Presentation to the National Academy of Sciences Expert Panel, "Surface Temperature Reconstructions for the Past 1,000-2,000 Years." : Supplementary Comments

Stephen McIntyre,
Toronto Ontario

Ross McKitrick, Ph.D.
Associate Professor
Department of Economics
University of Guelph

April 3, 2006

We are writing to provide some brief supplementary comments in response to the panel presentations on March 2-3, 2006.

1. We note that the panel's terms of reference have been slightly revised. We re-iterate a point made in our handout that every canonical multiproxy study is affected by problems of either data availability or lack of accurate and clear methodological descriptions (or both), and that authors, journals and funding agencies have, in general, been exceedingly uncooperative towards efforts to obtain materials needed to replicate results.

   In addition to deficient data archiving and data citation by multiproxy authors, there are many serious defects in data archiving by primary collectors, whose results are relied upon in multiproxy studies. Without making any attempt to survey these deficiencies, we report that much tree ring measurement data collected by Jacoby and d'Arrigo has not been archived, including measurements taken nearly 20 years ago, and, in some cases, results have been selectively archived. Similarly, much ice core data collected by Lonnie Thompson has not been archived or has been archived incompletely, including results that are nearly 20 years old. In the latter case, unreconciled and inconsistent grey versions have entered into multiproxy studies.

2. Subsequent to the panel, Wahl and Ammann issued a revised version of their submission to Climatic Change, which responded, in part, to our criticisms of their refusal to disclose key verifications statistics. Table 1S of the revised version now includes r^2 and CE verification statistics for all MBH98 reconstruction steps. Table 1S (reproduced below) confirms the claims of McIntyre and McKitrick [2005a, 2005b] that the 15th century MBH98 reconstruction fails verification r^2 and CE tests, and extends this finding to subsequent steps. The values reported here are virtually identical to those reported in Table 1 of McIntyre and McKitrick [NAS Panel 2006]. This is the paper cited approvingly by Mann in his presentation.

Table 1S Pearson's r^2 and CE Scores for MBH Reconstruction Emulations

Proxy Network MBH – periods	NH Mean r^2 Calibration-period	NH Mean r^2 Verification-period	NH Mean CE Verification-period
1400-1449	0.414	0.018	-0.215
1450-1499	0.483	0.010	-0.314
1500-1599	0.487	0.006	-0.253
1600-1699	0.643	0.004	-0.259
1700-1729	0.688	0.00003	-0.161
1730-1749	0.691	0.013	-0.063
1750-1759	0.714	0.156	0.077
1760-1779	0.734	0.050	-0.070
1780-1799	0.750	0.122	0.040
1800-1819	0.752	0.154	0.069
1820-1980	0.759	0.189	0.103

From Wahl and Ammann [*Climatic Change* 2006], Table 1S.

3. One of the panellists asked Mann for the value of his verification r^2 statistic for the 15th century step. Mann said that they did not calculate this statistic. This was untrue, based on the text of MBH98, as reported in McIntyre and McKitrick [NAS Panel 2006]. In addition, McIntyre showed that MBH98 source code calculated the verification r^2 statistic at exactly the same time as the RE statistic (see http://www.climateaudit.org/?p=273 and more generally http://www.climateaudit.org/index.php?cat=12). The r^2 values was one of the issues raised by Barton and was the one that prompted the President of NAS to propose that an NAS panel be convened. We do not understand why the Panel accepted without challenge Mann's claim that he did not calculate the r^2.

