SUPPLEMENTARY INFORMATION– PART 1

Global-scale temperature patterns and climate forcings over the past six centuries: A comment.

Stephen McIntyre[1], Ross McKitrick[2]

[1]Toronto Ontario Canada; [2]Department of Economics, University of Guelph, Guelph Ontario Canada N1G2W1.

This SI provides additional supporting material to our Comment on Mann et. al.[1] (MBH98), in which we focus on indicators affecting early 15th century values, primarily the first principal component (PC1) of the North American tree ring roster and secondly, a Canadian tree ring series from Gaspé, Quebec. SI Part 2 is a listing of the computer code used to generate the figures and statistics in the article and this Supplementary Information Part 1.

1. The MBH98 Tree-Ring PC Algorithm and Red Noise Simulations

MBH98 carried out principal component calculations on tree-ring site "chronologies"[2] in 6 regions, stating that:

> Certain densely sampled regional dendroclimatic data sets have been represented in the network by a smaller number of leading principal components (typically 3–11 depending on the spatial extent and size of the data set).

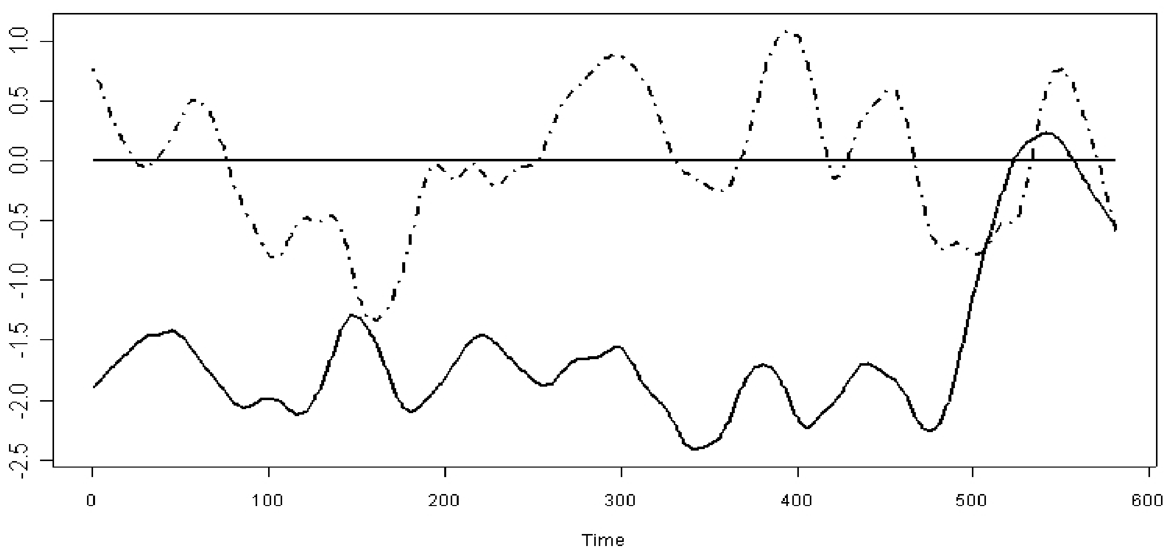
No other description of this methodology was contained in MBH98.[3]. The FTP site for MBH98, started in July 2002, contained the FORTRAN program[4] used to calculate the principal components for the North American region (NOAMER). Examination of this program showed that the site chronologies were first transformed by "short-segment standardization"[5], which entailed subtracting the 1902-1980 mean, then dividing by the 1902-1980 standard deviation and again by the 1902-80 detrended standard deviation. Then the principal components were calculated using singular value decomposition on the transformed data rather than on the covariance or correlation matrix. WDCP tree ring chronologies are already scaled to a mean index value of 1000. For series in which the 1902-1980 mean is similar to the 1400-1980 mean, the subtraction of the 1902-1980 mean has little impact on weightings. Series in which the 1902-1980 mean is different than the 1400-1980 mean are de-centred in this operation, and the variance is thereby inflated relative to other series. As a result, they become overweighted in the PC1 by the principal components algorithm, which weights series according to a variance

---

[1] Mann, M.E., Bradley, R.S. & Hughes, M.K. (1998) *Nature*, 392, 779-787.
[2] A site chronology is a standardized index of the average annual ring widths measured from a group of trees at a single place. The World Data Center for Paleoclimatology is the primary archive.
[3] The SI at < http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v392/n6678/abs/392779a0_fs.html> listed sites for 5 of 6 regions.
[4] <ftp://holocene.evsc.virginia.edu/pub/MBH98/TREE/ITRDB/NOAMER/pca-noamer.f.>.
[5] Applying a phrase employed by a referee.

maximization rule. The overweighting in MBH98 is increased further by using singular value decomposition on the transformed data rather than the covariance matrix. The result is that the principal components calculated in MBH98 are dominated by series in which the 1902-1980 mean differs from the 1400-1980 mean, i.e. those with a hockey-stick shape.

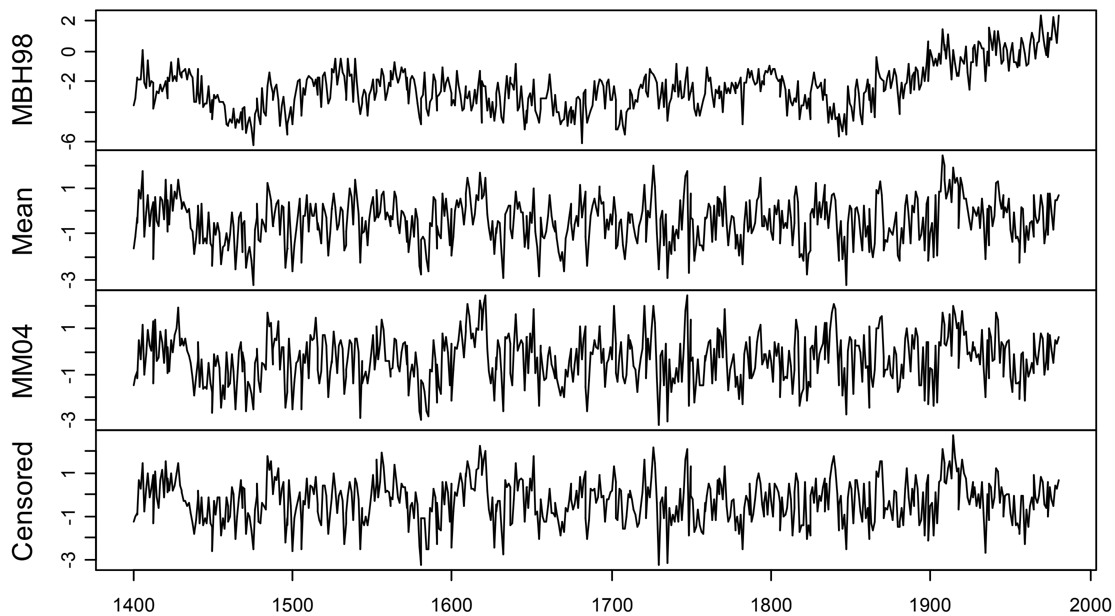
As an illustration of the effect of the MBH98 algorithm, we generated 70 series of stationary red noise and compared the PC1 from the MBH98 algorithm to that of a calculation without “short-segment standardization”. To generate the data, we took the 70 NOAMER sites available back to 1400 and fitted a lag 1 autoregression model to each. The coefficients ($\beta^1,...,\beta^{70}$) were all of magnitude less than 1. Then we generated 70 random vectors $a_t^1,...,a_t^{70}$ of length 1081, using the AR1 series $a_t^i = \beta^i a_{t-1}^i + e_t^i$, i=1,…,70. Each series was initialized at zero and the e_t^i vectors were N(0,1). The first 500 values were then dropped from each series, yielding 70 vectors of stationary red noise, each of length 581 (corresponding to the length of the 1400-1980 period). The (conventional) first principal component from these 70 series, after smoothing, exhibits the expected stationary sawtooth pattern (SI Figure 1, dashed line). When MBH98 method was applied (using rows 503-581 as the “short segment”), the PC1 consistently had a hockey-stick shape as shown in the example in SI Figure 1 (solid line). The reason for the hockey stick-shape is that some of the underlying vectors randomly trail up or down at the end of their length, and these are selected for high weighting by the MBH98 method. Ten simulations were carried out and a hockey stick shape was observed in every simulation.



