ROBUSTNESS OF THE MANN, BRADLEY, HUGHES RECONSTRUCTION OF SURFACE TEMPERATURES: EXAMINATION OF CRITICISMS BASED ON THE NATURE AND PROCESSING OF PROXY CLIMATE EVIDENCE

EUGENE R. WAHL[1] and CASPAR M. AMMANN[2]

[1] Environmental Studies and Geology Division, Alfred University, Alfred, New York, U.S.A.
One Saxon Dr., Alfred, NY 14802 wahle@alfred.edu 607.871.2604 607.871.2697 (fax)
[2] National Center for Atmospheric Research*, Boulder, Colorado, U.S.A.
The authors contributed equally to the development of the research presented.

Abstract. The Mann et al. (1998) Northern Hemisphere annual temperature reconstruction over 1400-1980 is examined in light of recent criticisms concerning the nature and processing of climate proxy-data used in the reconstruction. A systematic sequence of analyses is presented to examine issues concerning the proxy evidence, utilizing both indirect analysis via exclusion of proxies and processing steps subject to criticism, and direct analysis of principle component (PC) processing methods in question. Altogether new reconstructions over 1400-1980 are developed in both the indirect and direct analyses, which demonstrate that the Mann et al. reconstruction is highly robust against the proxy-based criticisms addressed. In particular, reconstructed temperatures are demonstrated to be virtually unaffected by the use or non-use of PCs to summarize proxy evidence from the data-rich North American region (both in terms of the time period used to "center" the proxy data before PC calculation and the way the PC calculations are done), as long as the full information in the proxy data is represented by the PC time series. Clear convergence of the resulting climate reconstructions is a strong indicator for this criterion.

* The National Center for Atmospheric Research is sponsored by the National Science Foundation, USA.

Also, recent "corrections" to the Mann et al. reconstruction that suggest 15th century temperatures could have been as high as those of the late-20th century are shown to be without statistical and climatological merit. Our examination does suggest that a slight modification to the original Mann et al. reconstruction is appropriate for the early 15th century (~ +0.05°), which leaves entirely unaltered the primary conclusion of Mann et al. (as well as many other reconstructions) that both the 20th century upward trend and high late-20th century hemispheric surface temperatures are anomalous over at least the last 600 years. Our new reconstructions are also used to evaluate the separate criticism of reduced downward magnitude in the Mann et al. reconstructions over significant portions of 1400-1900, suggesting that from the perspective of the proxy data themselves, such losses may be smaller than those reported in other recent work.

1. Introduction

The Northern Hemisphere mean annual temperature reconstruction of Mann, Bradley, and Hughes (Mann et al., 1998, 1999; hereafter referred to as "MBH98" and “MBH99” respectively, or, when referring to both articles, as "MBH”) is one of a growing set of high-resolution (annually resolved) reconstructions of global/hemispheric surface temperatures that cover all or significant portions of the last millennium (e.g., Rutherford et al., in press; Moberg et al., 2005; Cook et al., 2004; Huang, 2004; Mann and Jones, 2003; Crowley et al., 2003; Esper et al., 2002; Briffa et al., 2001; Crowley and Lowery, 2000; Jones et al., 1998; Bradley and Jones, 1993; see also Jones and Mann, 2004). The MBH reconstruction was the first to assimilate multiple kinds and lengths of climate proxy data sets into eigenvector-based multivariate climate field reconstruction techniques (cf. Kaplan et al., 1997; Cane et al., 1997; Evans et al., 2002) of global/hemispheric temperatures. Although it quite closely resembles previous (Bradley and

Jones, 1993) and also more recent reconstructions (cf. Rutherford et al., in press; Mann and Jones, 2003), its "hockey stick"-shaped result of a slow cooling trend over the past millennium (the "stick") followed by anomalous 20th century warming (the "blade") has been widely cited and featured in international assessments of global climate change, most prominently in the Intergovernmental Panel on Climate Change Third Assessment Report (Folland et al., 2001). As an important scientific work, among others, with notable policy implications, the MBH reconstruction has been subjected to significant scrutiny, which can be categorized into three areas.

First, the MBH reconstruction has been examined in light of its congruence/lack of congruence with other long-term annual and combined high/low frequency reconstructions. In particular, the amplitude of the Northern Hemispheric mean surface temperatures in MBH is significantly less at some times during the millennium than the amplitude of some other long-term reconstructions (notably, Moberg et al., 2005; Esper et al., 2002; Briffa et al., 2001; Harris and Chapman, 2001; Huang et al., 2000). New work by Rutherford et al. (in press) suggests that the extra amplitude of the tree ring-based reconstruction of Esper et al. (2002; cf. Cook et al., 2004) can be explained as the expectable result of a significantly restricted sampling (14 extra-tropical, continental tree-ring sites) of the spatially variable hemispheric temperature field. In contrast, the MBH99 reconstruction is calibrated using data from a minimum of 36 sites over AD 1000-1399 [12 actual predictands (employing principle component--or PC--summaries of dense proxy networks), including southern tropical and northern high-latitude ice core data along with temperate- and high-latitude tree ring data, with at least one proxy site on each of five continents] and a maximum of 415 proxy and instrumental records over 1820-1980 [112 actual predictands, with greater spatial coverage on all continents except Africa, and a small number of tropical oceanic sites that include coral proxies] (MBH). The comparison of the MBH reconstruction, derived from multi-proxy (particularly tree ring) data sources, with widespread bore-hole-based reconstructions (Harris and Chapman, 2001; Huang et al., 2000) is still at issue in the literature (Chapman et al., 2004; Schmidt and Mann, 2004; Mann and Schmidt, 2003; Mann et al., 2003; Rutherford and Mann, 2004), as is the comparison with the high/low frequency composite reconstruction of Moberg et al. (2005) (Mann, M.E., S. Rutherford, P.D. Jones, C. Ammann, E. Wahl, T. Osborn, K. Briffa, 'Comment on "Highly variable Northern Hemisphere temperatures reconstructed from low-and high-resolution proxy data" by A. Moberg et al.', in review with *Nature*).

Second, a related area of scrutiny of the MBH reconstruction technique arises from an AOGCM-based study (von Storch et al., 2004) that also examines the potential for loss of amplitude in the MBH method (and other proxy/instrumental reconstructions that also calibrate by using least squares projections of the proxy vectors onto a single- or multi-dimensional surface determined by either the instrumental data or its eigenvectors). This kind of analysis using the virtual climates of AOGCMs allows the long-term quality of climate reconstructions to be assessed directly, since the model-based reconstruction can be compared to the "real" model temperatures (known everywhere in the 3D-domain). In the real world, such a comparison is only possible, at best, over the recent ~ 150 years and over a restricted spatial domain. However, a number of issues specific to the modelling situation could arise in this context, including: how realistically the AOGCM is able to reproduce the real world far-field patterns of variability; the magnitude of forcings and the sensitivity of the model system that determine the magnitude of temperature deviations (see Shindell et al., 2003, 2001; Waple et al., 2002; MBH99); the possible presence of non-realistic drift in the model output (Goosse et al, 2005); and the extent to which the model is sampled with the same richness of information that is contained in proxy records (not only temperature records, but series that correlate well with the primary patterns of variability--including, for example, precipitation in particular seasons). Thus, the extent to which the results of von Storch et al. (2004), which suggest a large under-expression of amplitude by the MBH method, are true to the real-world application of this method is not well established. Examinations of these modelling-specific considerations have been undertaken by Mann et al. (in press) and will be examined elsewhere (Ammann, C.M., E.R. Wahl, and C. Tebaldi, 'Natural climate forcing magnitude and low-frequency bias of regression-based climate field reconstructions', in preparation).

A third area of scrutiny has focused on the nature of the proxy data set utilized by MBH, along with the characteristics of proxy pre-processing algorithms used to enhance the climate signal-to-noise characteristics of the proxy data (McIntyre and McKitrick, 2003, 2005a, b; hereafter referred to as "MM03", "MM05a", and "MM05b" respectively, or as "MM" jointly). In MM03, a reconstruction for Northern Hemisphere mean surface temperature from 1400-1980 is presented that is clearly inconsistent with the hockey-stick result of anomalous 20th century warming, showing not only relatively high temperatures in the 20th century, but also sustained temperatures in the 15th century that are *higher* than any sustained temperatures in the 20th century--descriptively, a “double-bladed hockey stick” (i.e., an upward trend on either end of the reconstruction period). In MM03, the authors describe this result as being developed using the MBH methodology, albeit with elimination of a large number of the proxy data series used by MBH, especially during the 15th century. In MM05b, a second version of the MM reconstruction is presented that the authors describe as virtually identical to the one presented in MM03. The version in MM05b is based on the substitution of a new version of a single tree ring series from northeastern Canada's St. Anne Peninsula (the "Gaspé" series) and on newly-computed PC summary series for North American tree ring data in MBH that are derived from the International Tree Ring Data Base (ITRDB), discussed below. Although the authors state that the MM03 and MM05b reconstructions are nearly identical, the high excursions of 15th century temperatures in MM03 are clearly larger than those in MM05b (by ~ 0.20°) and are much more clearly differentiated from late 20th century temperatures, especially in relation to instrumental temperatures in the 1990's. Thus, the most serious departure from the single-bladed hockey stick depiction of highly anomalous temperatures in the later 20th century is presented in MM03. Associated validation statistics for these reconstructions are not given in either MM03 or MM05b, and thus it is not possible to gauge directly from the published articles the climatological meaningfulness of the time series and thereby evaluate the authors' strong criticism of the single-bladed hockey stick result.