4. In the same answer, Mann said that calculation of the verification r^2 statistic "would be silly, and incorrect reasoning." The Panel also allowed this remarkable claim to go unchallenged. We are unaware of any independent authority for this assertion. There are numerous reasons why the panel should reject this assertion. First, Mann said that they considered this statistic, and the IPCC TAR said that the reconstruction showed skill in more than one statistic. Second, in other articles, Mann considers the verification r^2 statistic when it appears significant (see references in MM, NAS Panel). Third, the RE statistic has no known distribution; as we showed in McIntyre and McKitrick [2005a, 2005d], "rules of thumb" for RE significance derived from linear regression do not necessarily carry over to the MBH98 multivariate methods (a point agreed to by Wahl and Ammann [2006]), hence sole reliance on the RE statistic is weak methodology. Fourth, as we pointed out in McIntyre and McKitrick [NAS Panel], citing Bürger and Cubasch [2005], one cannot use the RE statistic both to select models and to test overfitting, since the model selection inference is based on a maintained hypothesis that the models are not overfit. Fifth, if the RE significance benchmarks of McIntyre and McKitrick [2005a, 2005c] are applied, then many seemingly significant reconstructions under "rules of thumb" RE standards do not achieve statistical significance, consistent with verification r^2 results.

5. The Ammann and Wahl submission to GRL, which attempted to demonstrate that they could establish a 99% RE benchmark of 0.0 on alternate grounds, was rejected shortly after the panel meeting. This leaves the 99% benchmark of 0.51 reported in McIntyre and McKitrick [2005d] as the most recent peer-reviewed consideration of RE benchmarks in an MBH context. Wahl and Ammann [2006] relied on the proposed RE benchmark from their rejected paper. Consequently, none of the tests for RE statistical significance in Wahl and Ammann [2006] are valid.

6. There is a formal relationship between the RE statistic and the verification r^2 discussed in Wilks 1995, relying on Murphy, 1988 (called "Skill Score"), both cited in McIntyre and McKitrick 2005a, in which the RE statistic in a stationary system (as required for regression) is necessarily less than the verification r^2 statistic. In these references, the RE statistic is considered only after the model is shown to pass a verification r^2 test. Spurious RE statistics can arise in non-stationary cases.

   A high r^2 score is *necessary* but not *sufficient* for evaluating model skill, whereas a low r^2 score is *sufficient* to reject a model. The same is true of the RE score. Neither statistic can validate a model that the other statistic has rejected.

7. Mann cited Wahl and Ammann [2006] as demonstrating that our "reconstruction" lacked "statistical or climatological merit". We re-iterate that we did not "present" an alternative reconstruction as a positive interpretation of climate history, but merely as a demonstration of specific sensitivity of MBH results to variations in PC methodology and to the

presence/absence of bristlecones. What Wahl and Ammann (and Mann) characterize as "our" reconstruction is merely an MBH98-type reconstruction with lesser weights on bristlecones (or without bristlecones). We **agree** with Wahl and Ammann (and Mann) that such a reconstruction lacks "statistical merit". However, that only proves the non-robustness of MBH results, which implies, in turn, either that all the proxies except bristlecones are no good, or the MBH98 method is no good, or both. Our contention is that the reconstruction **with** bristlecones is also no good, as evidenced by the failure of verification r^2 and CE statistics.

8. Mann claimed that he had properly adjusted bristlecone data for CO_2 fertilization. In our presentation, (Figure 11), we showed that MBH99 did not make **any** adjustment to MBH98 values – a point inconsistent with Mann's statements to the panel. Since the MBH98 results were carried over into MBH99, any adjustment in MBH99 is only to pre-1400 values. Even the adjustment in MBH99 for values prior to 1400 is not carried out in accordance with the description of the phenomenon contained in Graybill and Idso [1993], but merely on the basis of supposed similarities of low-frequency curves. As we observed in our presentation to the NAS panel, Biondi et al [1999] had also observed similar low-frequency similarities to a different series and "adjustment" to their series would have led to different results. The real issue is that any robust reconstruction should not need to rely on individual proxies whose supposed validity relies on ad hoc adjustments.

9. In response to a question about whether the Divergence Problem evidenced possible saturation of proxies and a potential inability to record high late 20th century temperatures, Mann showed several series (presumably from Osborn and Briffa, 2006, one of which was almost certainly the Yamal series), showing high late 20th century levels, from which he concluded that the proxies had not yet shown signs of saturation or non-linearity. The series in Osborn and Briffa [2006] were not selected at random from the population of all proxies. Over a large population (387) of temperature-sensitive sites, there has been a decline in ring widths and densities in the late 20th century. Within the population of 387 sites, individual chronologies can be found which have late 20th century upticks. Not every tree ring series declines. Selection of a couple of series with upticks is not a thorough response to the Divergence Problem. (For additional references, see Supplementary Bibliography.)