SI FIGURE 1. First principal component from 70 stationary red noise series, smoothed with Lowess (f=0.1). Dashed line: Conventional method (princomp command in the programming package R). Solid line: MBH98 algorithm, which standardizes on a short segment (1902-1980) prior to applying singular value decomposition.

2. The Different NOAMER PC1 Versions

SI Figure 2, top panel, shows the NOAMER PC1 for the AD1400 network from MBH98. It corresponds to Figure 1b in the Comment. SI Figure 2, second panel, is the simple mean of the proxies. SI Figure 2, third panel, shows the NOAMER PC1 calculated using the command "princomp" in the statistics package R, without applying short-segment standardization. The second and third panels closely resemble one another (correlation 0.92).



SI FIGURE 2. Top panel: PC1 of the post-1400 NOAMER tree ring network, calculated by MBH98 using short-segment standardization. Second panel: simple mean of proxies. Third panel: PC1 using standard software without short-segment standardization. Bottom panel: Unreported PC1 calculated by MBH after censoring Graybill-Idso high-altitude series. All normalized to 1902-1980.

The data for the fourth panel is from a directory at the MBH98 FTP archive.[6] It is an unreported alternate version of the NOAMER PC1 calculated by "censoring" 20 high-altitude sites (as discussed in the Comment). It too bears a striking resemblance to the MM04 PC1 (correlation 0.95), but it was calculated by MBH themselves. The hockey stick shape of the top panel is not reflected in any of the other versions. The top panel does, however, resemble the mean of the Graybill-Idso sites.

---

[6] <ftp://holocene.evsc.virginia.edu/pub/MBH98/TREE/ITRDB/NOAMER/BACKTO_1400-CENSORED>

3. Identification of Overweighted and Censored MBH98 Sites

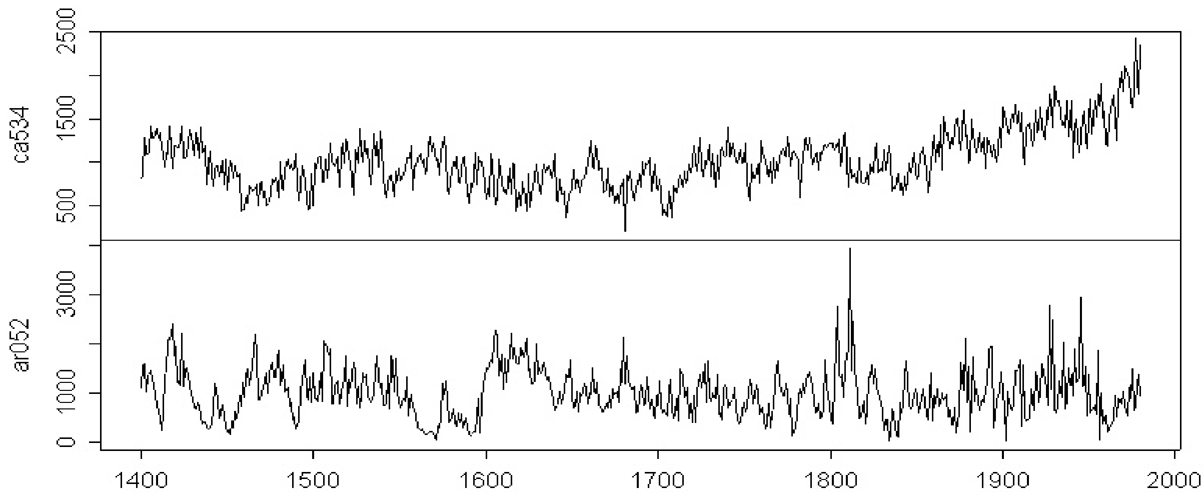
Overweighted sites were identified as follows. We bound the weighting factor eof01.out (length 70) in the directory BACKTO_1400 at Professor Mann's FTP site to the identification code vector noamer-itrdb-ad1400.txt (length 70) and sorted in decreasing absolute value of the eof01. The name, elevation and author of each site with >25% of the weighting of the top-weighted site was manually looked up at the WDCP FTP site using the identification code. The site names and locations were manually compared to Table 9.1 of Graybill and Idso resulting in the first 6 columns of the following table:

ID Code	Name	Species	Elevation (m)	Author	Graybill-Idso (1993) #	Exclusion in MBH98 CENSORED
ca528	Flower Lake	PIBA	3291	D.A. Graybill	13	TRUE
ca529	Timber Gap Upper	PIBA	3261	D.A. Graybill	14	TRUE
ca530	Cirque Peak	PIBA	3505	D.A. Graybill	12	TRUE
ca533	Campito Mountain	PILO	3400	D.A. Graybill and V.C. Lamarche	5	TRUE
ca534	Sheep Mountain	PILO	3475	D.A. Graybill	11	TRUE
ca555	Yolla Bolly	PIBA	2460	B. Buckley		FALSE
co523	Windy Ridge	PIAR	3570	D.A. Graybill	4	TRUE
co524	Almagre Mountain	PIAR	3536	D.A. Graybill	1	TRUE
co525	Hermit Lake	PIAR	3660	D.A. Graybill	3	TRUE
co535	Frosty Park	PIFL	3218	D.A. Graybill		TRUE
co545	Niwot Ridge	PIFL	3169	D.A. Graybill		TRUE
nv510	Charleston Peak	PILO	3425	D.A. Graybill	6	TRUE
nv511	Mount Jefferson	PIFL	3300	D.A. Graybill	7	TRUE
nv512	Pearl Peak	PILO	3170	D.A. Graybill	9	TRUE
nv513	Mount Washington	PILO	3415	D.A. Graybill	8	TRUE
nv514	Spruce Mountain	PILO	3110	D.A. Graybill		TRUE