Neither of the first two areas of scrutiny (congruence with other reconstructions and potential loss of least squares-based reconstruction amplitude in model virtual worlds) have fundamentally challenged the MBH conclusion of an anomalous rise (in terms of both duration and magnitude) in temperatures during the late 20th century (Jones and Mann, 2004; von Storch et al., 2004; Moberg et al., 2005). The conclusions emphasized by MM, however, warrant a more complete validation because they question the fundamental structure of climate evolution over time, and thus the corresponding interpretation of cause and effect (e.g., Crowley, 2000, MBH98), over much of the last millennium. In particular, MM present a unique climate reconstruction which implies that the early 15th century experienced the warmest sustained surfaces temperatures in the Northern Hemisphere in nearly 2000 years. In this paper, we examine the proxy data-based criticisms of the MBH method and results by MM, based on our own independent re-creation and use of the MBH method to reconstruct Northern Hemisphere mean surface temperature over 1400-1980. In addition, we also address aspects of logic in the arguments presented in MM05a and MM05b and the absence of important validation statistics in MM03 and MM05b, since these issues have bearing on the efficacy of the MM results and have not been addressed in the scientific literature. We also use the results we develop to briefly address the issue of loss of amplitude, from the perspective of the information contained in the proxy data themselves.

1.1 DETAILS OF MM CRITICISMS

MM have focused on two primary aspects of the MBH reconstruction: the methodological employment of proxy PC summaries (MM05a/b), as well as the use of proxy records deemed unjustifiable (MM03 and MM05b). In addition to the criticism of specific proxy records in MM03, in MM05a the authors strongly criticize the method applied by MBH to generate PC summaries of spatially-dense proxy networks in North America as an antecedent to the actual calibration of the proxies to the instrumental climate field, which is a common technique in high-resolution paleoclimatology, derived from the field of dendroclimatology (Fritts, 1976). In essence, this technique attempts both to downweight the impact of data-dense regions on statistical calibrations between proxies and instrumental values and to lower the potential impact of calibration "overfitting". In paleoclimatology, the latter situation occurs when a large number of proxy regressors in a least squares projection improves the fit between proxies and climate values during the calibration period, but in the process this fit becomes so specific to the calibration period that it does not work well when applied in an independent "verification" period. The use of the first few PCs of data-dense regional proxy networks helps alleviate this problem, and also aids in isolating the climate signal in the proxy variations from non-climate noise.

Relative to MBH, in MM05a the authors question the period over which the proxy values are transformed into standardized anomalies before the PC extraction procedure is performed, along with the method of transformation; and claim that the MBH method as originally applied (standardization and transformation over the calibration period, 1902-1980, before PC extraction) leads to proxy PC summary series that inappropriately weight for hockey stick bladed-shapes for PC1 (the leading component of variance) in the 20th century. [The temporal information captured by PC1 of the North American tree ring network is a crucial source of information in calibration for the 11th-14th century reconstruction (MBH99), and in verification for the early 15th century reconstruction in MBH98, as shown here.] A further aspect of this critique is that the single-bladed hockey stick shape in the 20th century in proxy PC summaries for North America is carried disproportionately by a relatively small number (15) of proxy records derived from bristlecone/foxtail pines in the western United States, which the authors mention as being subject to question as local/regional temperature proxies after approximately 1850 (cf. MM05a/b; Hughes and Funkhauser, 2003; MBH99; Graybill and Idso, 1993). It is important to note in this context that MBH do not claim that all proxies used in their reconstruction are closely related to local-site variations in surface temperature, rather they invoke a less restrictive assumption that "whatever combination of local meteorological variables influence the proxy record, they find expression in one or more of the largest-scale patterns of annual climate variability" against which the proxy records are calibrated in the reconstruction process (Mann et al., 2000). For example, tree ring records in semi-arid areas could represent precipitation modified by large-scale climate patterns rather than temperature. Nevertheless, the salient implication of these critiques in MM05a/b is that inappropriate variance is carried into the calibration process by proxy summary records that exhibit a single-bladed shape in the 20th century, and thus these summary data series will be given excessively high weight in the calibration process because eigenvector1/PC1 of the instrumental temperature field is strongly associated with the upward trend in 20th century instrumental temperatures. This relatively high weighting will then be carried by these proxies throughout the reconstruction period, affecting the overall shape of the reconstruction set throughout the 14th-20th centuries. [In the logically extreme case, such a biased weight given to an artificial signal that is not part of the proxy records themselves could introduce a climate reconstruction that is not supported by the original, underlying data.]

Finally, in MM05a, the authors examine appropriate threshold values for significance for the Reduction of Error (RE) statistic, which is commonly used as a validation statistic in high-resolution paleoclimatology (Cook et al., 1994). They discuss this issue in the context of their argument concerning the appropriate method for deriving PC summary series from proxy data.

2. Methods

2.1 REPLICATION OF MBH98 METHODOLOGY

Before directly addressing the issues related to the MM criticisms, we first introduce our own reproduction of the MBH algorithm for reconstructing climate. It is important to note that in this replication we retain the primary MBH assumptions concerning the stationarity of the relationship between proxy climate indicators and eigenvector spatial patterns of temperature anomalies and the relative temporal stability of the large-scale patterns themselves. Implementation of the MBH method was based on published descriptions (MBH; Mann et al,

2000) and on additional material (clarifications and data) available via the "ftp" site of M.E. Mann at the University of Virginia (ftp://holocene.evsc.virginia.edu/pub/sdr/temp/nature/MANNETAL98/). Dr. Mann also answered clarifying questions concerning a few details of the method. The core calibration and reconstruction components of the MBH method were first translated into matrix algebra formulae. The entire algorithm was then coded in the publicly available "R" software (R Development Core Team, 2004), which has particular advantages for implementing matrix algebra equations. The calibration component was simplified to have a consistent annual time step throughout, rather than the temporally heterogeneous method used in MBH of calculating PCs of the instrumental data based on a monthly time step and subsequently forming annualized PC series from averages of the monthly values (MBH). Although MBH98 state that the extra degrees of freedom generated by using the monthly time step are needed for eigenvector/PC decomposition of the global instrumental data field (which has 1082 components with information on a 5°x5° grid; Jones and Briffa, 1992), this is in fact not a requirement statistically (cf. Zorita et al., 2003). A second simplification eliminated the differential weights assigned to individual proxy series in MBH, in light of testing which showed that the results are insensitive to this implementation. The MBH technique of specifying 11 separate calibrations between the proxy data and specific numbers of retained PCs from the global instrumental data grid (for different sets of proxy data richness going back in time) was followed exactly; thus, specific calibrations were used for each of the following periods: 1400-1449, 1450-1499, 1500-1599, 1600-1699, 1700-1729, 1730-1749, 1750-1759, 1760-1779, 1780-1799, 1800-1819, and 1820-1980. The standard calibration period of 1902-1980 and independent verification period of 1854-1901 also followed MBH exactly. For the purposes of testing our re-creation of the MBH algorithm, the instrumental and proxy data used were the same as those employed in the original MBH reconstructions. The full code and data series are available online (see link below).

2.2 RECONSTRUCTION SCENARIOS REFLECTING MM CRITICISMS

After testing verified that our algorithm could re-create the original MBH results (Fig.1), we developed six reconstruction scenarios that examine the different criticisms of the MBH approach made by MM. Scenario 1 examines the MM03 climate reconstruction. Scenarios 2-4 examine the influence of the North American tree ring data, with inclusion/exclusion of the bristlecone/foxtail pine records, by utilizing the proxies *individually* rather than through employing PC summaries. Scenarios 5 and 6 examine the influence resulting from various approaches to generating PC summaries. Scenario 6 also examines the MM05b climate reconstruction.

1) Examination of MM03 Northern Hemisphere temperature reconstruction

Reconstruction was done over 1400-1499 by exclusion of all the North American proxy data from the ITRDB along with additional series from the North American Southwest and North American tree line, with calibration over the standard period. This scenario mimics the reconstruction scenario developed for this period in MM03 (cf. Rutherford et al., in press). It allows examination of the statistical validity of the MM03 result that Northern Hemisphere surface temperatures in the 15th century were higher than even later-20th century temperatures. Failure to pass validation tests for either the calibration or verification periods would indicate that the MM03 double-bladed hockey stick reconstruction for the 15th century cannot be ascribed climatological meaning; passing validation tests for both calibration and verification would indicate climatological meaning for the reconstruction.

2) Replacement of North American ITRDB PC summaries with full proxy network

Reconstructions were done for different configurations representing the proxy networks of 1404-1449, 1450-1499, 1700-1729 and 1800-1819 (cf. Table 2) by employment of *all* proxy data series available for each period, i.e., *without* the use of second-order PC summarizations of the proxies in data-rich regions of North America and Eurasia and excluding all data for the years 1400-1403. Calibration was restricted to 1902-1971. (The truncations of the reconstruction and calibration periods account for criticism in MM03 concerning infilling of values in a small number of proxy records by persistence from the earliest year backward and the latest year forward). The purpose of evaluating the reconstructions in the different proxy network configurations is to check whether the use of all the proxy data without PC summarizations is sensitive to the richness of the proxy data set, focusing on four reconstruction periods with specific qualities: a) the key period at issue in MM (15th century) and b) the warmest (18th century) and coolest (19th century) pre-20th century reconstructions when the proxy set is close to its full richness (415 series) and includes a few long instrumental records. The number of proxy records successively increase for the four scenarios (95, 139, 378, 405). The reconstructions were not restricted to the decades for which they represent the optimal network, but all extend to 1971, so that any outcome that might cast doubt on the hockey stick-blade result in the 20th century would be included in the analysis.

The Gaspé proxy series highlighted in MM05b as potentially problematic during the first half of the 15th century is retained in this scenario (unlike in scenarios 5 and 6 below). This retention is based on "individual proxy" tests over the entire 15th century that yielded highly similar reconstructions whether this series is included over 1404-1449 or not, and whether it occurs once or is replicated over 1450-1499 [occurring both as an individual record and as part of the ITRDB set used to calculate North American PC summaries--as in MBH98 (cf. MM05b)]. The Twisted Tree/Heartrot Hill boreal treeline proxy highlighted as potentially problematic in the early part of the reconstruction period by MM03 (cf. Fig. 4 there and associated text) is not included in the proxy set used in this scenario. These two criteria make the individual proxy set used in this scenario identical to the one that underlies our re-creation of the original MBH98 reconstruction.

