

## ***Interactive comment on “Millennial temperature reconstruction intercomparison and evaluation” by M. N. Juckes et al.***

**W. Eschenbach**

willis@taunovobay.com

Received and published: 20 December 2006

**ABSTRACT:** The MITRIE paper presents a comparison and evaluation of proxy-based temperature reconstructions. It also introduces its own reconstruction called the Union Reconstruction (UR), based on the CVM method. The UR is argued to be superior to previous reconstructions but, in fact, suffers from exactly the same problems that affect previous reconstructions. As such it is hard to recommend the publication of this paper, presenting as it does, yet another minor variation on existing reconstructions that remains compromised by the same problems (spurious correlation with temperature being the most significant) as the earlier reconstructions. There is evidence that the CVM measure generates spurious results (Granger 1974). One reason for the spurious results is that there are substantial grounds to question the validity of

particular records used as ‘temperature proxies’ in the paper. In addition, there are problems with the models and methods used in the reconstruction. More detailed comments expanding on these themes follow in this and succeeding reviews. Supplementary Online Material (SOM), including figures, proofs, and references, is available at <http://tinyurl.com/yd9bmc>

A) STATISTICAL SIGNIFICANCE AND SPURIOUS CORRELATION: The paper says “The composite tracks the changes in northern hemisphere temperature well, capturing the steep rise between 1910 and 1950 and much of the decadal scale variability. This is reflected in the significance scores (Table 3) which are high both for the full series and for the detrended series.”

The correlation of the Union Reconstruction (UR) with the NH data, while high, is not significant because of the autocorrelation of the two series. There are at least three methods available to establish the significance of correlation between two series each of which has significant autocorrelation. 1) The most common method is the Durbin-Watson Test, which examines the autocorrelation of the residuals between the two series. A Durbin-Watson statistic of less than 1.5 indicates that there is significant correlation in the residuals, indicating that the model is mis-fit to the data. The Durbin-Watson test statistic for the correlation of the UR and the NH data is only 1.4, marking it as not significant. (Shaw 1985, Draper 1998) 2) A second method is that of Quenouille (Quenouille 1952), which gives an effective number of degrees of freedom for two autocorrelated series. This reduced number of degrees of freedom is used in the normal way to calculate a standard “p” value for the significance of the correlation. The Quenouille method for the UR/NH data correlation gives a result of  $p=0.11$ , again showing the correlation is not significant.

Finally, there is the Monte Carlo method, which compares the results of random “red noise” realizations with the UR results. In order for this test to be valid, the red noise proxies must be used in the same way as the proxies in the CVM method. For the CVM method to work, one requirement is that the proxy results that have negative

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

correlations with the NH data must be “flipped” so that they have a positive correlation with the NH data. The reason that the UR CVM proxy has the calculated correlation with the NH data is that the Chesapeake proxy (which has a negative correlation with temperature) is flipped before being averaged. This, of course, is the correct procedure (although it is not mentioned in the text and has not yet been found in the code), and contributes significantly to the correlation seen in the reported results. However, when a Monte Carlo analysis is performed to determine the significance of the correlation, the exact same procedure must be followed - the random red-noise series that have a negative correlation with temperature must be flipped before they are used in the calculation. If this is not done, as in this paper, then the Monte Carlo simulation is not following the same procedure as the CVM method used for the UR, and thus the results of the test are not representative of the UR. When this is done, there are a number of red noise proxies that outperform the UR. An R script for a red noise (random walk) process that outperforms the UR is available at <http://tinyurl.com/ylk4sq>

(It is worth noting that a Monte Carlo method cannot prove that a result is significant, only that it is not significant. If a given method outperforms the same method used with a given red-noise proxy, all the Monte Carlo test proves is that the reconstruction has outperformed a given form of red noise. It does not show that it outperforms all types of red noise. On the other hand, if a given form of red noise outperforms the proxies, this shows that the correlation may well be random, since we cannot reject the null hypothesis. This is particularly true if the red noise is simple, such as a random walk in this case.)

Thus, based on the results of three separate evaluation methods (Durbin-Watson, Que-nouille, and Monte Carlo) we cannot reject the null hypothesis that the correlation of the UR with the NH data is random. This, of course, means that we can place no reliance on the UR as a reconstruction of historical temperatures.

B) BRISTLECONE/STRIPBARK PROXIES: The NAS panel was quite clear about not using Bristlecone/Stripbark tree ring proxies in historical temperature reconstructions.

Despite the NAS recommendations, and other information indicating problems with bristlecones (e.g. Graybill and Idso, Rob Wilson, Wegman Report) that supports the NAS Panel's recommendation, the UR contains no less than four such proxies: Methuselah's Walk, Indian Garden, CA Bristlecone ?C13, and Boreal.

The NAS recommendation is not an isolated example. Biondi et al. (including MBH author Hughes) wrote: "The average of those sites [a network of high-elevation temperature-sensitive tree-ring sites in the Great Basin and Sierra Nevada of Hughes and Funkhouser, unpublished], plotted in Figure 5, is based on many ring-width series, each one being 500 years or longer, without individual growth surges or suppressions and from "strip-bark" five-needle upper forest border pines of great age. Such a record is not a reliable temperature proxy for the last 150 years as it shows an increasing trend in about 1850 that has been attributed to atmospheric CO<sub>2</sub> fertilization [Graybill and Idso, 1993]"

The problem with bristlecones is that since "Such a record is not a reliable temperature proxy for the last 150 years", it cannot be used as a temperature proxy at all, because the record is not reliable during the calibration period. Thus, there is no way to calibrate the earlier period of the proxy record.

While there is no requirement that the MITRIE study follow the recommendation of the NAS Panel or of Biondi et al., there definitely is an obligation to justify the action if the recommendation is not followed. To do so, it is necessary to explain how to calibrate a proxy that has unreliable data during the calibration period.

**C) PROXY SELECTION AND PROCESSING:** It is vital to have clear proxy selection rules. This prevents later accusations of "cherry picking" only those proxies which support a desired conclusion, and makes the conclusions of the study more robust.

The statements about proxy selection in the paper are: "Here, we will restrict attention to records which span the entire reconstruction period from to AD 1000 to AD 1980 (with some series ending slightly earlier, as discussed below).", and "These se-

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

ries have been chosen on the basis that they extend to 1980 (the HCA composites and the French tree ring series end earlier), the southern hemisphere series have been omitted apart from the Quelcaya glacier data, Peru, which are included to ensure adequate representation of tropical temperatures. The MBH1999 North American PCs have been 20 omitted in favour of individual series used in other studies. Finally, the Polar Urals data of ECS2002, MBH1999 and the Tornetraesk data of MSH2005 have been omitted in favour of data from the same sites used by JBB1998 and ECS2002, respectively (i.e. taking the first used series in each case)."

The implied a priori selection rules are 1) proxies must span the period AD 1000 to AD 1980 (with exceptions); 2) individual series are used instead of proxy compilations; 3) older data are used in preference newer data, when both exist for a given site, and 4) proxies will be from the Northern Hemisphere (with exceptions). There is no rule in the MITRIE paper about archived vs. un-archived proxies. However, Dr. Juckes later said that there was a rule that the data be "published data". (See <http://tinyurl.com/yforq7> for a discussion of this issue). There are a number of problems with the rules as stated. (These and further issues will be discussed in Multidisciplinary Review 2 and then 3.)

---

Interactive comment on Clim. Past Discuss., 2, 1001, 2006.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)