SI TABLE 1. 16 SITES MOST HIGHLY WEIGHTED IN MBH98 NOAMER PC1, ORDERED BY IDENTIFICATION CODE

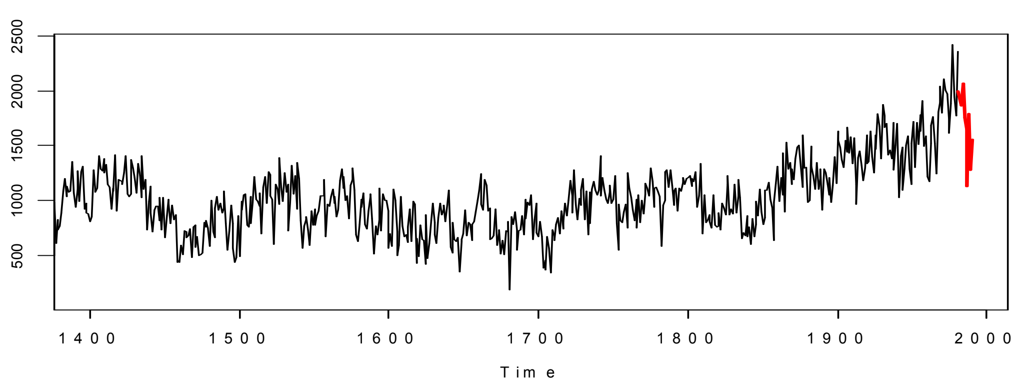
There is no identification of the series in the BACKTO_1400-CENSORED directory at the MBH98 FTP site. The "censored" sites were identified as follows. In the directory FTP/ITRDB/NOAMER, the file noamer-censored.inf has 192 series and the file noamer.inf has 212 series (as compared to 232 series in the original Nature SI). The 20 excluded series were identified by matching. There is a roster of the 70 series in the BACKTO_1400 file (as noted above) and these were recovered. The excluded series were matched against the 70 series in the BACKTO_1400 roster and all 20 series were matched, leaving 50 series unmatched in the exclusion file. The length of the weighting vector eof01.out in the BACKTO_1400-CENSORED file was exactly 50 – evidencing that the 50 series used in the BACKTO_1400- CENSORED calculations were obtained by "censoring" the 20 series. The series so censored were compared manually to the Graybill-Idso table in column 7.

The top-weighted site, Sheep Mountain CA (ca534), exhibits the distinctive hockey stick shape of the final MBH98 Northern Hemisphere temperature index, while another NOAMER site, Mayberry Slough AR (ar052), has a growth peak in the early 19th century (SI Figure 3). The MBH98 algorithm assigns 390 times the weight to Sheep Mountain compared to Mayberry Slough in the PC1.



SI FIGURE 3. Two NOAMER site chronologies used in MBH98. The first series is given 390 times the weight of the second series. The transformation of data by MBH98 prior to principal components calculation causes series with high 20th century growth rates to be over-weighted.

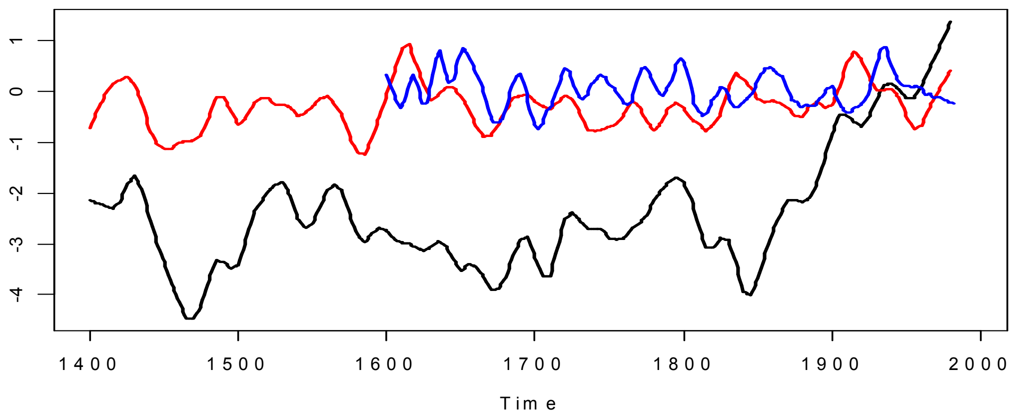
Ironically, the Sheep Mountain CA series drops sharply in the 1980s (see below), after the calibration period.



SI FIGURE 4. Site chronology for ca534 (Sheep Mountain CA) showing the post-1980 decline (red) in values.

4. Nonlinearity of NOAMER PC1

Figure 5 graphs smoothed versions of the MBH98 PC1 (black line) against the conventional PC1 (red line) and adds for comparison a smoothed version of the North American temperature index calculated by Briffa (1992).[7] The nonlinearity of the MBH98 line in the 20th century is conspicuous together with the unphysical negative values in all early periods.



SI FIGURE 5. All curves standardized on 1902-1980 and smoothed (lowess f=0.05) and. Black – MBH98 NOAMER PC1. Red – NOAMER PC1 without standardization on short segment (virtually identical to "censored" PC1). Blue – North American temperature reconstruction from tree rings in Briffa (1992).

5. The Gaspé Series

The St. Anne River, Gaspé ring width series was used twice by MBH98, as an individual proxy (treeline11) and in the NOAMER principal components roster (as series cana036). In its use as treeline11.dat (but not as cana036), MBH98 extrapolated the first four years. This is the only example of extrapolation of early values in MBH98. Such unique handling would seem to call for explicit justification, but instead was left disclosed.[8] This ad hoc handling allowed the inclusion of this series in the AD1400 roster, where it had a substantial impact as discussed below.

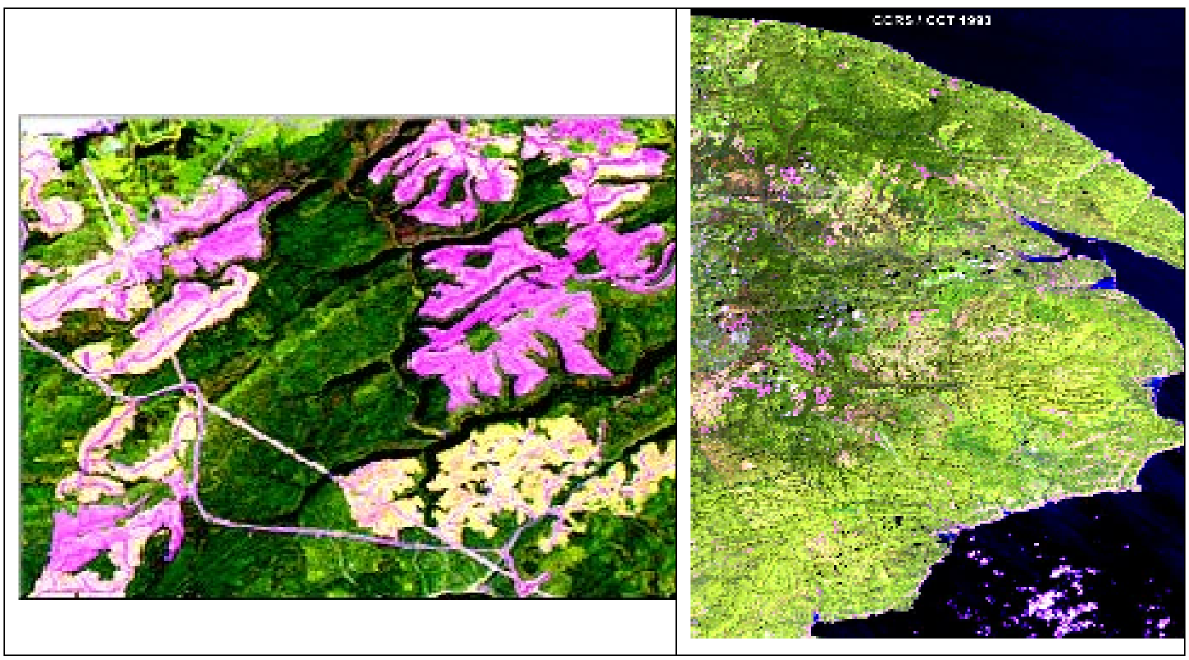
The underlying dataset commenced in 1404. It is based on only one tree up to 1421 and only 2 trees up to 1447. The early portion of the series therefore fails standard