Scenario 2 allows examination of the robustness of the climate information in the proxy data set in relation to the use/non-use of PC summaries of the proxies, by determining whether the data *per se* (isolated from any PC pre-processing) carry only a single-bladed hockey stick shape in the 20th century. If no second blade is found in the 15th century, then the issue of whether this overall result is somehow an artifact of the use of second-order PC proxy summarizations becomes essentially moot, as does the corollary presumed need to recalculate the MBH PC summarizations used in the reconstruction (MM05a/b). Such a result would also make moot the criticism in MM05a of the standard benchmark for the RE statistic (indicative of reconstruction skill when >0, indicative of lack of skill when ≤0), which is specific to the context of use of PC summarizations of proxy data.

3) Replacement of PC summaries and elimination of bristlecone/foxtail pine proxy records Reconstruction was done over 1404-1971 using the same structure as scenario 2, but *excluding* the 15 bristlecone/foxtail pine records from western North America identified in MM05a/b as carrying little local/regional temperature information during the calibration and verification periods (although note, a close local proxy-temperature correlation is not strictly required; see above).

Scenario 3 allows examination of the nature of the bristlecone/foxtail pine records as meaningful proxies in relation to the reconstruction of global eigenvector patterns of surface temperature (in contrast to the relationship between these proxies' ring widths and local/regional surface temperatures), and whether their exclusion has discernible impact on Northern Hemisphere mean surface temperature reconstructions, especially in the 15th century. If excluding these proxies degrades reconstruction performance in either the calibration or verification periods, then they do carry empirically meaningful information at the scale of global/hemispheric surface temperature variance patterns and this result would argue for their continued inclusion in the MBH-style method. However, independent of whether or not the bristlecone/foxtail pine records carry meaningful large-scale temperature information, should their exclusion have relatively little impact on the magnitude and trajectory of Northern Hemisphere temperature reconstructions over the last 600 years, then the question of whether they inappropriately lead to the single-bladed hockey stick result in the 20th century (i.e., an inappropriate diminishing of amplitude in the 15th century) ceases to be of significant import.

4) Replacement of PC summaries with strong truncation of proxy richness in North America Reconstruction was done over 1404-1971 using the same structure as scenario 2, but randomly *excluding* 50 out of 60 proxy data series in the Southwestern United States (defined as 100-123° W, 32-42° N) from the ITRDB; along with deterministically *excluding* four spatially repetitive ITRDB proxy series in Northwestern North America (defined as 100-124° W, 42-50° N), three spatially repetitive ITRDB proxy series in the Southeastern United States (defined as 76-92° W, 32-37° N), and four spatially repetitive non-ITRDB proxy series ("Stahle" series, MBH) in the Southwestern United States and Northwestern Mexico. Two repetitions of this scenario were done based on independent random exclusions. The Sheep Mountain bristlecone pine series, which MM05b characterize as the (inappropriately) top-weighted proxy in PC1 of the North American ITRDB series in MBH98, is not included in either selection to guard against such a potential influence. Only two other bristlecone/foxtail series are included in each selection.

Scenario 4 allows an important logical check on the use of scenarios 2 and 3, by determining whether the initial intent of employing PC summations of proxy data from the North American region (to reduce the influence of this spatially restricted, but data-rich region in calibration) is violated by the use of all the individual proxies in calibration. Since this technique is the key method used to examine the robustness of the climate information in the proxy data in relation to the use/non-use of proxy PC summaries, it is crucial that any bias caused by overweighting of the North American proxy data in calibration be identified (if present). If there is little or no difference in the reconstruction results whether the full North American ITRDB *cum* "Stahle" set of proxies is used versus employing strongly truncated subsets of these proxies (~83% reductions for the particularly rich Southwestern data, ~66% reductions for the North American data as a whole), then there is no empirical indication of the presence of spurious overemphasis being given to North America in the full "all proxy" scenarios. Repetition with independent random exclusion for the American Southwestern region allows a further check of the existence/non-existence of an impact. [The division of the truncated proxies into three sets follows the general distribution of direction of ENSO-related teleconnections in the three regions (Cole and Cook, 1998; Ragagopalan et al., 2000). Deterministic exclusions of proxies for the North American Northwest and the United States Southeast were used, rather than random exclusions, due to the small number of spatially-repetitive proxies in each of these regions.]

5) Inclusion of PC summaries with restriction of Gaspé series over 1400-1499

Reconstruction was done over 1400-1499 by *including* PC summaries of the North American proxy data from the ITRDB: replacing the original MBH-calculated PCs with newly calculated PCs based on the MBH data set, and excluding the Gaspé series as an individual proxy over 1400-1449 and as part of the data set used to calculate the North American ITRDB PCs. Calibration was done over 1902-1980. Reconstruction to 1400 and calibration to 1980 (rather than truncating at 1404 and 1971, respectively, as in scenarios 2-4) follows the periods used in MM05b. Three ways of calculating the PC summaries were employed: a) the original MBH reference period for determining standardized anomalies (1902-1980), which were input into the "svd" routine in R (utilizing the singular value decomposition method of PC extraction); b) the MM05a/b reference periods for determining standardized anomalies (1400-1980 and 1450-1980 for the 1400-1449 and 1450-1499 reconstruction periods, respectively), which were input into "svd"; and c) the MM05a/b reference periods for determining anomalies (centered but unstandardized), which were input into the "princomp" routine in R (which extracts PCs using the 'eigen' routine on the variance-covariance matrix of the data [default setting]). Two through five proxy PC series are used with each of these calculation methods, providing 12 reconstructions altogether.

6) Inclusion of PC summaries with restriction of the Gaspé series and elimination of bristlecone/foxtail pine records Reconstruction was done over 1400-1499 using the same structure as scenario 5, with the additional elimination from the North American ITRDB PC calculations of the 15 bristlecone/foxtail pine records eliminated in scenario 3.

Scenarios 5 and 6 allow direct examination of the MM05a/b criticism that a single-bladed hockey stick shape in the 20th century (with no second blade in the 15th century) is an artifact of the use of PC proxy summarizations, especially when done in conjunction with the elimination of the bristlecone/foxtail pine records. This direct examination is the logical counterpart of the indirect examination of this issue by exclusion of PC summarizations in scenarios 2-4. The coupling of indirect and direct tests allows the possibility of drawing highly robust conclusions (when both methods agree), as demonstrated powerfully in paleoenvironmental studies by Prentice et al. (1991) concerning the millennial-scale synchrony of vegetation and climate change over the last glacial-interglacial transition in North America.

A further motivation for scenarios 5 and 6 arises from the recognition of a mathematical inconsistency in the published replication of the MBH North American proxy PC calculations in MM05a/b, in relation to the method used by MM in their re-calculation of the PCs according to the MM centering convention (cf. Supplementary Information website, MM05b). In MM05b, the authors report that they were able to exactly replicate the MBH North American PC series using "svd" with standardized anomalies of the proxy data formed according to the MBH centering convention. In the R code for the re-calculation of these PCs at the MM Supplemental Information website, the method used for this purpose is the "princomp" routine on the same proxy data matrix, formed into anomalies (but not standardized) using the MM centering convention. The effect of using "princomp" without specifying that the calculation be performed on the correlation matrix (an alternate argument of "princomp") forces the routine to extract eigenvectors and PCs on the variance-covariance matrix of the unstandardized proxy data, which by its nature will capture information in the first one or two eigenvectors/PCs that is primarily related to the absolute magnitude of the numerically largest-scaled variables in the data matrix (Ammann, C.M. and E.R. Wahl, 'Comment on “Hockey sticks, principal components, and spurious significance” by S. McIntyre and R. McKitrick', in submission to *Geophysical Research*

Letters). We have demonstrated that this method of PC extraction has the effect of shifting the actual temporal information common to the North American ITRDB proxy data into higher-order PCs, especially the fourth PC (Ammann and Wahl, in submission). When "svd" is used on the standardized data, the temporal structure of the data is captured in the first two PCs, and the MBH/MM centering conventions only have the effect of reversing the order of the PC in which a single-bladed 19th-20th century hockey-stick shape occurs (PC1 with the MBH centering convention and PC2 with the MM centering convention). When the first two PCs are summed (either arithmetically or as vectors) for both centering conventions, the resulting time series are nearly identical in structure, with a small difference in scale over the pre-calibration period (Ammann and Wahl, in submission). Scenarios 5 and 6 allow systematic examination of the effects of these different PC calculation methods on the reconstruction of Northern Hemisphere temperatures over 1400-1499, in the context of the two salient truncations of the 15th century proxy data set proposed in MM05b. The results from scenario 6 further allow examination of the statistical validity of the primary MM05b climate reconstruction, in the same manner that the MM03 reconstruction is examined in scenario 1.

2.3 RETENTION OF INSTRUMENTAL EIGENVECTORS/PC SERIES IN CALIBRATION

The numbers of PCs from the global instrumental data grid against which the various proxy sets were calibrated were held the same as those used in the temporally corresponding calibrations of the main MBH98 re-creation. This methodology is considered appropriate for scenario 1 and the scenarios that employ PC summaries as proxies (5 and 6), since the numbers of proxy series assimilated into calibration are either equal to or similar to those used in the main MBH98 re-creation. For scenarios 2-4, which assimilate larger numbers of individual proxy series into calibration (especially scenarios 2 and 3), the method is justified by experiments in which results for the two 15th century calibrations were replicated by retaining the first four instrumental PCs (rather than the MBH specification of one PC for 1404-1449 and two for 1450-1499). The results of these replications demonstrate robustness to the number of retained instrumental PCs; the calibration and verification statistics are, at most, only slightly different from those of the original experiments (with no change of meaning), and the actual reconstructed temperature series that result are virtually indistinguishable from those of the original experiments. Thus justified, the use of the same number of instrumental "training" PCs in the scenarios vis-à-vis the corresponding calibrations in the MBH re-creation allows the pure effects of eliminating/changing proxy summary PCs and truncations of proxy richness to be isolated.

