changes in tree ring density (co509x and wy023x) have variances roughly thirty times smaller than the other chronologies indicating changes in tree ring width.

[7] To further check the controls on tree ring variance, the variance of each NOAMER chronology is compared with that of the nearest instrumental temperature record using the *Jones and Moberg* [2003] instrumental compilation between 1870 and 1980. Because no meaningful relationship is discernible (there is actually a weak anti-correlation between the tree ring chronology and instrumental variances), the best approach appears to be to normalize the variance of the NOAMER records prior to performing PCA. Thus, a third normalization is proposed where records are adjusted to zero-mean and unit variance over their full 1400 to 1980 duration, a standard practice in PCA [*Preisendorfer*, 1988, p. 22; *Rencher*, 2002, p. 393] here referred to as “full normalization”. Up to multiplication by a constant, full normalization is equivalent with PCA of the correlation matrix. Another point raised by MM05 is that many of the strongest trends in the tree ring chronologies may be unrelated to temperature change [*Graybill and Idso*, 1993] — in future studies this may warrant the exclusion or down-weighting of certain records, but this is an additional step which would have to be explicitly stated.

[8] Figure 2 shows the leading principal component (PC1) after normalizing the NOAMER chronologies according to the MBH98, MM05, and fully normalized conventions. To measure the degree of anomalous behavior in recent temperatures, MM05 define a hockey-stick-index as the 1902 to 1980 mean minus the 1400 to 1980 mean, all divided by the 1400 to 1980 standard deviation. As might be expected, the MBH98 PC1 has the largest hockey-stick-index at 1.6, MM05 the smallest at 0.3, and full normalization an intermediate index of 0.8. The

amount of variance explained by PC1 also varies according to the normalization convention: 38%, 19%, and 17% for the MBH98, MM05, and full normalization respectively. Monte Carlo estimates indicate that these variances are each well above the 99% confidence level for auto-correlated noise. Note that none of the higher order principal components explain more than 10% of the proxy variance, indicating that these are relatively weak components of the proxy variability.

[9] The sensitivity of PCA to what normalization is applied is made clear by the low squared-cross-correlation between the MBH98 and MM05 PC1s ($r^2 = 0.17$). In contrast, the annual average of the records is nearly insensitive to which of the three normalizations are applied; each pair of record averages has a squared-cross-correlation exceeding 0.95. To avoid ambiguity in future studies, it may be preferable to use simple averages rather than PCA when estimating spatial means such as Northern Hemisphere temperatures.

[10] It is useful to compare the record averages with the PC1 results after scaling both to Northern Hemisphere instrumental temperatures [Jones and Moberg, 2003] (see Figure 2). The pre-1902 values of the MBH98 PC1 are more negative than the corresponding record average. Conversely, the pre-1902 values of the MM05 PC1 are less negative, an observation somewhat at odds with the statement in MM05 that their PC1 is “very similar to the unweighted mean of all the series”. These off-sets between PCs and record averages further indicate that the MM05 results are biased in the opposite direction to those of the MBH98 results. The fully normalized PC1 and average closely resemble one another ($r^2 = 0.95$), indicating that the fully normalized PC1 describes variability common to much of the NOAMER data-set.

[11] A second issue involves the MM05 estimate of significance levels for the reduction of error statistic, $RE = 1 - \sum (y - x)^2 / \sum y^2$, using Monte Carlo methods. In this case, y is instrumental Northern Hemisphere temperatures and x is the PC1 of random, proxy-like records. An approximate distribution for the null-hypothesis of no relationship between x and y is obtained by binning many random realization of RE. Records whose actual RE value

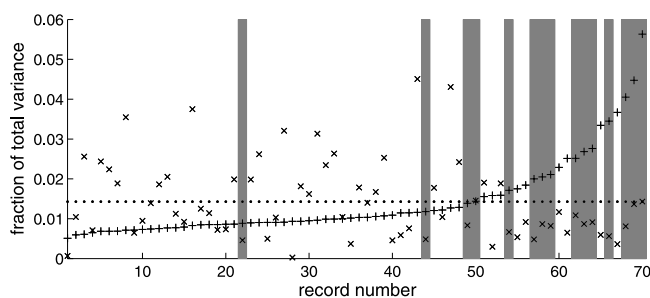


Figure 1. North American tree ring variance expressed as a fraction of total variance after applying three separate normalizations: MBH98 (plus signs), MM05 (crosses), and full normalization (dots). Records are sorted according to MBH98 variance. Shaded records are identified by MM05 (their Table 1) as dominating the MBH98 leading principal component.