10. We did not attempt to consider every multiproxy study within the confines of our presentation. None of the other studies has been subjected to proper critical analysis. In particular, we also have serious concerns about last year's reconstruction of Moberg et al [2005]. It was been cited as being free of tree ring problems, but it is not free of other problems. Until last month, a complete data set was not available. As a result of a Materials Complaint that we made to Nature, one Corrigendum has already been issued by Moberg et al. [Moberg et al, 2006].

    Moberg's low-frequency roster is a very small set (**only 11** proxy series). Moberg provided no explicit selection criteria and not all proxies have been calibrated to temperature (an increased presence of coldwater diatoms offshore Oman seems like a curious index for higher temperatures – see http://www.climateaudit.org/?p=93). The highest MWP proxy index values are just slightly lower than modern proxy index values. In the absence of explicit selection criteria, one cannot help but wonder whether the selection has been tailored either consciously or unconsciously to achieve that result.

    Moberg's averaging methods require normal distributions. However, although Nature's statistical policy requires evidence of normality, this policy does not appear to have been adhered to, as Moberg data sets showed marked evidence of non-normality with the data sets contributing most strongly to 20th-medieval differences also having the most marked non-normal distributions. (see bottom of http://www.climateaudit.org/?p=346 and more generally http://www.climateaudit.org/index.php?cat=6)

11. We are concerned that the committee may have been over-influenced by certain selections of proxies which show a 20th century "ramp", as evidence of an overall pattern within proxy populations, and thereby conclude, that, even if the statistical methods of the various multiproxy studies were inadequate, no other candidates are on the horizon. For example, at one point in the presentations, one of the Panellists mentioned that he was impressed by the seeming consistency of proxy series that were being shown to him. However, the Panel should realize that they were not randomly selected. It is beyond the scope of a short reply to provide a comprehensive survey demonstrating this, but we will cite two recent studies showing a very different result from the Hockey Stick multiproxy studies.

    Millar et al [2006] is a highly detailed and sophisticated study of the MWP in the Sierra Nevadas, California, using ecological niche modelling, at sites very close to the bristlecone and foxtail sites (the ring width chronologies from which appear to indicate a very "cold" MWP). Millar et al state:

    > The ecologic patterns and climatic estimates at Whitewing and San Joaquin Ridge corroborate studies showing significant Medieval warmth in the California region ...**we modeled paleoclimate during the time of sympatry [the MWP] to be significantly warmer (+3.2 °C annual minimum temperature) and slightly drier (-24 mm annual precipitation) than present…** [emphasis added]

    Naurzbaev et al. [2004] (which includes MBH coauthor Hughes) also applied a sophisticated approach in which they collected samples along latitudinal and altitudinal transects and applied these results to the consideration of subfossil medieval trees. Their objective was "low frequency" information rather than the annual information of the tree ring "site chronologies" They reported:

    > Trees that lived at the upper (elevational) tree limit during the so-called Medieval Warm Epoch (from A.D. 900 to 1200) show annual and summer temperature **warmer by 1.58 and 2.3 deg C, respectively**, approximately one standard deviation of modern temperature. Note that these trees grew **150–200 m higher (1–1.28C cooler)** than those at low elevation but the same latitude, implying that this may be an underestimate of the actual temperature difference. [emphasis added. MM note – the combined effect is 2.58-3.58 deg C., which is a similar result to Millar et al 2006]

    We also refer the panel to literature on treeline changes, where higher medieval treelines have been reported all over the world. Some examples are California [Lloyd and Graumlich, 1997]; Siberia [Shiyatov, 1995]; Fennoscandia [Hiller et al,2000]. See Graumlich [1994] and Bray , 1971 for other examples.