3. Results

3.1 MBH ORIGINAL RECONSTRUCTION

The re-creation of the MBH98 results is shown in Figure 1. Our Northern Hemisphere mean surface temperature reconstructions (WA) are nearly identical to the original MBH98 reconstruction (Fig. 1, gray and black lines). The slight differences are related to the methodological simplifications of using a consistent annual time step throughout the reconstruction process and equal weighting of the proxies. Validation statistics for the WA reconstruction in relation to the original MBH98 results are reported in Table 1, comparable statistics for the MM-examination scenarios (1-6) are reported in Table 2. As a measure of merit, the “Reduction of Error” (RE) statistic (Cook et al., 1994) is employed, which quantifies reconstruction skill and explained variance in relation to the calibration period mean (when positive) or absence of skill (when negative). Unlike the Pearson r^2 statistic, RE penalizes reconstructions that do not capture changes in the mean value of the instrumental reference series between the calibration and verification periods, even though the reconstructions track the transient fluctuations of the instrumental values well (cf. Fig. S1 in Rutherford et al., in press). Beyond their calibration quality, and just as importantly, the reconstructions also pass temporally independent cross-verification tests (covering 1854-1901), employing the spatially more restricted portion of the instrumental record available during this time (219 out of the 1082 global calibration grid cells). These results confirm that the MBH reconstruction, per se, is closely reproducible if the procedural sequence of MBH is followed and all original proxy data are applied. Additionally, the reconstruction is robust in relation to two significant methodological simplifications--the calculation of instrumental PC series on the annual instrumental data, and the omission of weights imputed to the proxy series. Thus, this new version of the MBH reconstruction algorithm can be confidently employed in tests of the method to various sensitivities. Without such confirmation that the overall procedure has been reliably repeated, which is not reported, e.g., by Zorita et al. (2003) or von Storch et al. (2004), any evaluation of issues with the procedure has to be taken with a degree of caution.

3.2 EVALUATION OF MM CRITICISMS

The 15th century reconstructions that result from elimination of significant proxy information in scenario 1 (MM03; cf. Rutherford et al., in press) are also shown in Figure 1 (orange line). Relative to the MBH98 original and the WA results shown in Figure 1, this scenario yields much warmer NH temperatures for the 15th century, similar to MM03, and at odds with both MBH and other empirical reconstructions (Jones and Mann, 2004). According to our calculations, however, *this result does not have climatological meaning* because the reconstructions clearly fail validation tests, returning negative RE scores for both calibration and verification (-0.42 and -0.57, respectively, for 1400-1449; –0.65 and –2.71, respectively, for 1450-1499) indicative of no reconstruction skill (Table 2). The WA reconstructions including the records excluded in scenario 1 perform far better, with positive calibration and verification RE scores (0.39 and 0.48, respectively, for 1400-1449; 0.47 and 0.44, respectively, for 1450-1499) (Tables 1 and 2). It should be noted that in this experiment, the MBH step of scaling the proxy-reconstructed (i.e. fitted) instrumental PCs--so that those estimated for the calibration period have the same variance as the actual instrumental PCs--was not used, consistent with the method of MM03 (MM Supplemental Information).

Results for scenarios 2 and 3 are shown in Figure 2. The "all proxy" scenarios (2), based on different sets of proxy data richness (green and blue curves in Fig. 2), are highly similar to each other, and are also generally similar, especially in overall trajectory, to WA, which used proxy PC summaries (smoothed gray curve). [The 1700-onwards reconstruction is not shown in Figure 2, since it is highly similar to the 1404-1971 reconstruction (green).] It is interesting to note that the 1404-1971 reconstruction is also highly similar to the same kind of "all proxy" scenario developed using an entirely independent reconstruction technique--regularized expectation maximization, or RegEM (Schneider, 2001)--reported by Rutherford et al. (in press, cf. Fig. 2 there). There is a noticeable reduction of amplitude relative to the 1902-1980 mean (defined to be zero anomaly) in the "all proxy" scenarios versus WA, in particular on the order of 0.10° over ~1425-1475, ~1680-1750, and ~1815-1835, and on the order of 0.15° over ~1575-1600 and ~1850-1900. The calibration statistics and performance are extremely good (and actually above the original) for all sets of proxy richness, and the verification statistics show meaningful skill--although with reduction from the original values, especially for the 1700-

onwards reconstruction. The increase in calibration skill and reduction of verification skill shown by the 1700-onward reconstruction in scenario 2, in relation to the less data-rich 1404- and 1450-onwards reconstructions, could be an indication of some calibration overfitting, as could the similar relationships of skill statistics between scenario 2 and the WA reconstructions using proxy summaries (Table 2). Overall, the single-bladed hockey stick result of the MBH original reconstruction is robustly retained, although there is suggestion of a slight decrease in the amount of "extension of the blade" from a 15th-19th century baseline. The close fit of the calibration-period reconstructions with the 20th century instrumental record in all the reconstructions indicates that extending the perspective of the hockey stick blade with the instrumental series into the 1990s is appropriate.

Results for the exclusion of the bristlecone/foxtail pine series developed according to scenario 3 are shown by the purple curve in Figure 2. The exclusion of these proxy records generally results in slightly higher reconstructed temperatures than those derived from inclusion of all the proxy data series, with the greatest differences (averaging ~ +0.10°) over the period 1425-1510. The highest values before the 20th century in this scenario occur in the early 15th century, peaking at 0.17° in relation to the 1902-1980 mean, which are nevertheless far below the +0.40-0.80° values reported for scenario 1. The verification RE scores for this scenario are only slightly above the zero value that indicates the threshold of skill in the independent verification period. This phenomenon, which cannot be attributed to calibration overfitting because the number of proxy regressors is *reduced* rather than augmented, suggests that bristlecone/foxtail pine records *do* possess meaningful climate information at the level of the dominant eigenvector patterns of the global instrumental surface temperature grid. This phenomenon is an interesting result in itself, which is not fully addressed by examination of the local/regional relationship between the proxy ring widths and surface temperatures, and which suggests that the "all proxy" scenarios reported in Figure 2 yield a more meaningful comparison to the original MBH results than when the bristlecone/foxtail pine records are excluded. Even in the absence of this argument, the scenario 3 reconstructions in the 15th century do not exhibit large enough excursions in the positive direction (in relation to the 20th century instrumental record) to yield a double-bladed hockey stick result that diminishes the uniqueness of the late 20th century departure from long-term trends.

The results for scenario 4 (not shown) are generally very close to the reconstructions represented by the green and blue curves in Figure 2. The performance of the reconstruction models is shown in Table 2. The longest-duration differences from the full "all proxy" scenarios are three limited periods (decadal length or less) of small positive offsets ($\leq +0.12°$) in the second half of the 15th century. Over the rest of the reconstruction period, scenario 4 yields interannually varying positive (relatively more) and negative (relatively fewer) anomalies from the "all proxy" scenarios, with maximum absolute differences on the order of $\sim 0.20°$. (The exact nature and timing of anomalies from the "all proxy" scenarios varies according to the random selection of Southwestern United States ITRDB proxies excluded from reconstruction, but within the boundaries mentioned.) These results, based on strong truncations of the North American proxy data used in reconstruction, indicate that the full "all proxy" scenarios show very little, if any, bias due to over-representation of North American proxy records. Thus, the logical appropriateness of using the "all proxy" scenarios as a valid "test bed" for examining the impact of inclusion/non-inclusion of proxy PC summaries is established.

The results for scenario 5 are shown in Figure 3 for the representative cases indicated by 5a-d in Table 2. [Reconstruction statistics and time series for all experiments in scenarios 5 and

6 are reported at the website containing the WA algorithm and data.] Scenario 5a (heavy lavender curve in Fig. 3) shows the pure effect of removing the Gaspé series as an individual proxy over 1400-1449 and as part of the North American ITRDB network used to calculate proxy PCs over 1450-1499. Dropping the Gaspé series over 1400-1449 leads to a reconstruction that is higher than WA by ~ 0.05° on average, with a maximum of ~ +0.15° (1402). Over 1450-1499, the reconstruction is virtually identical WA. The divergence of scenario 5a from WA in the early 15th century is comparable to the difference shown in MM05b (their Figures 1a and 1b). The single highest reconstructed temperature in scenario 5a is 0.21° in 1406, contrasted with 0.12° in the same year in WA. This value is below the maximum high excursions of the mid-20th century (0.31°) and ~ 0.50° lower than the highest late-20th century temperatures. Scenario 5b (thin lavender curve in Fig. 3) shows the additional effect of changing from MBH (5a) to MM (5b) centering (both based on standardized proxy data) in the proxy PC calculations (see above in "Methods"). The reconstructed temperatures in 5b are slightly higher on average (~ 0.05°) than those in scenario 5a over the 15th century, with the same difference in terms of peak high excursions (0.26° in 1402). This slight shift gauges the differential impact on reconstructed climate in scenario 5 related to using the full series length for proxy centering (MM) as opposed to a centering in relation to the calibration period (MBH convention) before performing PC calculations, and is significantly smaller than the impact for this methodological difference suggested in MM05a/b.

