

Supplementary Information # 3

Calibration and Verification

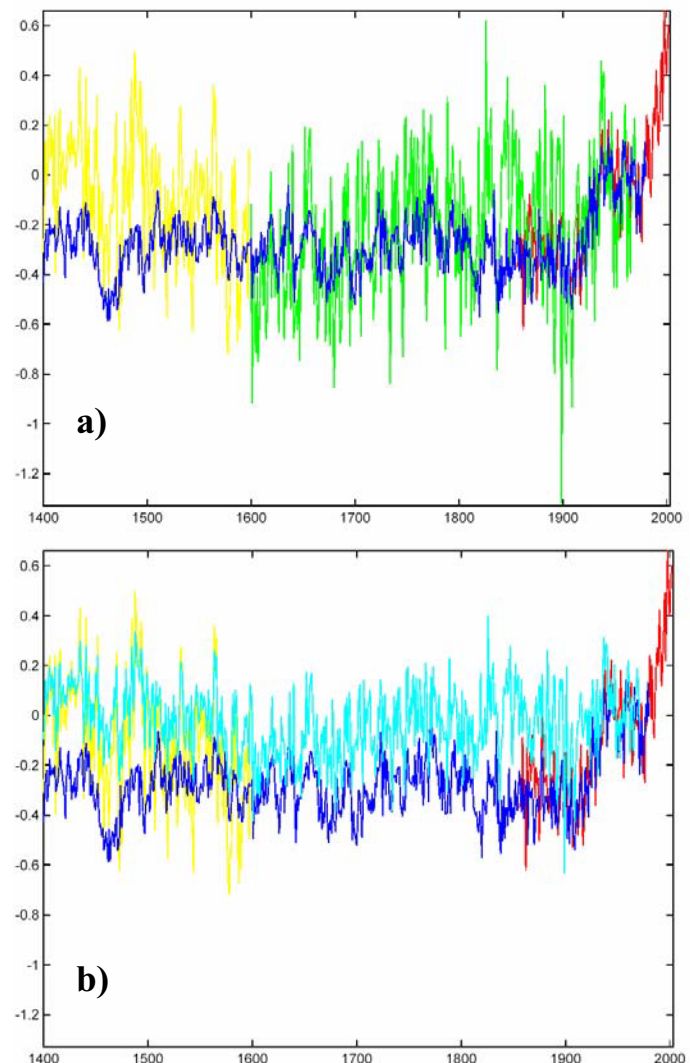
Here is a comparison of calibration/verification statistics for different experiments. Calibration is 1902-1980 unless otherwise indicated.

Table 1.

Experiment	Calibration <i>RE</i>	Verification (1854-1901) <i>RE</i>
1. MBH98, network used to reconstruct back to AD 1400	0.76	0.69
2. MBH98, network used to reconstruct back to AD 1820	0.42	0.51
3. MBH98 redo, 95 indiv prox back to AD 1403 (1902-1971 calib)	0.60	0.39
4. MBH98 redo, 94 indiv prox back to AD 1400 (1902-1971 calib)	0.59	0.33
5. MM04 reproduction, RPC standardization enforced	0.21	-0.16
6. MM04 reproduction, RPC standardization not enforced	-0.52	-0.76

Negative calibration *RE* in exp. (6) results from the lack of enforcement (MM04) of the protocol of MBH98 in which RPCs are scaled to have same variance as corresponding PC series during calibration interval. If this protocol is followed (exp 5), positive calibration *RE* is obtained, but verification *RE* is still negative. Pre AD 1600 reconstruction is similar for both exp (5) and exp (6), and fails statistical verification in both cases. This is due both to a clear mean offset from the actual 1854-1901 NH instrumental record, and a spurious increase in variance (the latter more pronounced for exp. 6) (See Figure 1--note that the calculation of verification statistics must make use of the mask of available 19th century observations, an important detail that MM04 failed to take into account--see discussion below).

FIGURE 1: Comparison of MBH98 (stepwise) reconstruction (blue), CRU 'variable grid' instrumental NH temperature series 1856-2003 (red), with (a) our reproduction of the MM04 reconstruction (yellow through 1600; green through 1971), and (b) an alternative version of the MM04 reconstruction (cyan) which enforces the standardization of MMBH98 that reconstructed PCs have the same variance as the actual annual-mean PC during the calibration period.