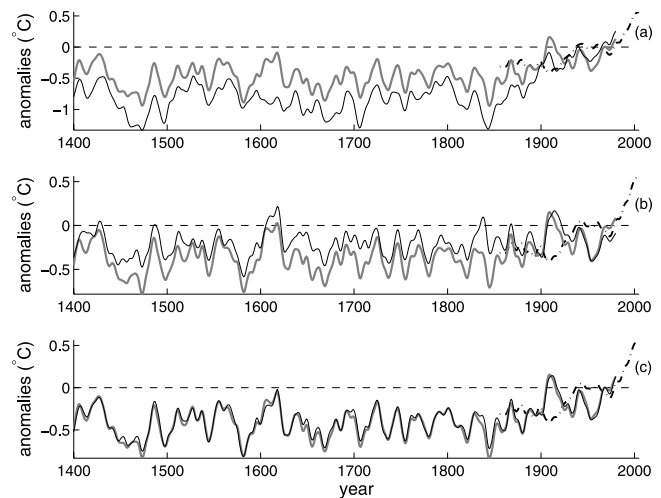


Figure 2. Leading principal components (black, thin line) and averages (gray, thick line) of the North American tree ring chronologies calculated using three separate normalizations: (a) the MBH98 normalization, (b) the MM05 normalization, and (c) full normalization. Shown for reference are the instrumental Northern Hemisphere temperature anomalies relative to the 1961 to 1990 average [Jones and Moberg, 2003] (dash-dot lines). Averages and PCA have their mean and variance scaled to the instrumental record between 1902 and 1980. Averages are similar to one another, but only the fully normalized PC1 is similar to the averages. For visual clarity, records are shown after smoothing by an 11 year Hanning window.

exceeds 99% of the randomly realized values are said to be significant. Inspection of the MM05 Monte Carlo code (provided as auxiliary material¹) shows that realizations of x are not adjusted to the variance of the instrumental record during the 1902 to 1980 training interval — a critical step in the procedure.

[12] The MM05 code generated realizations of x having roughly a fourth the variance of y , biasing RE realizations toward being too large. MM05 thus estimate a RE critical value substantially higher (RE = 0.6) than that of MBH98 (RE = 0.0) and incorrectly conclude that the AD1400 step of the MBH98 temperature reconstruction is insignificant. When the MM05 algorithm is corrected to include the variance adjustment step and re-run, the estimated RE critical value comes into agreement with the MBH98 estimate. (Data and computer codes used for PCA analysis and the estimation of critical values are provided as auxiliary material.)

[13] In summary, MM05 show that the normalization employed by MBH98 tends to bias results toward having a hockey-stick-like shape, but the scope of this bias is exaggerated by the choice of normalization and errors in the RE critical value estimate. Those biases truly present in the MBH98 temperature estimate remain important issues, and corrections for these biases will be taken up elsewhere.

¹Auxiliary material is available at <ftp://ftp.agu.org/apend/gl/2005GL023395>.

[14] **Acknowledgments.** Useful comments were provided by W. Curry, O. Marchal, M. Raymo, and C. Wunsch. Support was provided by the NOAA Postdoctoral Program in Climate and Global Change.

References

- Cook, E. R., and L. A. Kairiukstis (1990), *Methods of Dendrochronology*, Springer, New York.
- Fritts, H. C. (1976), *Tree Rings and Climate*, Elsevier, New York.
- Graybill, D., and S. Idso (1993), Detecting the aerial fertilization effect of atmospheric CO₂ enrichment in tree-ring chronologies, *Global Biogeochemical Cycles*, 7, 81–95.
- Jones, P., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Clim.*, 16, 206–223.
- Mann, M., R. Bradley, and M. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787.
- McIntyre, S., and R. McKittrick (2005), Hockey sticks, principal components, and spurious significance, *Geophys. Res. Lett.*, 32, L03710, doi:10.1029/2004GL021750.
- Preisendorfer, R. (1988), *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York.
- Rencher, A. C. (2002), *Methods of Multivariate Analysis*, John Wiley, Hoboken, N. J.
-
- P. Huybers, Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. (phuybers@whoi.edu)