In MM05a/b, PC calculations were not only based on series that were centered over a different period, but MM omit the step of standardizing the individual records. As shown in Ammann and Wahl (in submission), this method causes the PC analysis to capture the variance in the data in a very different way, with the first PCs mostly picking up time series with the largest amplitudes, but not what is common among time series. Only subsequent PCs (after the series have been indirectly standardized to the same variance level) capture variability that is in most of the individual series (Ammann and Wahl, in submission). Thus, the number of PCs that are required to summarize the underlying proxy data changes depending on the approach chosen. Here we verify the impact of the choice of different numbers of PCs that are included in the climate reconstruction procedure. Systematic examination of the Gaspé-restricted reconstructions using 2-5 proxy PCs derived from MM-centered but unstandardized data demonstrates changes in reconstruction as more PCs are added, indicating a significant change in information provided by the PC series. When two or three PCs are used, the resulting reconstructions (represented by scenario 5d, purple curve in Fig. 3) are highly similar (supplemental information). As shown below, these reconstructions are functionally equivalent to reconstructions in which the bristlecone/foxtail pine records are directly excluded (cf. purple curves for scenarios 6a/b in Fig. 4). When four or five PCs are used, the resulting reconstructions (represented by scenario 5c, dotted thin lavender curve in Fig. 3) are virtually indistinguishable (supplemental information) and are very similar to scenario 5b. The convergence of results obtained using four or five PCs, coupled with the closeness of 5c to 5b, indicates that information relevant to the global eigenvector patterns being reconstructed is no longer added by higher-order PCs beyond the level necessary to capture the temporal information structure of the data (four PCs using unstandardized data, two PCs using standardized data). More generally, the overall similarity of scenarios 5a-c demonstrates that when the full information in the proxy data is represented by the PC series, regardless of the precise ordering of the PCs and which centering convention is used, the impact of PC calculation methods on climate reconstruction in the MBH method is extremely small.

It is important to note that not all of the reconstructions developed in scenarios 5a-d are equivalent in terms of climatological meaning. Applying PC calculations over 1400-1449 as used in scenario 5d yields *negative RE scores in verification* (Table 2; similarly when three proxy PCs are used, supplemental information), indicating that these reconstruction *cannot be ascribed climatological meaning*. Thus, only the lavender curves in Figure 3 (scenarios 5a-c) represent potentially appropriate adjustments to the original MBH results in the early 15th century, based on question of adequate replication of the individual tree chronologies included in the Gaspé record during this time (MM05b). These scenarios pass both calibration and verification tests, although with somewhat lower RE scores and poorer reconstruction of the verification period mean than WA (Table 2); in particular, verification performance is differentially poorer when the MM centering convention is used (5b/c). The failure to verify of scenario 5d, including only two PCs derived from unstandardized data in the MM centering convention, demonstrates that the bristlecone/foxtail pine records are important for the 1400-1449 network--the information they add to PC4 in this way of calculating the PCs is necessary for a successful climate reconstruction. Restricting the PCs in MM05a/b to only the first two (5d) *indirectly* bans the bristlecone/foxtail pine records from contributing.

Scenario 6 illustrates this observation by *direct* exclusion of the bristlecone/foxtail pine records from the data set used to calculate North American ITRDB proxy PCs (note, the Gaspé data are also removed as in scenario 5). As shown in Figure 4 and Table 2, such direct exclusion leads to functionally and statistically equivalent results to scenario 5d. The 6a (using the MBH centering convention) and 6b (using the MM centering convention) reconstructions are nearly identical to the 5d reconstruction and, again, *cannot be ascribed climatological meaning* for the period 1400-1449, based on negative RE scores in verification. For the period 1450-1499, scenarios 6a/b (and 5d) yield performance measures very much like those of the WA reconstruction (Table 2), thus, from a strictly statistical perspective inclusion of the bristlecone/foxtail pine data in the proxy PC calculations neither enhances nor degrades reconstruction performance during the second half of the 15th century. Over 1450-1499, scenarios 6a/b have maximum high reconstructed temperatures of ~ 0.20°, virtually identical to average mid-20th century values, but well below late-20th temperatures. The average difference between the WA reconstruction and scenarios 6a/b over this period is ~ +0.13° (Fig. 4). Given these observations, from a climate reconstruction point of view one can argue that, in general, the bristlecone/foxtail pine records do not introduce spurious noise and their inclusion is justifiable. Their inclusion by standardization of the individual proxy records (independent of the centering convention) or, even if non-standardized series are applied, by using at least four PCs (until the resulting climate reconstructions converge), leads to reconstruction models that demonstrate skill in both calibration and independent verification.

Scenario 6c (orange curve in Fig. 4) parallels scenario 1 (fitted instrumental PCs not rescaled), and is comparable to Figure 1c in MM05b. Like scenario 1, scenario 6c fails both calibration and verification tests over 1400-1449 (RE values for calibration and verification are -0.34 and -0.56, respectively; Table 2). Although the highest temperatures in this scenario for the early 15th century are similar to those reported in MM05b (max 0.53°), which would on face value suggest the possibility of a double-bladed hockey stick result, these values *cannot be ascribed climatological meaning* and thus *cannot represent a correction to the original MBH reconstructions*. Over 1450-1499, the scenario 6c reconstructions do pass calibration and verification tests, with a maximum reconstructed temperature of 0.27°. However, the RE scores in both verification and calibration at the grid-cell level for this reconstruction are far lower than those exhibited by scenarios 6a/b (Table 2).

The results for the other combinations of proxy PC calculation method and retained proxy PCs not shown conform closely to the representative cases reported in scenarios 5 and 6, and are archived at the supplemental information website.

4. Discussion and Summary

4.1 PRINCIPLE COMPONENT SUMMARIES OF PROXY CLIMATE DATA AND THE USE/NON-USE OF SPECIFIC PROXY RECORDS

Our results show that the MBH climate reconstruction method is not only reproducible, but also proves robust against important simplifications and modifications. The results of this study demonstrate that the primary issue raised in MM05a--that the method used by MBH to form PC summaries of climate proxies from data-rich regions results in calibrations that inappropriately weight proxies with a single-bladed hockey stick-like trajectory over the 20th century--is moot in relation to the overall MBH result of uniquely high Northern Hemisphere temperatures in the late 20th century (relative to the entire 14th-20th century period). Indirect examination of this issue by use of all the continuous individual proxy data over this period, without any form of summarization (PC or otherwise), results in a reconstruction that is highly like the MBH original. Such an approach produces a small reduction of amplitude over the pre-20th-century period, but the temporal structure exhibits a clear single-bladed hockey stick shape in the 20th century (Fig. 2). A parallel result is derived from direct examination of the impact on reconstruction quality of the *methods* used to calculate proxy PC series based on the North American ITRDB data set in the 15th century [Ammann and Wahl (in submission) and here], apart from issues concerning specific proxy time series. *Thus, it is the information content of the proxies themselves that drives the shape of the MBH reconstruction, not methodological issues concerning PC summarizations of the proxy series*. This conclusion is robust to several forms of check, most powerfully the agreement of the indirect and direct examinations, and also to strong truncations of the individual North American proxy data assimilated into reconstruction (especially from the data-rich Southwestern United States). The latter check demonstrates that the "all proxy" scenario is robust in calibration to the relative density of North American data used in these reconstructions. The subsidiary issue in MM05a concerning appropriate thresholds of the RE statistic for determining reconstruction skill is also made moot by the results of the individual proxy scenarios, since the MM05a analysis is specific to the use of PC summaries.

The bristlecone/foxtail pine proxies from the Southwestern United States are shown to add necessary verification skill to the climate reconstructions for 1400-1449 when PC summaries are used and significantly greater verification skill to the reconstructions for 1400-1499 when no PC summaries are used--indicating that in these cases the records carry important climate information at the level of eigenvector patterns in global surface temperatures. These results are valid notwithstanding issues concerning these proxies' empirical relationship to local/regional surface temperatures after 1850 (cf. MM05a/b; Hughes and Funkhouser, 2003; MBH99; Graybill and Idso, 1993). These results enhance the validity of the MBH assumption that proxies used in the reconstruction process do not necessarily need to be strongly related to local/regional surface temperatures, as long as they register climatic variations that are linked to the empirical patterns of the global temperature field that the MBH method (and other climate field reconstructions) target. The fact that these additions of skill occur *only* in the verification period (1854-1901) for scenarios 2 and 3 (Table 2) and are far more pronounced in verification for scenarios 5 and 6

(Table 2; supplementary information) leads to the further conclusion that these proxies do not generate spurious increases in calibration fit that thereby downweight the value of other (presumably more locally climate-correlated) proxy series. Over 1450-1499, the bristlecone/foxtail pine proxies neither enhance nor degrade reconstruction performance when PC summaries are used. Thus, in this situation it is logically appropriate to retain these proxies over the entire 15th century, since they are necessary for verification skill in the first half of this period and have no impact on calibration and verification performance in the later half.

Removal of the Gaspé record from the MBH reconstruction during the early 15th century when proxy PC summaries are used, as proposed in MM05b, represents a potentially appropriate adjustment of the original MBH results. With this correction, the adjustment of the MBH time series over 1400-1449 averages ~ + 0.05-0.10°, depending on the centering convention used for the proxy PC calculations, as shown by the lavender curves in Figure 3. This adjustment yields a maximum high excursion of 0.21-0.26° over the entire 15th century, which is similar to the mean of mid-20th century temperatures, but ~ 0.50° below late-20th century temperatures. Since the calibration and verification statistical performance for these adjusted reconstructions is somewhat poorer than that achieved by the WA reconstruction, it could be argued empirically that WA represents the best reconstruction achieved in our analysis for the 15th century, and that no change to the MBH reconstruction is appropriate. However, if adjustment is considered appropriate, then the MBH centering convention for proxy PC calculation might be favored over the MM centering convention, based on better statistical performance in verification (associated with the lower values for adjustment given above).

4.2 ROBUSTNESS OF MBH98 RESULTS IN RELATION TO MM CRITICISMS

Our results do not support the large upward "correction" to reconstructed 15th century temperatures described in MM03 and MM05b, and leave unaltered the single-bladed hockey stick conclusion of strongly anomalous late 20th century temperatures. This conclusion is retained even if the bristlecone/foxtail pine records were (inappropriately) eliminated for the 15th century, because the maximum high excursion during the entire 15th century would be ~ 0.35°, which is just slightly above the highest values that occurred during the mid-20th century, but still well below late-20th temperatures.

