

Huybers' and von Storch-Zorita Comments, and Our Replies

The Comments and Replies are available at <http://www.climateaudit.org/?p=413>.

Steve has posted extended discussions on the two comments, which are also linked from the above URL. This note is just a quick summary of the counterpoints.

1. Huybers

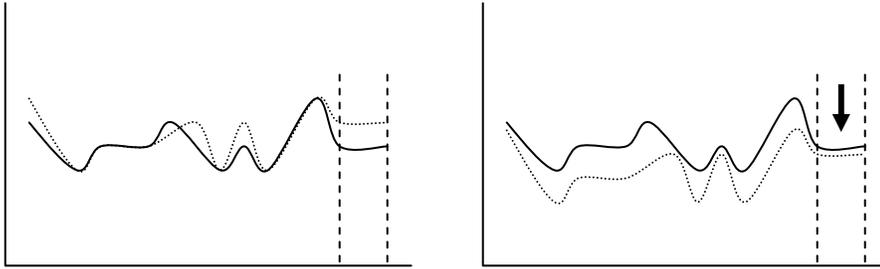
Huybers makes two arguments.

1. Mann's PC method is biased toward finding hockey sticks in a data set. We (M&M) illustrated the bias by comparing a principal component (PC) produced by Mann's algorithm to that produced by a standard algorithm, which does an algebraic decomposition of the covariance matrix. However the underlying data should be transformed to have a standardized variance prior to taking a PC, which implies we should have used a PC based on the decomposition of the *correlation* matrix rather than the *covariance* matrix. The correlation PC—as computed by Huybers—looks much more like a hockey stick, indicating that the M&M comparison exaggerates the bias from the MBH method.
2. To compute a significance benchmark for the RE statistic, M&M used pseudo-“predicted values” generated by feeding red noise into the MBH98 algorithm. This is the correct procedure, however the MBH98 algorithm generates PCs with a much smaller variance than the temperature data it is being used to predict. So there ought to be a step in which the variance of the pseudo-predicted values is scaled up to match that of the temperature data. Although inserting this step does not change the R^2 statistic (which indicates the early portion of MBH98 is insignificant either way), it moves the significance benchmark for the RE statistic to near 0.0, suggesting that the early portion of MBH98 actually may be significant.

Our reply, in brief, is as follows.

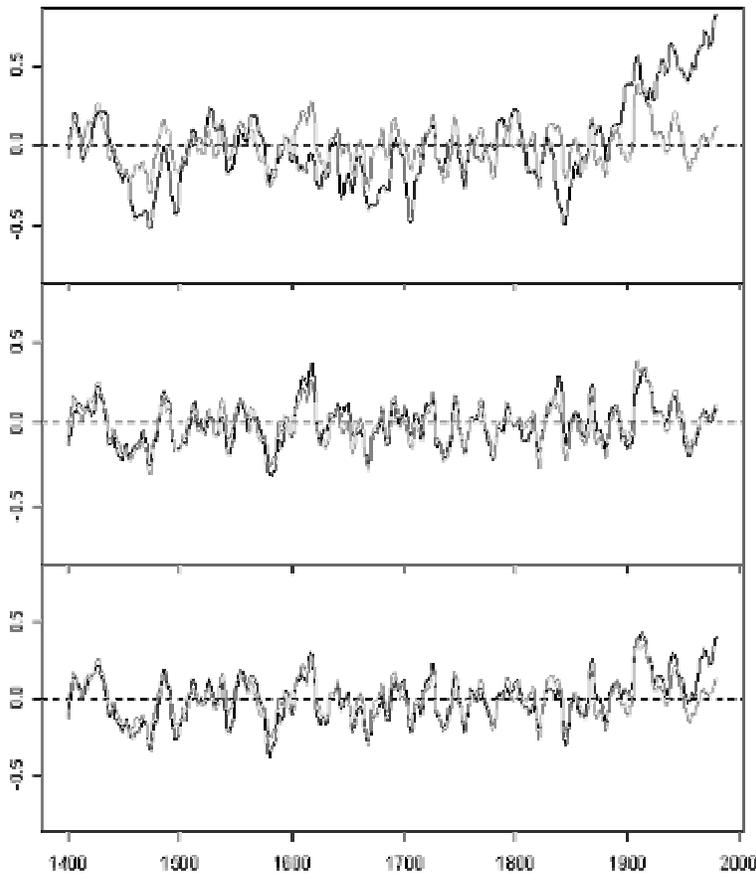
#1: In PC analysis, while it is true that the decomposition can be applied to the covariance or the correlation matrix, textbooks (including those cited by Huybers) indicate a clear preference for using the covariance matrix, unless the data are in widely differing units, which is not the case here. The tree ring data are pre-standardized into dimensionless index numbers, and are just the sort of data for which the covariance matrix is the standard, preferred option. So if they are pre-standardized, why would further standardization change anything? Actually it doesn't. The two kinds of PC are nearly identical: the appearance of difference arises because of a graphing trick in Huybers' Comment.

Here's a schematic version.



Suppose that the pair of lines on the left (solid, dotted) represent two series you want to compare. They are both centered on the same mean. They clearly track each other closely up to the final segment in between the dashed vertical lines. You wouldn't think of them as radically different. But now suppose you decide to force the two lines to have the same average over the final segment, as in the right version. The dashed line has to shift down, opening up long series of offsets between them. The same series now look much less alike. But of course that's just due to the recentering.