[7] Briffa, K.R., Jones, P.D. and Schweingruber, F.H., 1992. Journal of Climate 5(7), 735-754.
[8] Its disclosure was required in the Corrigendum by Mann et al.

minimum signal criteria[9]. This portion of the series was not used by Jacoby and d'Arrigo (1989) or D'Arrigo and Jacoby (1992)[10], whose analysis began effective 1600.

Although Gaspé, Quebec is described as a "northern treeline" series by Mann et al. (2003) and in the cited reference, it is in fact nowhere near the northern treeline, which is at least a thousand km away. Natural Resources Canada[11] describes the region as follows (photos below): "Forest covers about 95 percent of the Gaspé Peninsula. Fir, birch and maple are typically found here, as well as the impact of human activity. This image segment shows clearcutting in the peninsula's interior, with recent cuts shown in bright pink. Less recent cuts, where the early vegetation is starting to grow back, is a yellow colour, while the older cutting sites (where the regrowing vegetation is more mature) is a light green. The uncut, standing forest is a dark green colour. From images such as this, forestry agencies and companies can determine how much cutting has taken place during a particular season, and whether natural regeneration needs to be "assisted". Logging roads can also be mapped, whether for logging, firefighting, recreation or other uses."

The St Anne River series is a cedar series, from a site located just to the west of the image below.



SI FIGURE 6. Two overhead images of sector of Gaspé peninsula, coloring described in text above. The figure at left is smaller scale.

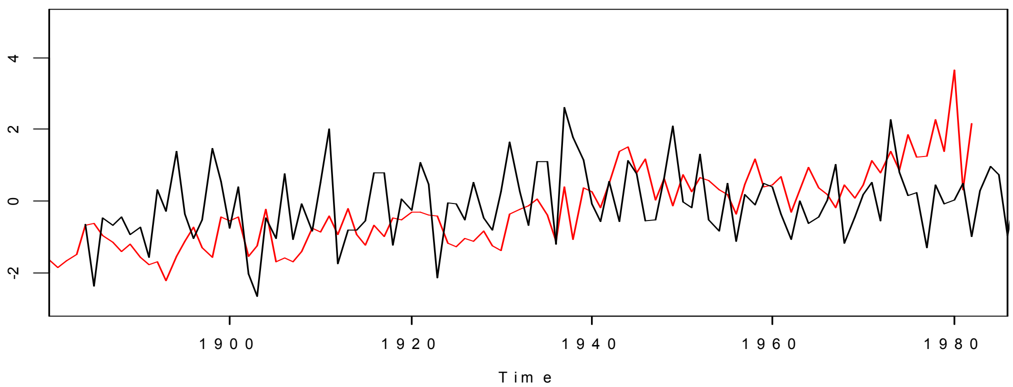
---

[9] Wigley, T.M.L., Briffa, K.R. and Jones, P.D. (1984). *Journal of Climate and Applied Meteorology* 23, 201-213.

[10] Jacoby and d'Arrigo, (1989). *Climatic Change* 14, 39-59; D'Arrigo, R.D. and Jacoby, G.C. (1992) in *Climate Since A.D. 1500,* (eds. Bradley, R.S. & Jones, P.D., 246-268, Routledge, 1992).

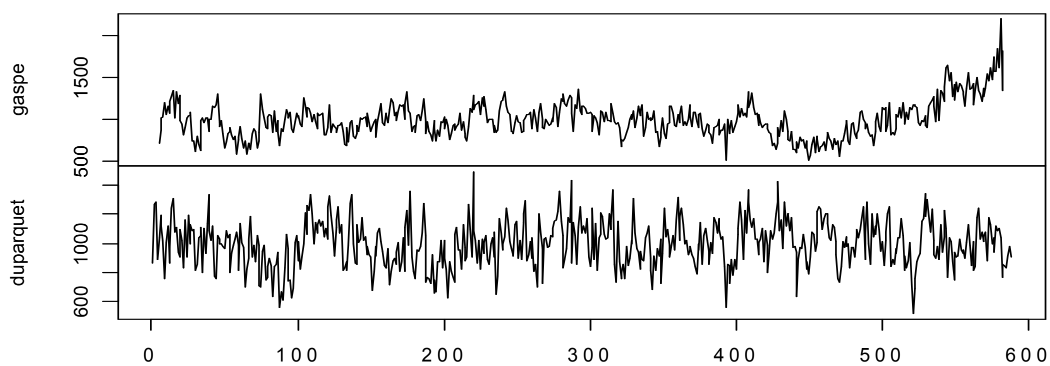
[11] <http://www.ccrs.nrcan.gc.ca/ccrs/learn/tour/36/36scene6_e.html>

The Gaspé, Quebec tree-ring series has an R^2 against the summer CRU grid-cell temperature for (47.5N, 67.5W) of only 0.03. The 20th century portion of this series thus exhibits nonclimatic factors or a nonlinear temperature response. The two series are plotted against each other below for the overlapping period 1884-1982.



SI FIGURE 7. Red – Gaspé tree-ring chronology. Black – CRU temperature series for (47.5N, 67.5W). Both series (1884-1982) normalized to 1902-1980.

The 20th century growth rate of the Gaspé site has a markedly different pattern from the only other eastern Canadian cedar chronology at Lac Duparquet as shown below, showing that the enhanced growth rate is not characteristic of all such sites.)