More generally, our results highlight the necessity of reporting skill tests for each reconstruction model, as is customary in quantitative paleoclimate reconstruction. Taking this consideration into account, there is strong scientific reason to conclude that the 15th century reconstructions reported in MM03, which show hemispheric temperatures as high, and higher, than those of the mid 20th century, do not have climatological meaning. This double-bladed hockey stick result, while computable using the MBH algorithm, does not pass standard validation tests (RE scores <0 for both calibration and verification). These validation results indicate that the annual climatological mean of the calibration period would be a *better* predictor over 1854-1980 (the period of available instrumental values that can be used for comparison) than the reconstruction models of this scenario (cf. Cook et al., 1994). The same result holds for the somewhat lower-amplitude double-bladed hockey stick reconstruction reported in MM05b over 1400-1449, which fails to pass calibration and verification tests for this period. Thus, the primary argument offered by MM for rejecting the uniqueness of high late-20th century temperatures is found to be without merit, based on examination of the empirical quality of the reconstruction models these authors propose.

Overall, the primary outcome from our results is that the work reported in MM03, MM05a, and MM05b does not provide substantive reason to invalidate the general conclusion of anomalous warmth in the later 20th century derived from the MBH reconstruction method and proxy data framework. We find that this result is neither an artifact of selection of the proxy series nor the result of formation or application of PC summaries in the reconstruction procedure. With only a slight upward adjustment (on average ~ +0.05°) to the original MBH reconstruction, which is potentially warranted by removal of the Gaspé record over 1400-1449, the MBH result of anomalous warmth in the later 20th century remains consistent with other paleoclimate reconstructions developed for the last 1-2 millenia (Rutherford et al., in press; Moberg et al., 2005; Cook et al., 2004; Huang, 2004; Jones and Mann, 2004; Mann and Jones, 2003; Esper et al., 2002; Briffa et al., 2001; Crowley and Lowery, 2000; Jones et al., 1998; Bradley and Jones, 1993), especially in light of the recent reconciliation of the Esper et al. (2002; cf. Cook et al., 2004) and MBH reconstructions reported by Rutherford et al. (in press).

4.3 POTENTIAL AMPLITUDE LOSS IN MBH RECONSTRUCTION

Issues concerning possible loss of amplitude in the MBH reconstruction (Moberg et al., 2005; Huang, 2004; von Storch et al., 2004) remain difficult to assess within the proxy framework alone (cf. MBH99, for initial examination of this issue in the MBH reconstructions). Empirically, one indication for such a loss could be derived from positive anomalies for the reconstructed versus instrumental means in the verification period (Table 2). Positive anomalies indicate that reductions of low frequency amplitude during the verification period (by values shown in Table 1) might reasonably be expected for the original MBH reconstructions between 1400-1449 (~0.04°), 1600-1699 (~0.04°), 1700-1749 (~0.02°), 1750-1759 (~0.03°), and 1760-

1779 (~0.01°). These values for amplitude loss are much less than those suggested by Moberg et al. (2005), von Storch et al. (2004), and Huang (2004), and may represent a lower bound on the actual amount of loss.

Ammann et al. (in preparation) and Mann et al. (in press) develop much more systematic tests of potential downward reconstruction bias in the MBH method, using AOGCM output in the style of von Storch et al. (2004), but in a model environment with significantly different depictions of the transient solar forcing series and a lower overall sensitivity.

The anomalies of reconstructed versus instrumental verification means for scenarios 1-4 indicate similar, and enhanced, tendencies for reconstruction amplitude loss (Table 2). These results carry two final implications in relation to the arguments made by MM. First, the enhanced loss of verification-period amplitude by the individual proxy scenarios coupled with the conclusion of support for the uniqueness of late-20th century temperatures indicate that, rather than being problematic, PC summaries as developed in the original MBH method actually enhance the quality of the resulting climate reconstructions. This conclusion, focused on the verification period results, demonstrates the usefulness of the PC summaries in a framework that is independent of appropriate/inappropriate variation carried by the PC summaries in the calibration period; it might additionally illustrate that by reducing the number of predictors concentrated in a relatively small area (as compared to the rest of the records spread over the globe), overfitting in the calibration period is omitted and thus the predictive skills of the reconstruction models are enhanced. Second, the probable loss of downward amplitude of nearly 0.20° by the 15th century networks indicated for scenario 1 (Table 2) suggests that even the highest excursions of the MM03-type reconstruction may be too large by this amount. In this case, the highest reconstructed temperature in scenario 1 becomes ~0.60°, indicating that it would not be *higher* than latest 20th century temperatures; thereby removing this particular implication of the MM03 reconstruction on the basis of downward amplitude loss alone, apart from the lack of climatological meaning indicated by validation scores.

4.4 SUMMARY

The results presented here show no evidence for removing the MBH Northern Hemisphere temperature reconstruction from the list of important climate reconstructions of the past six centuries, on the basis of alleged "flaws" in its underlying methodology. Indeed, our analyses act as an overall indication of the robustness of the MBH reconstruction to a variety of issues raised concerning its methods of assimilating proxy data, and also to two significant simplifications of the MBH method that we have introduced. Compared to the original, a small adjustment in the climate reconstruction (~ +0.05°) resulting from removal of a single tree ring series (Gaspé) between 1400 and 1449 appears justifiable. However, the shape of a single bladed "hockey stick"-like evolution of Northern Hemisphere temperature over the last 600 years is strongly confirmed within the MBH reconstruction framework. Questions of potential loss of downward amplitude in the MBH method remain, but the evidence developed here from the perspective of the proxy data themselves suggests such losses may be smaller than those shown in other recent work.

To stimulate further experimentation with the MBH reconstruction method, we have established a user-friendly version of our simplified MBH algorithm in the public-domain R language, accessible at a website sponsored by the National Center for Atmospheric Research (NCAR), USA (http://www.cgd.ucar.edu/ccr/ammann/millennium/MBH_reevaluation.html).

The relevant proxy and instrumental data utilized are also housed at this site, along with a simple description of the MBH algorithm and scenarios for its use.

Acknowledgments

We are grateful to L. Mearns and the NCAR Weather and Climate Impact Assessment Science Initiative for support, and acknowledge D. Nychka and C. Tebaldi of NCAR for statistical and programming aid. The National Center for Atmospheric Research is sponsored by the National Science Foundation, USA.

Appendix

The reductions of verification RE scores for WA versus MBH in Table 1 are possibly artifacts of using a spatially restricted instrumental grid to calculate the verification statistics. As mentioned in "Results", the number of grid cells available over the 1854-1901 verification period is 219, versus the 1082 cells used in calibration. Of the 219 global total, 172 cells are in the Northern Hemisphere, versus 707 Northern Hemisphere cells that are actually reconstructed in calibration (and thereby throughout the entire length of the reconstruction period). For the 1820-1980 proxy network for which comparable "sparse grid" verification reconstructions are available, the WA average reconstructed Northern Hemisphere mean temperature (entire verification period) is highly similar to that reported by MBH (-0.224° and -0.229°, respectively), while the corresponding standard deviations show a reduced variation in WA compared to MBH (0.111° and 0.146°, respectively). The corresponding MBH instrumental standard deviation is 0.165°. [MBH sparse grid data are from "Global Temperature Patterns in Past Centuries: an Interactive Presentation"; NOAA Paleoclimatology Program.] Thus, there is appropriate variation in the MBH sparse grid reconstructions that is underrepresented in the parallel WA reconstructions.

However, this loss of variation by WA does *not* occur in the reported verification period annual mean reconstructions, calculated from the full 707 Northern Hemisphere grid-cell set (standard deviations of 0.080° for WA and 0.084° for MBH). Thus, there is apparently a slight differential impact of the two reconstruction procedures in terms of the variability present in the sparse verification grid, which does not "pass through" at the scale of the full reported reconstruction grid. This observation suggests that the differences between the WA and MBH verification RE values shown could be overestimates of the true (but unmeasurable) differences occurring at the level of the full Northern Hemisphere reconstruction grid, driven by the non-random sampling of the sparse sub-grid used to calculate verification performance statistics.

References

Bradley, R. S. and Jones, P. D.: 1993, '"Little Ice Age" summer temperature variations: their nature and relevance to recent global warming trends', *Holocene* **3 (4)**, 367-376.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: 2001, 'Low-frequency temperature variations from a northern tree ring density network', *J. Geophys. Res.* **106**, 2929-2941.

Cane, M. A., Clement, A. C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., Zebiak, S. E., and Murtugudde, R.: 1997, 'Twentieth-century sea surface temperature trends', *Science* **275**, 957-960.

Chapman, D. S., Bartlett, M. G., and Harris, R. N.: 2004, 'Comment on "Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions", by M. Mann and G. Schmidt', *Geophys. Res. Lett* **31**, doi:10.1029/2003GL019054.

Cole, J. E. and Cook, E. R.: 1998, 'The changing relationship between ENSO variability and moisture balance in the continental United States', *Geophys. Res. Lett.* **25(24)**, 4529-4532.

Cook, E. R., Briffa, K. R., and Jones, P. D.: 1994, 'Spatial Regression Methods in Dendroclimatology: A Review and Comparison of Two Techniques', *Int. J. Clim.* **14**, 379-402.

Cook, E. R., Esper, J., and D'Arrigo, R. D.: 2004, 'Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years, *Quat. Sci. Rev.* **23**, 2063-2074.

Crowley, T. J. and Lowery, T.: 2000. 'How Warm Was the Medieval Warm Period? A comment on "Man-made versus natural climate change"', *Ambio* **29**, 51-54.

Crowley, T. J., Baum, S. K., Kim, K.-Y., Hegerl, G. C. and Hyde, W. T.: 2003, 'Modeling ocean heat content changes during the last millennium', *Geophys. Res. Lett.* **30(18)**, 1932, doi:10.1029/2003GL017801.

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.

Evans, M. N., Kaplan, A., and Cane, M. A.: 2002, 'Pacific sea surface temperature field reconstruction from coral $d^{18}O$ data using reduced space objective analysis', *Paleoceanography* **17**, 7,1-7,13.

Folland, C. K., Karl, T. R., Christy, J. R., Clarke, R. A., Gruza, G. V., Jouzel, J., Mann, M. E., Oerlemans, J., Salinger, M. J., and Wang, S.-W.: 2001, 'Observed Climate Variability and Change', in Houghton, J. T., et al. (eds.), *Climate Change 2001: The Scientific Basis*, Cambridge Univ. Press, Cambridge, 99-181.