12. It is a trivial exercise to collect proxies showing elevated MWP values, calculate their averages in some sense and thereby obtain a reconstruction with a higher MWP than modern period (consistent with Naurzbaev et al [2004] and Millar et al [2006]), as shown below:

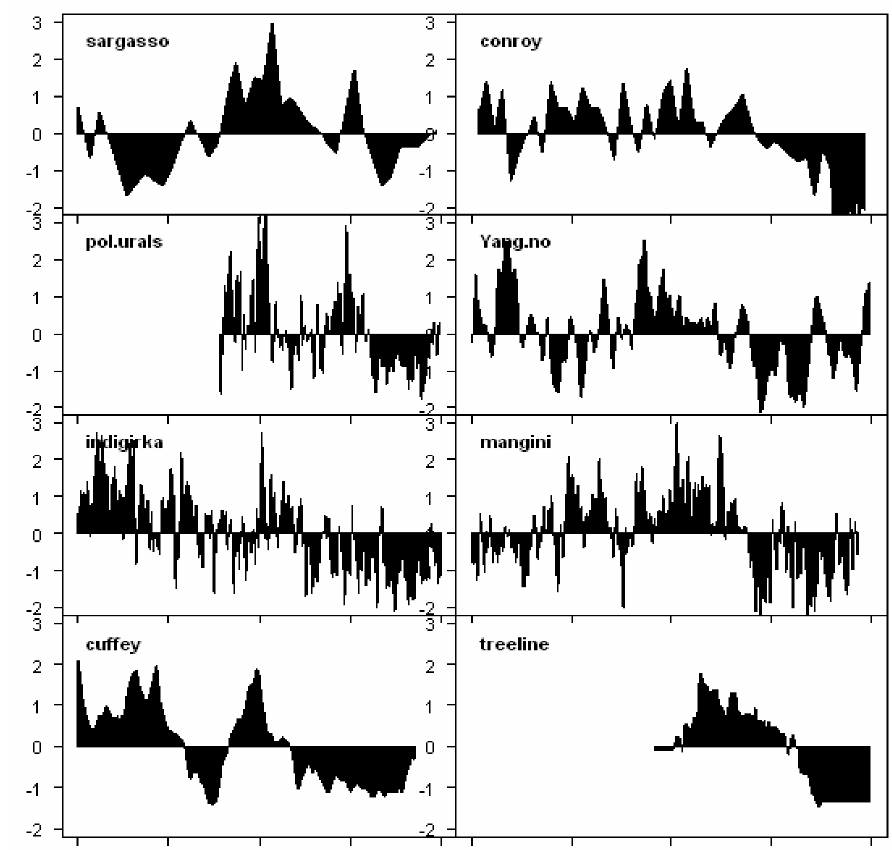

Figure 1. Top left down; then top right down. (1) Sargasso Sea [Keigwin, 1996], used in Crowley and Lowery 2000, Moberg et al 2005; (2) Updated Polar Urals ring widths [Esper et al, 2002]; (3) Indigirka ring widths [Sidorova and Naurzbaev 2002, high-frequency only used in Moberg et al, 2005]; (4) Greenland temperature reconstruction [Alley, 2000 using Cuffey and Clow, 1997]; (5) Conroy Lake sediments [Gajewski, 1998, used in Moberg et al 2005]; (6) Chinese composite of Yang et al [2002] without Thompson's Dunde and Guliya $\delta O18$ series; (7) Speleothem temperature reconstruction [Mangini et al, 2005]; (8) Treeline at Polar Urals [digitised from Briffa et al, 1996].

If these series are standardized and averaged, according to a common multiproxy procedure, and then fitted to NH temperature, one obtains a "reconstruction" with an elevated MWP and a "blade", but which leaves a very different overall impression than the multiproxy studies that have received so much recent attention.