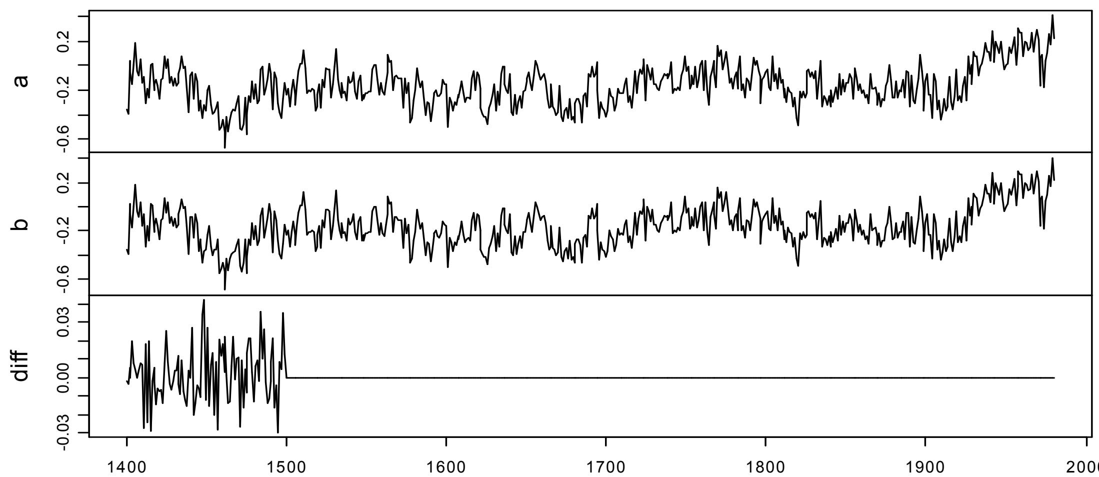


SI FIGURE 8: Tree ring width site chronologies. Above: St. Anne River, Gaspé, Canada (cana036). Below: Lac Duparquet, Canada (cana106).

Other than being referred to by Jacoby and d'Arrigo, this site has never been published (Edward Cook, pers. comm.) Careful and detailed study (not carried out in MBH98) would be required to support its early 15th century effect discussed below.

6. The Stahle/SWM Series

Mann et al. (2003) stated that the Stahle/SWM network (i.e. the tree ring roster from the US Southwest and Mexico, collected by D. Stahle) was a “key” indicator. However, the presence or absence of the SWM network in the 15th century has very little impact: less than 0.035 deg C. in all years, as shown below.



SI FIGURE 9: Northern Hemisphere temperature reconstructions. a: our emulation of MBH98 (same as Figure 1d); b: emulation without Stahle/SWM PC1 in the 15th century; c: difference between the two reconstructions.

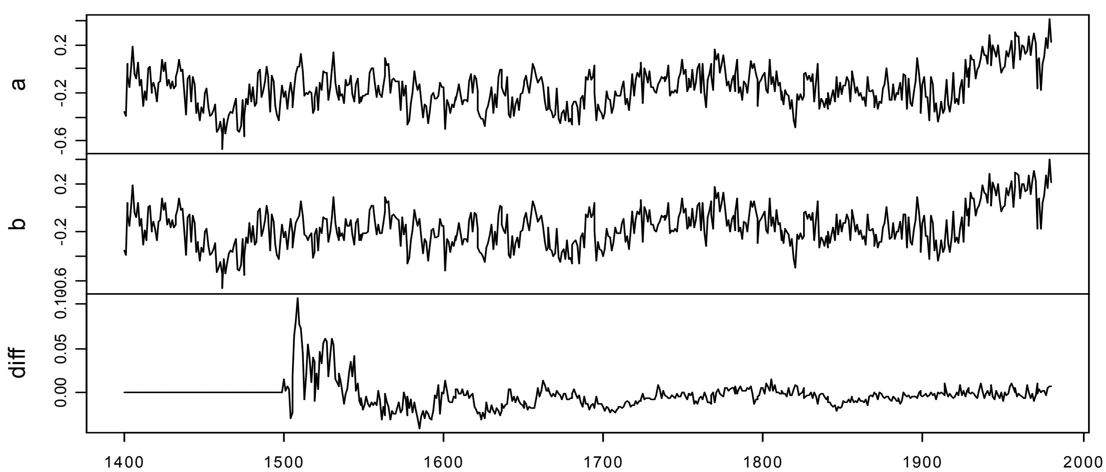
Regardless of the actual impact, our exclusion of the 15th century portion of this series was justified on quality control grounds. MBH98 attributed the network to Stahle and Cleaveland (1993) but have recently stated [Mann, Bradley and Hughes, Corrigendum, *Nature* forthcoming] that it derives from Stahle et al. (1998).[12] For each site in the SWM network, MBH98 use two series: earlywood and latewood widths, although Stahle et al. (1998) do not use latewood widths. Of the SWM sites listed in the original MBH98 Supplementary Information (SI), only 2 are available before 1450. In the FTP data actually used in MBH98, there are 3 sites (6 series) extending back prior to 1450. Two of these sites (4 series) have identical values for the first 120 years for earlywood widths and the first 125 years for latewood widths, each differing thereafter, suggesting that the data are either spliced versions of different sites or different editions of the same site. Either way at least one of the two sites is clearly ineligible pre-1450, leaving only two potentially eligible sites – an insufficient number to constitute a network in usual MBH98 usage. One of them is Spruce Canyon CO, which is already used in the NOAMER roster and should therefore be dropped from the SWM group. The data for the

[12] Stahle, D.W. et al. (1998). *Bulletin of the American Meteorological Society,* 79(10), 2137-2152; Stahle, D.W. and Cleaveland, M.K. (1993). *J. Climate*, 6, 129-139.

remaining SWM site, Cerro Barajas, as used in MBH98, includes physically impossible negative values in the early portion of the series, which are not present in the version archived at WDCP. The unsuitability of the SWM PC1 in the AD1400-1450 period is clearly established and we note that Stahle et al. themselves did not apply their network prior to 1706.

7. The Twisted Tree series

With respect to the other "key indicator", the Twisted Tree series as used by MBH98 does not begin until 1459 so it is irrelevant to the pre-1450 interval. However, there is no justification for not using the updated version that we used. MBH98 cites Jacoby and d'Arrigo (1989)[13] as a source for 11 northern treeline sites. Jacoby and d'Arrigo obtained additional samples from several sites in or about 1992 and carried out further quality control, which is reflected in the site chronologies archived at the World Data Centre for Paleoclimatology (WDCP) in the 1990s. MBH98 used older 1989 versions, some of which differ from the final archived versions. The version of the Twisted Tree series used in MBH98 was the older, unarchived edition that ended in 1976 and included a pre-16th century portion based on a single tree. The archived version, which we used, ends in 1992 and only commences in 1529, when 3 trees become available.[14] The impact of using the updating version is only somewhat material in the early 16th century, when it reaches 0.1 deg C in a few years as shown below.



