REPLY TO AMMANN AND WAHL:
"COMMENT ON "HOCKEY STICKS, PRINCIPAL COMPONENTS AND SPURIOUS SIGNIFICANCE"

Stephen McIntyre
512-120 Adelaide St. West,
Toronto, Ontario Canada M5H 1T1
stephen.mcintyre@utoronto.ca

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

Ammann and Wahl (AW herein) build their entire comment on two methodological arguments (principal components using the correlation matrix and rescaling impact on RE benchmarking), which repeat almost verbatim two arguments previously made by Huybers [2005], neither adding new justification, nor discussing, much less rebutting, anything in our Reply [McIntyre and McKitrick 2005c; "MM05c" herein]. Nor do they address the subsequent discussion in Bürger and Cubasch [2005]. Additionally, AW not only repeat results that we had previously published, but claim them as their own and then accuse us of having failed to report them. In their abstract and summary, AW make claims that are unsupported in their text, then assert our results are "unfounded," despite the fact that results from their own code yields validation statistics (unreported by AW) that strikingly confirm claims in McIntyre and McKitrick [2005a] (MM05a) concerning spurious significance in the Mann et al. [1998] (MBH98) reconstruction.

Correlation or Covariance PCs?

Huybers argued that tree ring chronologies should be divided by their standard deviation prior to PC calculations (his "full normalization"), acknowledging that this option was equivalent to PC analysis using the correlation matrix, rather than the covariance matrix. In McIntyre and McKitrick [2005b], we had previously reported on this option, noting both that the bristlecone hockey stick occurred in the PC2 (rather than the PC4) and that the resulting NH reconstruction was intermediate between MBH98 and that from using two covariance PCs. AW raise the identical point as Huybers. We already provided comprehensive counterarguments to Huybers' position [MM05c], and there is no need to repeat them here.. We observed in [MM05c] that "Any valid climate reconstruction should not depend on whether a correlation matrix or covariance matrix is used in tree ring PC analysis" And notwithstanding this, for networks expressed in common units, such as ITRDB site chronologies, PC analysis using the covariance matrix was the recommended procedure in statistical texts, even in Huybers' own cited references.

Bürger and Cubasch have emphasized the need for robust methods, observing that "if it [MBH98] is robust, certain refinements such as rescaling should *not* affect the essence of the final result." PC centering is identified by them as one such "refinement". Although they allow for a choice between centered and uncentered PCs, they do not expressly mention the further choice between covariance and correlation PCs, although it is obviously consistent with their taxonomy.

AW simply re-iterate division of chronologies by their standard deviation, as originally advocated by Huybers, but do not provide any statistical references or new justification for the repeated claim and do not rebut either our Reply to Huybers or the further points in Bürger and Cubasch.

Cross-Validation Statistics

In MM05a, we asserted that the MBH98 reconstruction lacked statistical significance, as evidenced by a failed r^2 and other cross-validation statistics. We observed that statistical significance for the RE statistic cannot be determined from a theoretical distribution, but only from simulations, and pointed out that the MBH98 Monte Carlo simulations for RE significance had been incorrectly done, yielding a 99% RE significance benchmark which was much too low (0.0 as opposed to 0.56 from our simulations).

Huybers did not dispute the r^2 finding but argued that our simulations failed to replicate an MBH98 re-scaling step (not reported in MBH98 itself); he did new simulations and claimed that a 0.0 RE benchmark had been restored. We mention in passing that this is the same rescaling step that Bürger and Cubasch say should "not" affect the essence of the final result. In our Reply to Huybers, we showed that Huybers' new simulations had failed to replicate the use of proxy networks in MBH98; we reported new simulations using networks of noise and the rescaling step proposed by Huybers, once again yielding a high RE 99% benchmark (0.54).

AW, again, repeat the exact same argument and then ignore the network issue. In their own simulations they do not employ pseudoproxy networks. As a result, their endorsement of Huybers' claim to have restored a benchmark of 0.0 is worthless.

Further, like Huybers, AW also failed to resolve (or even consider) the evidence of the failed MBH98 r^2 and other cross-validation statistics. Table 1 below compares three sets of cross-validation statistics for the 15th Century step under dispute: the values reported in MBH98, those reported in MM05a and those that we obtained from running archived AW code with MBH98 temperature PCs and weights. All tabulated values are insignificant. The near identity of cross-validation r^2 (and other) statistics under AW code to ours is independent verification of key claims we published in MM05a, and it refutes AW allegations that our claims were "unfounded" It is noteworthy that they did not report these results, the omission of which leads to an inaccurate representation of the research record, especially when accompanied by AW allegations that our results as "unfounded".

Table 1. MBH98 (and Emulations) 15th Century Step Cross-Validation Statistics

	RE	r^2	CE	Sign Test	Product Test
MBH98	0.48	n.r.	n.r.	n.r.	n.r.
MM05a Emulation	0.46	0.02	-0.26	0.46	1.54
A&W Code	0.47	0.02	-0.24	0.54	0.91

Misrepresentations and Unsupported Points

AW derive rhetorical force by a series of misrepresentations and false claims about our work, as well as from a series of unsupported assertions, especially in their summary and abstract. Space does not permit a complete listing so we discuss only the most egregious here.

1. AW falsely asserted that a hockey stick shape appears in "***all*** the summaries… of the ITRDB [North American] network" (emphasis added). As we pointed out in MM05b and in our Reply to Huybers, there is no hockey-stick pattern in any North American PCs in the summary in the BACKTO_1400-CENSORED directory at Mann's FTP site – a summary excluding a small group of bristlecone

sites. The hockey stick pattern is characteristic of bristlecone growth and does not appear in summaries in which bristlecones are not present.