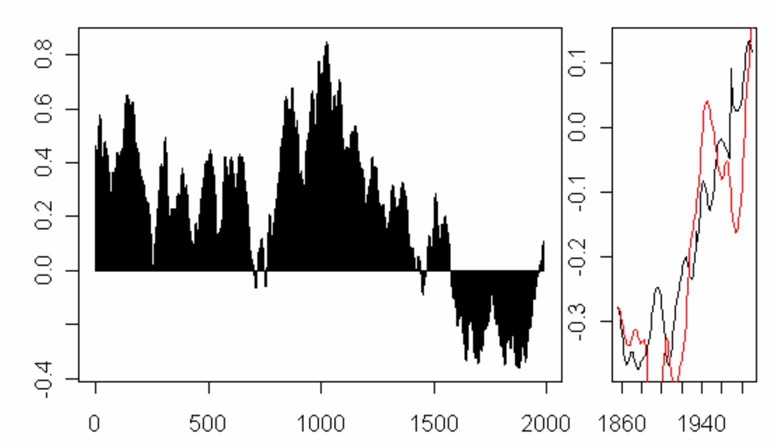

Figure 2. "Apple-picking " Reconstruction. Left – average of 8 proxy series in Figure 1; right – comparison to instrumental CRU NH temperature average (1961-1990 basis) – note change in scale from left to right. Both panels in "deg C".

We do not put this forward as an "alternate" reconstruction, but merely to illustrate the different impact of "apple picking" to make apple pie, rather than cherry picking to make cherry pie.

By not responding to all other points raised by Mann and other presenters, it should not be implied that we necessarily agree, as we have selected only a few matters for this letter.

Yours truly,

Stephen McIntyre

Ross McKitrick

DATA CITATIONS:

Sargasso Sea: ftp://ftp.ngdc.noaa.gov/paleo/contributions_by_author/keigwin1996/fig4bdata column 2, linear interpolations to annual basis

Updated Urals: Email from Brooks Hanson, deputy editor of Science, Feb. 2006, responding to request regarding Esper et al [2002]

Indigirka: Email from Anders Moberg, Feb. 23, 2006, after Materials Complaint to Science and Corrigendum by Moberg et al [2006], providing unpublished data from Sidorova and Naurzbaev [2002]

Greenland: ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/greenland/summit/gisp2/isotopes/gisp2_temp_accum_alley2000.txt

Conroy Lake: ftp://ftp.ngdc.noaa.gov/paleo/pollen/recons/liadata.txt

Variation on Yang composite: Calculated from data supplied by email by Bao Yang.

Mangini: ftp://ftp.ncdc.noaa.gov/pub/data/paleo/speleothem/europe/austria/spannagel2005.txt

PolarUral treeline: Digitized from Figure xx of Briffa et al [1996]

REFERENCES

Alley, R.B. 2000. The Younger Dryas cold interval as viewed from central Greenland. Qu*aternary Science Reviews* 19, 213-226.

Ammann, C. and E. Wahl (2006), Comment on ‘‘Hockey sticks, principal components, and spurious significance’’ by S. McIntyre and R. McKitrick, *rejected*.

Biondi, F., D.L. Perkins, D.R. Cayan and M.K. Hughes, 1999. July temperature during the second millennium reconstructed from Idaho tree rings. *GRL* 26, 1445-1448.

Bray, J.R., 1971. Vegetational distribution, tree growth and crop success in relation to recent climatic change. *Adv. Ecol. Res.* 7: 177-233.

Briffa, K.R., Jones, P.D., Schweingruber, F.H., Karlen, W. and Shiyatov, S.G., 1996. Tree-ring variables as proxy-climate indicators: problems with low-frequency signals. In: Climatic Variations and Forcing Mechanisms of the Last 2000 Years (Eds. P.D. Jones, S. Bradley and J. Jouzel). NATO ASI Series Vol. 141, 9-41.

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

Crowley, T.J. and Lowery, T.S., 2000. How warm was the Medieval warm period? *Ambio* 29, 51-54.

Cuffey, K.M., and G.D. Clow. 1997. Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition. *Journal of Geophysical Research* 102, 26383-26396.

Esper, J., Cook, E.R. and Schweingruber, F.H., 2002. Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* 295: 2250-2253.