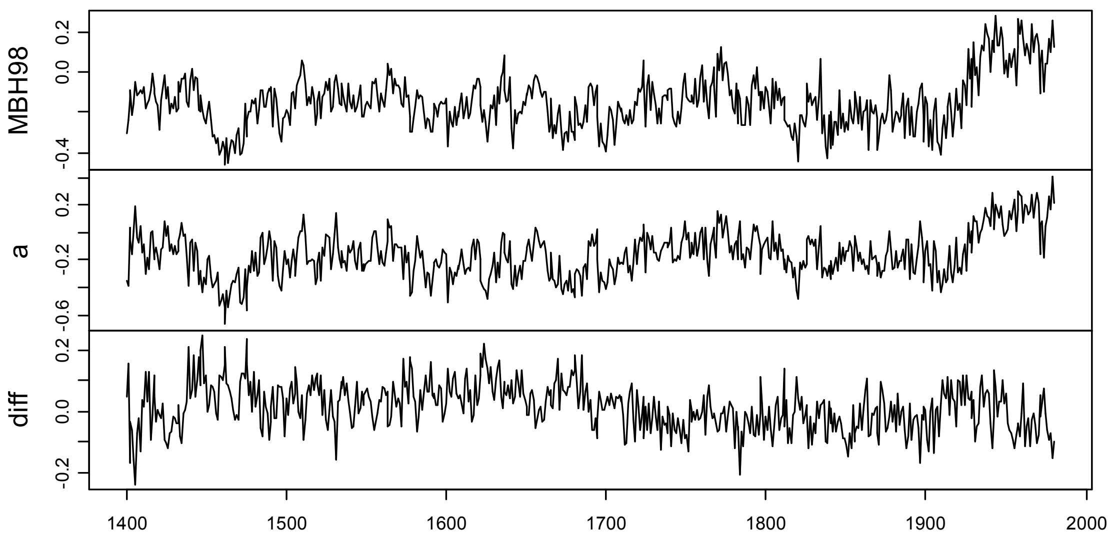
SI FIGURE 10: Northern Hemisphere temperature reconstructions. a: our emulation of MBH98 (same as Figure 1d); b: emulation using updated Twisted Tree series, which commences later than the obsolete version used in MBH98; c: difference between the two reconstructions.

---

[13] Jacoby, G.C. and D'Arrigo, R.D. 1989. *Climatic Change* 14: 39-59.
[14] We considered this example of a 3 tree minimum in assessing the Gaspé series.

8. Emulation of the Original MBH98 Graph

The figure below shows the MBH98 northern hemisphere temperature index, together with our emulation and the difference. The differences are generally less than 0.1 deg C, but do reach 0.2 deg C in some years. Remaining methodological differences between MBH98 and us are immaterial to our conclusions, since our emulation of MBH98 shows uniquely high 20th century values of the MBH98 Northern Hemisphere temperature index, but not when using the same methodology after switching to conventional PC analysis and removing the Gaspé extrapolation. It is our view that our emulation is sufficiently precise that the impact of key methodological issues, such as principal component methodology and Gaspé extrapolation, can be fairly assessed.



SI FIGURE 11: Northern Hemisphere temperature reconstructions. a: MBH98 (as in Figure 1a). b: our emulation of MBH98 (same as Figure 1d); c: difference between MBH98 and our emulation.

We have diligently attempted to obtain copies of the computer programs used in MBH98 in order to effect a complete reconciliation, but have been refused. Given the widespread public reliance on the results of MBH98, we find this more than surprising. We have provided our own software in Part 2 of this Supplementary Information. The description of methodology in MBH98 is sketchy and, based on recent statements, is occasionally inaccurate. Based on a materials complaint we submitted, *Nature* has requested from Professor Mann a Corrigendum that will include an upgraded SI to contain an improved description of methodology, but, as of the date of this submission, this SI is not available to us. Several possible areas that may contribute to the discrepancies are: (1) Mann et al. have recently said that they used 159 series in their reconstruction (a figure nowhere mentioned in the original article). A full specification requires listing of the number of principal components used in each of 6 regions by calculation interval. We have estimated this specification as best as we could, but there may be further differences.[15] (2) Mann et al. standardized the CRU grid cells, but did not report the standardization factors. When we did the calculation, the CRU dataset had been updated. We calculated grid cell standardization factors from the new CRU dataset and there will undoubtedly be differences with the older dataset used by MBH98. It is hard to see why the changes in standardization factors would be material to the calculations, but there could be an effect.

Other differences may result from some other undisclosed procedure in MBH98. For example, we discovered the short-segment standardization used in MBH98 only by inspection of Fortran routines at the FTP site. There is no discussion of this methodology in MBH98 and it is such an odd practice in a principal components calculation that it would have been impossible to conjecture that they did it. It was mere chance that the Fortran routine was left on their FTP site, since most computer programs used in MBH98 are unavailable.

The emulation algorithm used here differs in some important respects from the algorithm used in McIntyre and McKitrick (2003) (MM03)[16], to which Mann et al. have made reference. In the earliest portion of tree ring regions, some sites have missing data. Conventional principal components algorithms fail with missing data. MBH98 did not describe the handling of missing data in tree ring regions and Mann et al. refused to explain this matter upon request. In MM03, we calculated principal components for tree ring regions for the maximum period for which all sites were available. Since some sites were unavailable in the 15th century, this resulted in the North American PC1 and PC2 and Stahle/SWM PC1 being unavailable in the 15th century in the reconstruction carried out in MM03. MM03 reported high early 15th century values using our emulation of MBH98 on the resulting dataset. In their response to MM03, Mann et al. (2003)[17] reported that they obtained very similar high early 15th century values in a run in which three "key indicators" – the NOAMER PC1, the Stahle/SWM PC1 and the Twisted Tree series - were excluded. The latter two indicators are now seen to have little effect on the differing early 15th century values – which are driven by the NOAMER PC1 and Gaspé series, as discussed in the Comment and elsewhere in this SI. We specifically modified our algorithm to carry out stepwise principal component calculations, and, in particular, as noted elsewhere, we confirm that we used the NOAMER PC1 and PC2 in the 15th century reconstructions discussed in this submission.

---

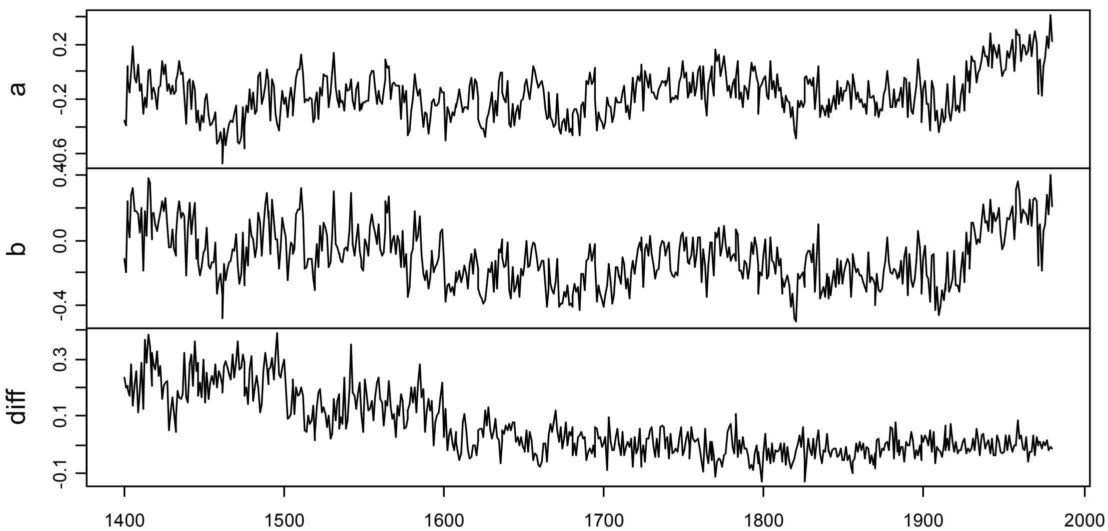
[15] See table at end of SI Part 2.

[16] McIntyre, S. & McKitrick, R. (2003). *Environment and Energy* 14(6) pp. 751-771.