2. AW accused us of not reporting the presence of a hockey stick pattern in lower order PCs. In MM05a para. 12 (MM05b – p. 75), we expressly reported that the bristlecone hockey stick appeared in the covariance **PC4**, further pointing out (unlike AW) that it only accounted for 8% of total variance. In MM05b (p. 76), we expressly reported that the bristlecone hockey stick appeared in the **PC2** using correlation PCs. Remarkably, AW presented the same observations as novel (see their Figures 1b and 1c) and reproached us over it.

3. AW falsely accused us of retaining only two PCs in our analyses using covariance PCs, ignoring a specific discussion in MM05b (p. 76) of the impact of retaining up to **5** covariance PCs.

4. AW say we failed to consider the effect of "proper" standardization on temperature reconstructions, while completely ignoring results presented in MM05b (p. 76) using correlation PCs – an equivalent methodology. AW then claimed - without **any** supporting calculations or discussion of our different findings - that the differences between MBH98 and a reconstruction using 2 correlation PCs are less than "five hundreds of a degree". However, for Case 5b of Wahl and Ammann (under review), which also uses 2 correlation PCs, the reported RE statistic (0.18) is 0.30 lower than the RE statistic for the corresponding MBH98 emulation (0.48) – a result which is inconsistent with the claim of a negligible difference.

5. AW assert, without any statistical support, that a supposed similarity between the PC1+PC2 (both presumably standardized, although the methodology is not discussed) under the MBH98 method, to the sum of the PC1+PC2 using correlation PCs, has some statistical meaning. Such a construct plays no role in the MBH98 method, since the PC series are not summed, but are used individually. If AW wish to promote the PC1+PC2 as a temperature proxy, then they should demonstrate the validity of the proxy. Given that AW Figure 1d shows that the PC1+PC2 does not have a hockey stick shape, it is likely that an MBH98-type reconstruction using the PC1+PC2, together with other MBH98 proxies, would yield results rather similar to results using 2 covariance PCs.

6. AW acknowledge that MM05a simulation procedures yielded representative autocorrelation structures for the North American tree ring data, but claimed, without **any** evidence or even an alternative version of MM05a Figure 1, that the autocorrelation of the PCs from the red noise network are unlike those of the original network. AW claimed that our simulations had inadequately modeled differing site chronology variances. But the MBH98 "PC" procedure includes the division of each chronology by its (detrended) calibration period standard deviation. We implemented this step in the MM05a simulations of MBH98 methodology. *Prima facie*, dividing each chronology by this standard deviation appears to almost entirely cancel out the tailoring of variances that AW argue for in the prior step. In the absence of a more coherent presentation, it is impossible to find any substance in their argument.

7. In their summary, AW baldly assert that NH reconstructions based on "no standardization" yield results similar to MBH98. We are unable to comprehend this claim, given the conflicting claims elsewhere on the supposed impact of "proper standardization".
8. In their summary, AW state that a NH reconstruction without PC summarization yields results similar to MBH98. No supporting evidence is presented anywhere in the text, and we doubt that any such reconstruction, without an alternative scheme for regional summarization would replicate the "reasonable geographical balance" central to MBH98 methodology, or be anything other than a backdoor insertion of bristlecones.
9. In their summary, AW claim that "all approaches that capture an acceptable amount of the variance" lead to essentially the same results. No evidence is provided for this claim in their text, which does not even discuss, much less define, "acceptable" variance, other than in the circular sense of "capturing" the bristlecone hockey stick. As discussed in MM05b, reconstructions without bristlecones or with downweighted bristlecones yield quite different results.
10. AW falsely claimed that we said that the hockey stick shape was **"introduced"** through erroneous MBH98 methodology, and that we claimed *simpliciter* that a "spurious hockey stick climate reconstruction is introduced by data transformation". This claim is absurd on its face, since we reported the bristlecone hockey stick in lower order PCs under centered methods. We asserted that the erroneous MBH98 method promoted a lower-order effect into the PC1, while inflating the first eigenvalue such that Mann et al. misconstrued the bristlecones as representing the "dominant component of variance". The existence of this MBH98 bias is now widely recognized. While we identified the bias, however, we emphasized the interaction between the flawed methodology and flawed proxies, expressly pointing out that patching the flawed methodology would not necessarily yield a valid reconstruction if flawed proxies dominated the reconstruction though other fixes.

In summary, AW have not raised any issues except ones already dealt with in our exchange with Huybers, and have failed to show that **any** claims in MM05a were "unfounded". Indeed, calculations using AW code confirm MM05a findings of failed MBH98 cross-validation r^2 and other statistics.

References:

Ammann, C. and E. Wahl (2006), Comment on ''Hockey sticks, principal components, and spurious significance'' by S. McIntyre and R. McKitrick, *GRL*, doi:10.1029/2005GL252464, in press.

Bürger, G., and U. Cubasch (2005), Are multiproxy climate reconstructions robust?, *GRL*, 32, L23711, doi:10.1029/2005GL024155.

Huybers, P. (2005), Comment on "Hockey sticks, principal components and spurious significance" by McIntyre and McKitrick, *GRL*, L20705, doi: 10.1029/2005 GL023395.

Mann, M.E., R.S. Bradley and M.K. Hughes (1998), Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779-787.

McIntyre, S. and R. McKitrick, (2005a), Hockey Sticks, Principal Components and Spurious Significance, *GRL*, 32, L03710, doi:10.1029/2004GL021750.

McIntyre, S. and R. McKitrick (2005b), The M&M Critique of the MBH98 Northern Hemisphere Climate Index: Update and Implications, Energy and Environment, 16, 69-99.

McIntyre, S., and R. McKitrick (2005c), Reply to comment by Huybers, *GRL*, 32, L20713, doi:10.1029/2005GL023586.