This is what Huybers did in his graph. The two PC1s discussed in Huybers' comment are also almost identical, except at the mid-20th century where they diverge, based almost entirely on the differing weights assigned to the bristlecones. When drawing his graph, he renormalizes the data over the 20th century portion, the one place they diverge. (We are uncertain how this rescaling can be reconciled to his recommendation for "full normalization" elsewhere.) The effect on the graph is as in the above figure on the right: it introduces offsets over the 1400-1900 interval, which he then considers a knock against our PC. But here's what the graphs would look like without the Huybers trick.



In the above Figure, the top panel shows the PC1 of the 70 full-length proxies as computed by Mann (black line) compared to the simple mean of the 70 full-length proxies (gray line). The gray line is repeated in the next two panels. The second panel shows the PC1 from a standard (covariance) PC calculation and the third is the PC1 from Huybers' correlation-based calculation. Obviously they are almost identical up to the post-1950 segment, and both contradict the 20th century growth pattern in Mann's result. Huybers forces the post-1900 segment to be aligned with the simple mean, which makes our graph look displaced (compared to the mean) across the previous 500 year interval.

The divergence between the second and third panels post-1950 is entirely due to the greater weight placed on the problematic bristlecones in Huybers' correlation PC. He seems to realize the downside of this outcome, since he acknowledges a potential need to downweight or eliminate the bristlecones based on "future" research. This is an unacceptable evasion, since there is more than enough research available to indicate that the bristlecones are invalid as temperature proxies and should be removed from the data set, in which case the covariance/correlation question is totally moot.

Finally, Huybers' presents several other arguments for preferring the correlation PC1. Not that it makes any difference (once the graph is done without tweaking) but, even on his own terms his arguments don't work. His cited texts don't back up his position (we looked them up). He tries to argue that whichever PC1 provides a closer approximation to the mean is more correct, but if you use the mean of all the proxy series (not just the full-length subset) it looks more like the covariance PC1; and in any case that's not a sensible test of a PC method. There are other issues related to the need to treat autocorrelation, duplicating ring width and density series, etc., all of which are explained in the Reply.

Point #2 is sort of correct. We did not do a variance rescaling since it is not mentioned in MBH98, and to make sure you're computing the right benchmark you need to use the exact algorithm you're comparing results from. The recent code released by Mann shows that he actually did do a variance rescaling, without saying so in the paper. It does not occur at the point in the algorithm conjectured by Huybers. However, we accept the need to re-do our RE benchmark. We agree that if the variance rescaling is done the way Huybers proposes, the RE benchmark comes out around 0.

But the rescaling does not affect the R^2 statistic (since the rescaling term cancels out) so by the R^2 criterion MBH98 remains insignificant, even though it appears to pass the RE test. Huybers does not draw attention to this apparent contradiction, instead he asserts that our RE benchmark was in error and the early portion of MBH98 really was significant.

But his new RE benchmark is not directly comparable to the MBH98 RE stat. When we did the RE benchmark in our GRL paper we simplified the computation by only including a single noise series, rather than 22 noise series (i.e. one for every proxy in the AD1400 network), which was a conservative assumption. For our reply here we re-did the RE benchmarking adding in variance rescaling at the appropriate step and adding in white noise vectors in place of the other 21 series in the network. This yields a new RE benchmark of 0.54, which still exceeds the MBH98 RE value of 0.51, indicating their results are not significant in the AD1400 step. Since our results indicate agreement between the R^2 stat and the RE stat, and since we implement all the (known) MBH98 algorithm steps in the process, we reject Huybers' claim that the early portion of MBH98 is significant. We stand by our original conclusion that the apparent significance of MBH98 was spurious.

2. von Storch and Zorita

VZ (as we call them, for short) also acknowledge that the Mann method mines for hockey sticks and can find them even in red noise data where no hockey sticks exist. They coin the term the "Artificial Hockey Stick" (AHS) effect. But VZ argue that the AHS may not matter because they can show an example using climate model-generated data where Mann's biased PC method does not affect the overall result.

The problem with their example is very simple. They set up a world in their model in which we would not expect the AHS to be very strong. We've never said the AHS appears in every setting, only certain ones: where the data are strongly autocorrelated or where one or more hockey stick shapes are already present. In these cases Mann's algorithm identifies a PC1 as the dominant pattern of variance even though it isn't. But where the underlying data are not heavily autocorrelated and there is not a strong hockey stick shape already lurking in the bin, no AHS is expected.

VZ generated "pseudoproxies" in their model which have a strong correlation to gridcell temperatures and weak autocorrelation. The AHS does not appear to be very influential. But the Mann proxy data has no correlation to gridcell temperatures (the average correlation of a proxy to its local gridcell temperature is -0.08!), the bristlecones have a strong non-climatic growth spurt in the 20th century, and some of the data are strongly autocorrelated. So the conditions are ideal for an AHS effect in MBH98. The VZ example is irrelevant to the assessment of its effects on Mann's results.

We argued in our reply that if VZ accurately modeled the conditions in which an AHS appears, they would generate a high RE statistic and a low R^2 . Unfortunately we were not given access to their code or to their R^2 stats, but our conjecture is that both their RE and R^2 stats are high, proving that they failed to replicate the conditions for observing spurious AHS-driven results.