[17] Mann, M.E., Bradley, R.S. & Hughes, M.K. (2003) "Response to M&M" http://www.cru.uea.ac.uk/~timo/paleo/.

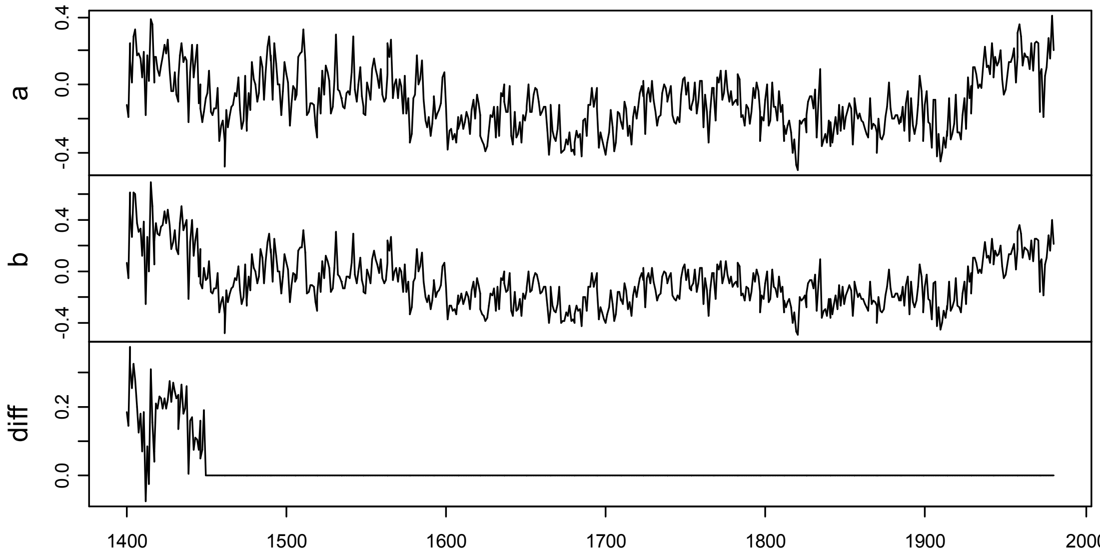
9. Effect of NOAMER PC1 and Gaspé

The specific effect of merely replacing the idiosyncratic MBH98 PC methodology with a conventional principal component algorithm is illustrated in SI Figure 12 below, showing an effect of between 0.1 and 0.4 deg C throughout the 15th century. This is because the MBH98 North American PC1 has strong negative values throughout the pre-1850 period (compare to SI Figure 5), which increasingly depress the resulting Northern Hemisphere temperature index back in time as the size of the roster decreases and their relative effect increases.



SI FIGURE 12: Northern Hemisphere temperature reconstructions. a: our emulation of MBH98 (as in Figure 1d and SI Figure 11). b: using conventional principal components (without short-segment standardization). The NOAMER PC1 and PC2 are used back to 1400; c: difference between the two runs.

Merely applying conventional PC methodology is sufficient to remove much of the distinctiveness of the late 20th century portion of the climate index. We then examined the specific effect of the ad hoc extrapolation of the Gaspé series. This resulted in a further depression of the overall temperature index by up to 0.4 deg C in the first half of the 15th century as shown in SI Figure 13 below. The temperature index resulting from these two steps had early 15th century values exceeding 20th century values and is illustrated in Figure 1e.

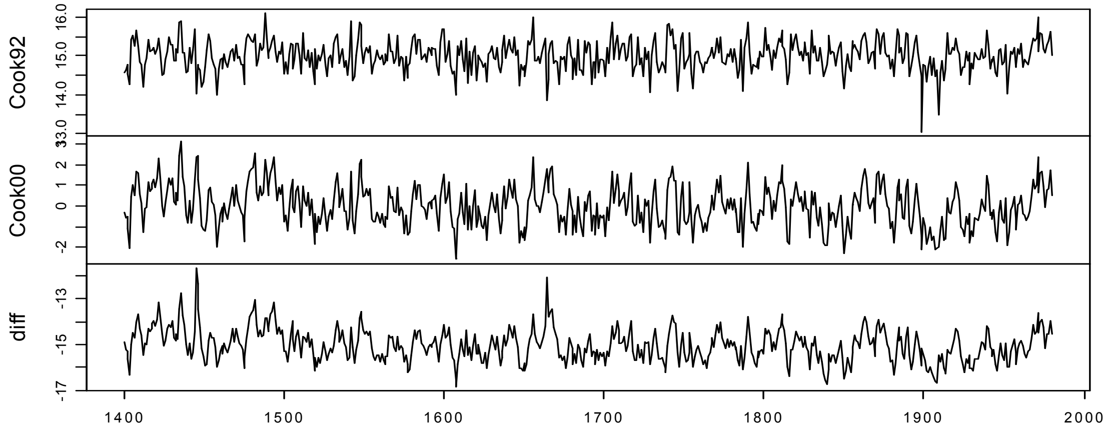


SI FIGURE 13: Northern Hemisphere temperature reconstructions. a: our calculation using conventional principal components (without short-segment standardization) as in SI Figure 12. b: After removing 4 years of Gaspé extrapolation (same as Figure 1e). c: difference between the two runs.

10. Other Data Quality Issues and Reconciliation to MM03

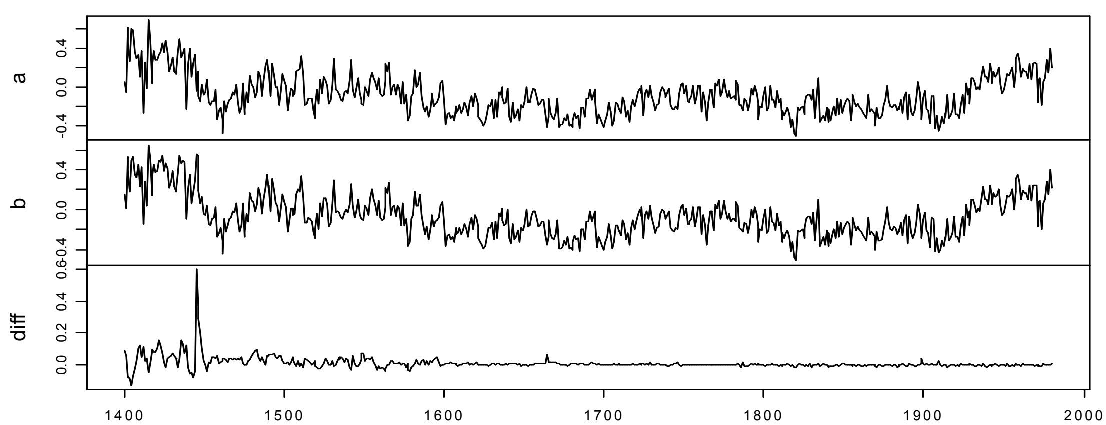
In McIntyre and McKitrick (2003)[18], we reported on a number of data quality issues in MBH98. Mann et al. are issuing a Corrigendum concerning many of these issues. One of the most important concerns in McIntyre and McKitrick (2003) was the use in MBH98 of obsolete versions of many series and truncated versions of several other series. Some series were obsolete in 1998, while others have been superceded afterwards. The effect of the use of the obsolete version of the Twisted Tree series has been already discussed. The impact of these series updates turns out to be less than that of the NOAMER PC1 and Gaspé series discussed above. The largest change results from Cook's new (2000) reconstruction of Tasmanian temperature. Cook's new series has significant increases in reconstructed temperature in several years in the first half of the 15th century, as shown below:

[18] McIntyre, S. & McKitrick, R. (2003). *Environment and Energy* 14(6) pp. 751-771.