Fritts, H.C.: 1976, *Tree Rings and Climate*, London, Academic Press.

Goosse, H., Crowley, T. J., Zorita, E., Ammann, C. M., Renssen, and Driessschaert, H.: 2005, 'Modeling the climate of the last millennium: what causes the differences between simulations?', *Geophys. Res. Lett.* (in press).

Graybill, D. and Idso, S.: 1993, 'Detecting the aerial fertilization effect of atmospheric CO_2 enrichment in tree-ring cohronologies', *Global Biogeochemical Cycles* **7**, 81-95.

Harris, R. N. and Chapman, D. S.: 2001, 'Mid-latitude (30-60N) climatic warming inferred by combining borehole temperatures with surface air temperatures', *Geophys. Res. Lett.* **28**, 747-750.

Huang, S., Pollack, H.N., and Shen, P.-Y.: 2000, 'Temperature trends over the past five centuries reconstructed from borehole temperatures'. *Nature* **403**, 756-758.

Huang, S.: 2004, 'Merging information from different resources for new insight into climate change in the past and future', *Geophys. Res. Lett*. **31**, L13205.

Hughes, M. K. and Funkhouser, G.: 2003, 'Frequency-dependent climate signal in upper and lower forest border trees in the mountains of the Great Basin', *Clim. Change* **59**, 233-244.

Jones P. D., and Briffa, K. R.: 1992, 'Global surface temperature variations during the 20th century: Part 1: Spatial, temporal and seasonal details', *Holocene* **1**, 165-179.

Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: 1998, 'High-Resolution Paleoclimatic Records for the Last Millennium: Interpretation, Integration and Comparison with Circulation Model Control-Run Temperatures', *Holocene* **8**, 455-471.

Jones, P. D. and Mann, M. E.: 2004, 'Climate Over Past Millennia', *Rev. Geophys*. **42**, RG2002, doi: 10.1029/2003RG000143.

Jones, P. D., Osborn, T. J. ,and Briffa, K. B.: 2001, 'The Evolution of Climate Over the Last Millennium', *Science* **292**, 662-667.

Kaplan, A., Kushnir, Y., Cane, M. A., and Blumenthal, M. B.: 1997, 'Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures', *J. Geophys. Res.* **102(C13)**, 27,835-27,860.

Mann, M. E., Bradley R. S., and Hughes, M. K.: 1998, 'Global-scale temperature patterns and climate forcing over the past six centuries', *Nature* **392**, 779-787.

Mann, M. E., Bradley R. S., and Hughes, M. K.: 1999, 'Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations', *Geophys. Res. Lett.* **26**, 759-762.

Mann, M. E., Bradley R. S., and Hughes, M. K.: 2000, 'Long-term variability in the El Nino Southern Oscillation and associated teleconnections', in Diaz, H. F. and Markgraf, V. (eds.)

El Nino and the Southern Oscillation: Multiscale Variability and its Impacts on Natural Ecosystems and Society, Cambridge University Press, Cambridge, UK, 357-412.

Mann, M. E. and Jones, P. D.: 2003, 'Global surface temperatures over the past two millennia', *Geophys. Res. Lett.* **30**, 1820, 10.1029/2003GL017814.

Mann, M. E. and Rutherford, S.: 2002, 'Climate Reconstruction Using "Pseudoproxies"', *Geophys. Res. Lett.* **29**, 139,1-139,4.

Mann, M. E., Rutherford, S., Bradley, R. S., Hughes, M. K., Keimig, F. T.: 2004 'Optimal Surface Temperature Reconstructions using Terrestrial Borehole Data', *J. Geophys. Res.* **108(D7)**, 4203, doi: 10.1029/2002JD002532.

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: in press, 'Testing the Fidelity of Methodologies Used in "Proxy"-Based Reconstructions of Past Climate', *J. Climate*.

Mann, M. E. and Schmidt, G. A.: 2003, 'Ground vs. Surface Air Temperature Trends: Implications for Borehole Surface Temperature Reconstructions', *Geophys. Res. Lett.* **30(12)**, 1607, doi: 10.1029/2003GL017170.

McIntyre, S., and McKitrick, R.: 2003, 'Corrections to the Mann et al (1998) Proxy Data Based and Northern Hemispheric Average Temperature Series', *Energy and Environment* **14**, 751-771.

McIntyre, S., and McKitrick, R.: 2005a, 'Hockey sticks, principal components, and spurious significance', *Geophys. Res. Lett.* **32**, L03710, doi:10.1029/2004GL021750.

McIntyre, S., and McKitrick, R.: 2005b, 'The M&M Critique of the MBH98 Northern Hemisphere Climate Index: Update and Implications', *Energy and Environment* **16**, 69-100.

Moberg, A., Sonechkin, D., Holmgren, K., Datsenko, N., and Karlén, W.: 2005, 'Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data', *Nature* **433**, 613-617.

Prentice, I. C., Bartlein, P. J., and Webb III, T. I.: 1991, 'Vegetation and Climate Change in Eastern North America Since the Last Glacial Maximum', *Ecology* **72**, 2038-2056.

Rajagopalan, B., Cook, E.R., Lall, U., and Ray, B.K.: 2000, 'Spatiotemporal Variability of ENSO and SST Teleconnections to Summer Drought over the United States during the Twentieth Century', *J. Climate* **13**, 4244-4255.

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., and Jones, P. D.: 2005. 'Proxy-based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain', *J. Climate* (in press).

Rutherford S. and Mann, M. E.: 2004, 'Correction to "Optimal surface temperature reconstructions using terrestrial borehole data"', *J. Geophys. Res.* **109(D11107)** doi:10.1029/2003JD004290.

Schmidt, G. A. and Mann, M. E.: 2004, 'Reply to comment on "Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions" by D. Chapman et al.', *Geophys. Res. Lett.* **31**, L07206, doi: 10.1029/2003GL0119144.

Shindell, D. T., Schmidt, G. A., Mann, M. E., Rind, D., and Waple, A.: 2001, 'Solar Forcing of Regional Climate Change During the Maunder Minimum', Science 294, 2149-2152.

Shindell, D. T., Schmidt, G. A., Miller, R. L. and Mann, M. E.: 2003, 'Volcanic and Solar Forcing of Climate Change during the Preindustrial Era', *J. Climate* **16**, 4094-4107.

Schneider, T.: 2001, 'Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values', *J. Climate* ***14***, 853-887.

Von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., Gonzalez-Rouco, F., and Tett, S.F.B.: 2004, 'Reconstructing Past Climate from Noisy Data', *Science* **306**, 679-682.

Waple, A., Mann, M. E., and Bradley, R. S.: 2002, 'Long-term Patterns of Solar Irradiance Forcing in Model Experiments and Proxy-based Surface Temperature Reconstructions', *Clim. Dyn.* **18**, 563-578.

Zorita, E., Gonzalez-Rouco, F., and Legutke, S.,: 2003, 'Testing the Mann et al. (1998) Approach to Paleoclimate Reconstructions in the Context of a 1000-Yr Control Simulation with the ECHO-G Coupled Climate Model', *J. Climate* **16**, 1378-1390, 2003.

Figure 1 Comparison of Wahl/Ammann and Mann/Bradley/Hughes (1998) Northern Hemisphere annual mean temperature reconstructions with scenario 1 reconstruction [emulation of McIntyre/McKitrick (2003) results]. Wahl/Ammann reconstruction is shown in gray, both at annual resolution and smoothed with 49-year Gaussian filter. Mann/Bradley/Hughes reconstruction is shown in black, smoothed with 49-year Gaussian filter. Scenario 1 reconstruction for 15th century is shown in orange at annual resolution. Zero reference level is mean value for 1902-1980 instrumental data. Instrumental data are shown in light lavender: for full Northern Hemisphere grid over standard calibration period (1902-1980), and for spatially-restricted Northern Hemisphere grid over verification period (1854-1901) (cf. Appendix).

Figure 2 Northern Hemisphere annual mean temperature reconstructions using only individual proxy records--without principle component summaries of proxy data (scenarios 2 and 3, described in text; cf. Table 2). Green--all individual records for 1404-1449 and 1450-1499 Mann/Bradley/Hughes (1998) proxy networks (1450-1499 network used to reconstructed entire period from 1450-1971) (scenario 2). Blue--all individual records for 1800-1820 Mann/Bradley/Hughes (1998) proxy network used to reconstruct 1800-1971 (scenario 2). Purple--same as for green curve, except that 15 bristlecone/foxtail pine records questioned in McIntyre/McKitrick (2005a/b) are removed from proxy roster (scenario 3). Gray line, zero reference level, and instrumental data same as in Figure 1.

Figure 3 Northern Hemisphere annual mean temperature reconstructions for 15th century with proxy principle component (PC) summaries retained, excluding the St. Anne's River (Gaspé) series as an individual proxy over 1400-1449 and as part of data set used to calculate PCs of

North American proxies from the International Tree Ring Data Base (ITRDB) (scenario 5, described in text; cf. Table 2). Heavy lavender--reconstructions using standardized anomalies for ITRDB proxies (for input into PC extraction) referenced to 1902-1980 mean values (scenario 5a). Thin lavender--reconstructions using standardized anomalies for ITRDB proxies referenced to mean values over 1400-1980 and 1450-1980, for 1400-1449 and 1450-1499 reconstructions, respectively (scenario 5b). Dotted thin lavender--reconstructions using *non*-standardized anomalies for ITRDB proxies referenced to mean values over 1400-1980 and 1450-1980, for 1400-1449 and 1450-1499 reconstructions, respectively (with retention of sufficient PC series to capture temporal information structure of ITRDB data) (scenario 5c). Purple--same as dotted thin lavender, *except* with too few PC series retained to capture information dynamic structure of ITRDB data (acting in effect as *exclusion* of bristlecone/foxtail pine records from PC calculations) (scenario 5d). Gray--Wahl/Ammann reconstruction (for reference).