Gajewski, K. 1988, Late Holocene climate changes in eastern North America estimated from pollen data. *Quaternary Research* 29:255-262

Graybill, D.A., and S.B. Idso. 1993. Detecting the aerial fertilization effect of atmospheric CO2

Graumlich, L., 1994. Long-term vegetation change in mountain environments. In Mountain Environments in Changing Climates (ed. M. Beniston), pp. 167-179

Hiller A., Boettger T., Kremenetski C., 2001. Mediaeval climatic warming recorded by radiocarbon dated alpine tree-line shift on the Kola Peninsula, Russia**.** *The Holocene*, 11, 491-497.

Keigwin, L.D., 1996. The Little Ice Age and Medieval Warm Period in the Sargasso Sea, *Science*, 274, 1504-1508.

Lloyd, AH and LJ Graumlich. 1997. Holocene dynamics of treeline forests in the Sierra Nevada. *Ecology* 78, 1199-1210.

Mangini, A., C. Spötl, and P. Verdes. 2005. Reconstruction of temperature in the Central Alps during the past 2000 yr from a d18O stalagmite record. *Earth and Planetary Science Letters*, 235, 741-751.

McIntyre, S., and R. McKitrick, 2005a. Hockey sticks, principal components, and spurious significance, *Geophys. Res. Lett.,* 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 Von Storch and Zorita, *GRL*, 32, L20713, doi:10.1029/2005GL023089.

McIntyre, S., and R. McKitrick 2005d, Reply to Comment by Huybers, *GRL*, 32, L20713, doi:10.1029/2005GL023586.

McIntyre, S., and R. McKitrick. 2006. Presentation to the Committee on Surface Temperature Reconstructions for the Past 1,000-2,000 Years. Meeting Handout. March 2, 2006.

Moberg, Anders, Dmitry M. Sonechkin, Karin Holmgren, Nina M. Datsenko and Wibjörn Karlén, 2005. Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, 433, 613-617.

Moberg A, Sonechkin DM, Holmgren K, Datsenko NM, Karlén W, Lauritzen S-E, 2006: Corrigendum: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data. *Nature* 439, 1014.

Millar, C.I., J.C. King, R.D. Westfall, H.A. Alden, and D.L. Delany. 2006. Late Holocene forest dynamics, volcanism, and climate change at Whitewing Mountain and San Joaquin Ridge, Mono County, Sierra Nevada, CA, USA. *Quaternary Research*. In Press. (Downloaded from http://www.fs.fed.us/psw/programs/snrc/staff/millar/Whitewing_txt.pdf).

Murphy, Allan H. (1988), Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient, *Monthly Weather Review* 116, 2417-2424.

Naurzbaev, Mukhtar M., Malcolm K. Hughes and Eugene A. Vaganov, 2004. Tree-ring growth curves as sources of climatic information, *Quaternary Research* 62, 126– 133.

Osborn, Timothy J. and Keith R. Briffa, 2006, The Spatial Extent of 20th-Century Warmth in the Context of the Past 1200 Years, *Science* 311, 831-834.

Shiyatov, S.G., 1995. Reconstruction of climate and the upper treeline dynamics, *Publications of the Academy of Finland* 6/95, 144-147.

Sidorova OV, Naurzbaev MM 2002: Response of /Larix cajanderi/ to climatic changes at the Upper Timberline and in the Indigirka River Valley, Lesovedenie 2, 73-75, in Russian),

Wahl, Eugene R. and Caspar M. Ammann, 2006 (under review). Robustness of the Mann, Bradley, Hughes Reconstruction of Surface Temperatures: Examination of Criticisms Based on the Nature and Processing of Proxy Climate Evidence.

Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences: an Introduction*, International Geophysics Series, **Vol. 59**, Academic Press, 464 pp.

Yang Bao, Achim Braeuning, Kathleen R. Johnson and Yafeng Shi, 2002, General characteristics of temperature variation in China during the last two millennia. *GRL* 10.1029/2001GL014485.