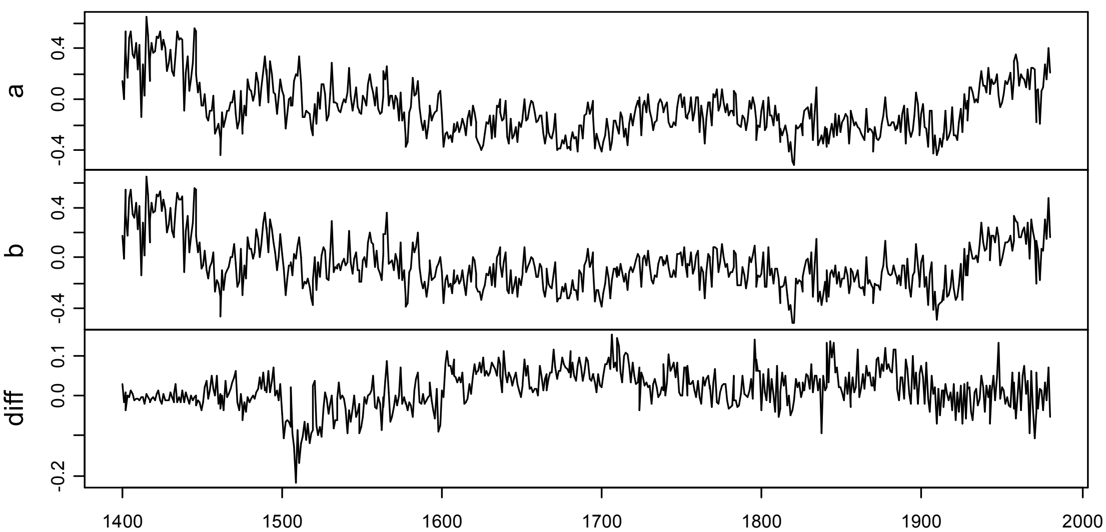
SI FIGURE 14: Cook's Tasmania temperature reconstructions. a: 1992 reconstruction used in MBH98; b: Cook's 2000 reconstruction c: difference between the two runs. Series b is an anomaly rather than a temperature. This does not affect calculations, since all proxies are standardized before regression.

The effect of the new data is material only for a few 15th century years and do not affect relative 15th and 20th century levels, as shown in SI Figure 15 below.



SI FIGURE 15: Northern Hemisphere temperature reconstructions. a: our calculation with conventional principal components and no Gaspé extrapolation (as in Figure 1e and SI Figure 13b). b: With updated Tasmania series. c: difference between the two runs.

The effect of all the other data corrections described in MM03 are shown below, primarily resulting from the use of updated data versions as described in MM03. There are consistent differences, but not rising to the levels of the ones selected for special analysis above.



SI FIGURE 16: Northern Hemisphere temperature reconstructions. a: our calculation as in Figure 15b). b: With updates as described in MM03. c: difference between the two runs.

11. Goodness-of-Fit Verification Statistics

Goodness of fit is measured by how well the proxies and temperatures fit each other over the calibration interval (1901-1980) and how well the predicted temperatures match the withheld observational data over the verification interval (1856-1900). We carried out calculations of RE and R^2 statistics both for the full stepwise reconstruction and for the reconstruction using only the AD1400 roster, to show the specific effect of the different PC1 versions and the Gaspé extrapolation. The results are summarized in Table 2 below. All calculations use the most recent CRU northern hemisphere data (downloaded March 2004).

The first column shows the values reported in MBH98 or Mann et al. (2004). The second column shows the RE value (0.586) for the MBH98 stepwise reconstruction compared to the most recent CRU data, which is lower than that reported in MBH98 (0.69). We surmise that the RE of their reconstruction using only the AD1400 roster fitted to the most recent CRU data will likewise be lower (probably around 0.39). MBH98 did not report the R^2 values of their reconstruction, but we note that the R^2 value of their stepwise reconstruction in the verification period is only 0.123 and we surmise that the R^2 value of their reconstruction using only the AD1400 roster would be lower still (between 0 and 0.1).

	MBH98		Emulation	Corrected	Gaspé
	Reported	CRU 2004		PCs	AD1450
STEPWISE VERSION					
RE	0.69	0.586	0.430	0.434	0.434
R^2 –calibration interval	NA	0.601	0.453	0.453	0.453
R^2- verification interval	NA	0.123	0.003	0.001	0.001
AD1400 ROSTER					
RE	0.51	NA	0.298	-0.026	-1.043
R^2 –calibration interval	NA	NA	0.332	0.343	0.280
R^2- verification interval	NA	NA	0.008	0.004	0.006

TABLE 2. VERIFICATION STATISTICS FOR STEPWISE RECONSTRUCTION AND RECONSTRUCTION USING ONLY AD1400 ROSTER FOR SEVERAL CASES. MBH98 – as reported; CRU 2004 – comparing MBH98 data to CRU Northern Hemisphere edition of March 2004; Emulation: using MM04 replication of MBH98 on MBH98 data; Corrected PCs: Emulation after replacing PCs with conventionally calculated versions; Gaspé AD1450: Same, after removing Gaspé extrapolation.

Our emulation of the MBH98 stepwise reconstruction had somewhat lower RE and R^2 values than either those reported in MBH98 or that we could verify against CRU 2004. Our emulation of the reconstruction using the AD1400 roster had lower RE values (0.3 versus 0.51), of which about half the difference is probably attributable to lower RE values for all reconstructions against CRU 2004.

As seen in the first 3 lines of the next two columns, applying the steps relating to principal components and Gaspé extrapolation described above results in virtually no difference in the RE and R^2 values under stepwise reconstruction. During the verification period, there is a roster of 112 proxies and any statistics resulting from the stepwise procedures provide little illumination for issues pertaining to the AD1400 roster.

Application of the steps relating to principal components and Gaspé extrapolation using only the AD1400 roster reveals a material difference to the RE statistic as shown in the last 3 lines of Table 2. Merely using conventional PC methods, the RE falls from 0.298 to -0.026; removing the Gaspé extrapolation reduces it further to -1.043. Higher RE values for the AD1400 roster are therefore obtained by MBH98 using series which fail the linearity requirements of MBH98 methodology (as discussed in the Comment), linearity requirements for the RE or R^2 statistic itself to be meaningful and/or requirements for the statistic to contain a meaningful dendrochronological signal.

The R^2 statistics in the verification period for our emulation of MBH98 (both using the AD1400 roster and a stepwise reconstruction) are effectively 0. A good model should obviously have good results in a number of statistical tests and the very low R^2 values in the verification period contradict claims of statistical skill for MBH98 methodology.