Calculation of Verification Statistics for NH mean series

MM04 have made a serious error in how they have calculated verification statistics (both RE and the flawed R^2 diagnostic they promote) for reconstructed NH mean temperature series. They have not taken into account the changing nature of the distribution of observations that contribute to the Jones et al (1999) 'variable grid' instrumental NH series available at the CRU website, which they used for statistical verification back to 1856 (note that the version of the CRU dataset used by MBH98 is available back to 1854). The MBH98 Northern Hemisphere reconstruction represents a 'frozen grid' spatial average over the region sampled by the instrumental record during the 1902-1980 calibration interval (1082 global gridpoints indicated by gray shading in Figure 2). In order to evaluate the fidelity of the reconstruction over the earlier interval (1854-1901) one must apply the mask of the (considerably sparser) available instrumental observations over that earlier interval (219 global gridpoints indicated by 'checkerboard' pattern in Figure 2). The reconstruction and instrumental record must be averaged over this sparser region to avoid a spatial bias in the comparison. Properly applying the mask of available observations during the verification interval before computing a NH mean from the MBH98 reconstruction, yields a very close relationship with the instrumental record over the verification interval (Figure 3), and the high RE statistics reported in Table 1.

MM04 make an incorrect "apples and oranges" statistical comparison of an instrumental 19th century NH record based on spatially sparse observations against a reconstructed NH mean over that interval that is based on averaging the spatial reconstructions over the far more widespread region of the Northern Hemisphere available during the 20th century. This explains their vast underestimate of the RE (and R^2) statistics for the MBH98 reconstruction, and biases their other RE and R^2 statistics as well.

FIGURE 2. Distribution of the (1082) nearly continuous available land air/sea surface temperature gridpoint data available from 1902 onward used for calibration by MBH98 (grey shaded) The squares indicate the subset of 219 gridpoints with nearly continuous records extending back to 1854 used for verification. Northern Hemisphere (NH) temperature is estimated as areally-weighted (ie, cosine latitude) averages over the Northern hemisphere domain.

The rectangle in the eastern equatorial Pacific (the 'Nino3' region) can be ignored.

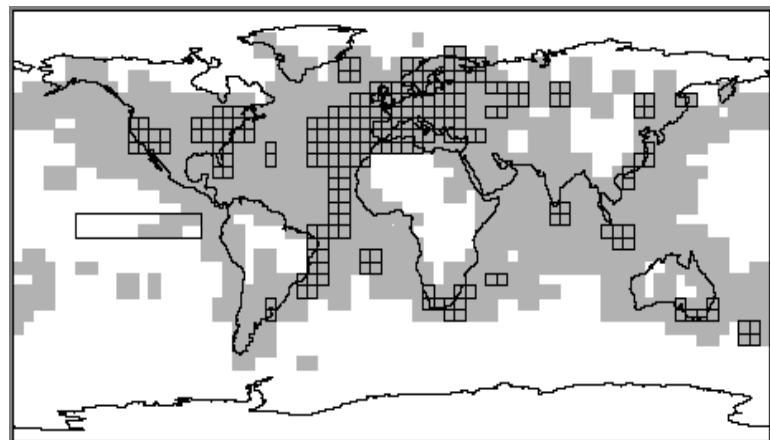


FIGURE 3. Instrumental NH mean temperature (black) averaged over 1082 gridpoints available from 1902-1980 and 219 gridpoints available from 1854-1901. MBH98 Northern Hemisphere mean reconstruction is shown restricted to the same 1082 gridpoints 1902-1980 (blue) and the same 219 gridpoints 1854-1901 (red). The calibration and verification scores are given in Table 1 (experiment 1). [from Mann, M.E., Gille, E., Bradley, R.S., Hughes, M.K., Overpeck, J.T., Keimig, F.T., Gross, W., Global Temperature Patterns in Past Centuries: An interactive presentation, *Earth Interactions*, 4-4, 1-29,2000.]