Figure 4 Gaspé proxy restrictions as in Figure 3, with additional exclusion of 15 bristlecone/foxtail pine records from data set used to calculate principle components (PCs) of North American proxies from the International Tree Ring Data Base (ITRDB) (scenario 6, described in text; cf. Table 2). Heavy purple--reconstructions using standardized anomalies for ITRDB proxies (for input into PC extraction) referenced to 1902-1980 mean values (scenario 6a). Thin purple--reconstructions using standardized anomalies for ITRDB proxies referenced to mean values over 1400-1980 and 1450-1980, for 1400-1449 and 1450-1499 reconstructions, respectively (scenario 6b). Orange--reconstructions using *non*-standardized anomalies for ITRDB proxies referenced to mean values over 1400-1980 and 1450-1980, for 1400-1449 and 1450-1499 reconstructions, respectively (with fitted instrumental PCs *not* rescaled by factor which equates variances of fitted and instrumental PCs over calibration period) (scenario 6c).

Gray--Wahl/Ammann reconstruction (for reference).

Table 1 RE Scores and Verification Mean Anomalies for MBH Reconstructions

Proxy Network MBH – periods	NH Mean RE Calibration-period MBH WA	All Grid Cells RE Calibration-period MBH WA	NH Mean RE Verification-period MBH WA	Reconstructed *minus* Instrumental Means Verification-period WA
1400-1449	0.42 / 0.39	0.08 / 0.06	0.51 / 0.48	0.039° *
1450-1499	0.41 / 0.47	0.09 / 0.09	0.51 / 0.44	-0.040
1500-1599	0.42 / 0.48	0.10 / 0.10	0.49 / 0.47	-0.035
1600-1699	0.67 / 0.64	0.14 / 0.14	0.53 / 0.46	0.038
1700-1729	0.71 / 0.69	0.14 / 0.18	0.57 / 0.50	0.019
1730-1749	0.71 / 0.69	0.15 / 0.18	0.61 / 0.55	0.015
1750-1759	0.74 / 0.71	0.18 / 0.24	0.57 / 0.61	0.028
1760-1779	0.74 / 0.73	0.26 / 0.32	0.70 / 0.54	0.009
1780-1799	0.74 / 0.74	0.27 / 0.35	0.69 / 0.59	-0.017
1800-1819	0.75 / 0.75	0.27 / 0.35	0.68 / 0.60	-0.026
1820-1980	0.76 / 0.75	0.30 / 0.37	0.69 / 0.62	-0.031

* Verification-period instrumental mean is -0.193°, relative to mean for period 1902-1980

Note: The reductions of verification RE scores for WA versus MBH in Table 1 are possibly artifacts of using a spatially restricted instrumental grid to calculate the verification statistics. Cf. Appendix.

Table 2 RE Scores and Verification Mean Anomalies for MM-Reconstruction Scenarios

Scenario with MBH Proxy networks	NH Mean RE Calibration period MM WA	All Grid Cells RE Calibration period MM WA	NH Mean RE Verification period MM WA	NH Offset in Verification period MM WA *Instrumental mean (-0.193°)*
1 Omission ITRDB & other N. Am. Data				
(reconstructed instrumental PCs not re-scaled)				
1400-1449	*-0.42* / 0.39	*-0.09* / 0.06	*-0.57* / 0.48	0.194° / 0.039°
1450-1499	*-0.65* / 0.47	*-0.20* / 0.09	*-2.71* / 0.44	0.170 / -0.040
2 no PC for ITRDB Data				
1404-1449	0.60 / 0.39*	0.06 / 0.06*	0.36 / 0.48*	0.113 / 0.039*
1450-1499	0.67 / 0.47	0.12 / 0.09	0.28 / 0.44	0.107 / -0.040
1700-1729	0.73 / 0.69	0.17 / 0.18	0.19 / 0.50	0.140 / 0.019
1800-1819	0.76 / 0.75	0.36 / 0.35	0.35 / 0.61	0.116 / -0.026
3 no PC for ITRDB and no Bristlecone/Foxtail				
1404-1449	0.62 / 0.39*	0.06 / 0.06*	0.06 / 0.48*	0.169 / 0.039*
1450-1499	0.71 / 0.47	0.12 / 0.09	0.10 / 0.44	0.138 / -0.040
4 no PC for ITRDB and Limitation of Records				
(a) 1404-1449	0.57 / 0.39*	0.06 / 0.06*	0.03 / 0.48*	0.167 / 0.039*
1450-1499	0.70 / 0.47	0.13 / 0.09	0.15 / 0.44	0.115 / -0.040
(b) 1404-1449	0.56 / 0.39*	0.06 / 0.06*	0.20 / 0.48*	0.139 / 0.039*
1450-1499	0.68 / 0.47	0.13 / 0.09	0.30 / 0.44	0.081 / -0.040
5 No Gaspé**				
(a) MBH centered and standardized 2 PCs (svd):				
1400-1449	0.28 / 0.39	0.04 / 0.06	0.34 / 0.48	0.093 / 0.039
1450-1499	0.47 / 0.47	0.09 / 0.09	0.44 / 0.44	-0.030 / -0.040
(b) MM centered and standardized 2 PCs (svd):				
1400-1449	0.32 / 0.39	0.04 / 0.06	0.18 / 0.48	0.135 / 0.039
1450-1499	0.49 / 0.47	0.09 / 0.09	0.46 / 0.44	-0.007 / -0.040
(c) MM centered, not-standardized 4 PCs (princomp):				
1400-1449	0.29 / 0.39	0.04 / 0.06	014 / 0.48	0.144 / 0.039
1450-1499	0.48 / 0.47	0.09 / 0.09	0.43 / 0.44	0.001 / -0.040
(d) MM centered, not-standardized 2 PCs (princomp):				
(indirect Bristlecone/ Foxtail exclusion)	0.19 / 0.39	0.02 / 0.06	*-0.18* / 0.48	0.190 / 0.039
1400-1449	0.46 / 0.47	0.09 / 0.09	0.42 / 0.44	0.008 / -0.040
1450-1499				
6 No Gaspé and no Bristlecone/Foxtail**				
(a) MBH-centered, standardized, 2 PCs (svd):				
1400-1449	0.24 / 0.39	0.03 / 0.06	*-0.13* / 0.48	0.182 / 0.039
1450-1499	0.48 / 0.47	0.09 / 0.09	0.43 / 0.44	0.008 / -0.040
(b) MM-centered, standardized, 2 PCs (svd):				
1400-1449	0.19 / 0.39	0.02 / 0.06	*-0.20* / 0.48	0.193 / 0.039
1450-1499	0.47 / 0.47	0.09 / 0.09	0.42 / 0.44	0.008 / -0.040
(c) MM-centered, *not*-standardized, 2 PCs (princomp):				
(reconstructed instrumental PCs not re-scaled)	*-0.34* / 0.39	*-0.08* / 0.06	*-0.56* / 0.48	0.196 / 0.039
1400-1449	0.32 / 0.47	0.005 / 0.09	0.14 / 0.44	-0.062 / -0.040
1450-1499				

* note slight difference in data coverage with scenario limited to 1404-1449

** statistics given for scenario iterations described in "Results"; statistics for *all* iterations of each scenario available at supplemental information website

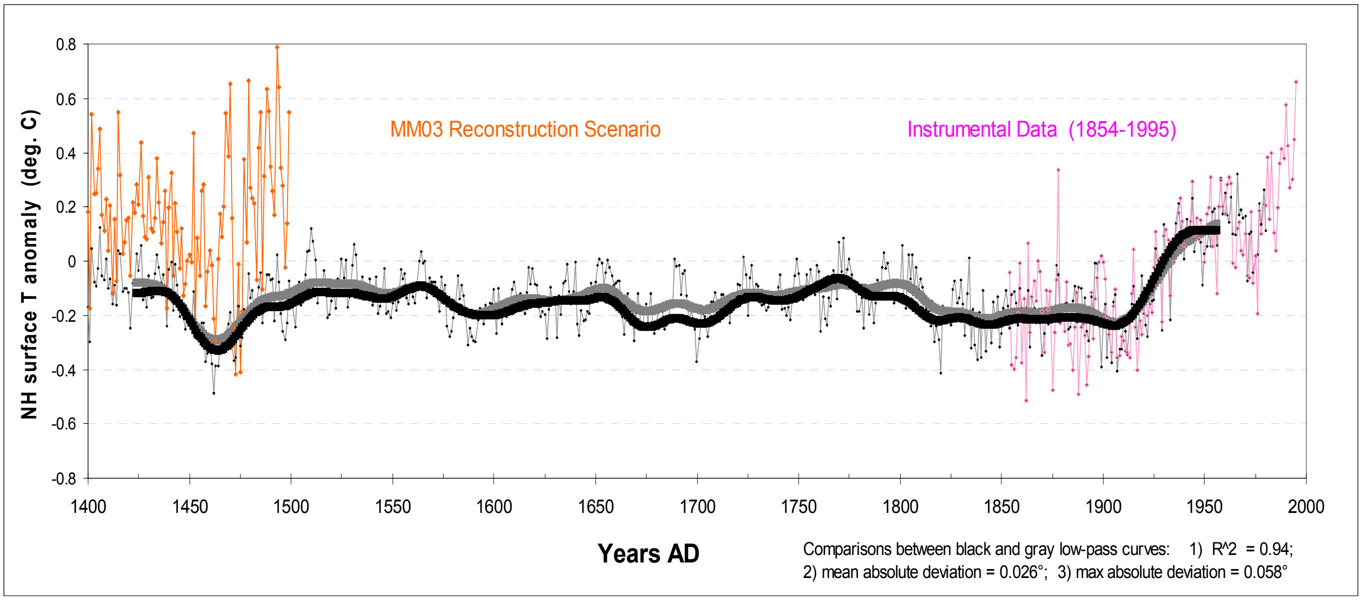

Figure 1

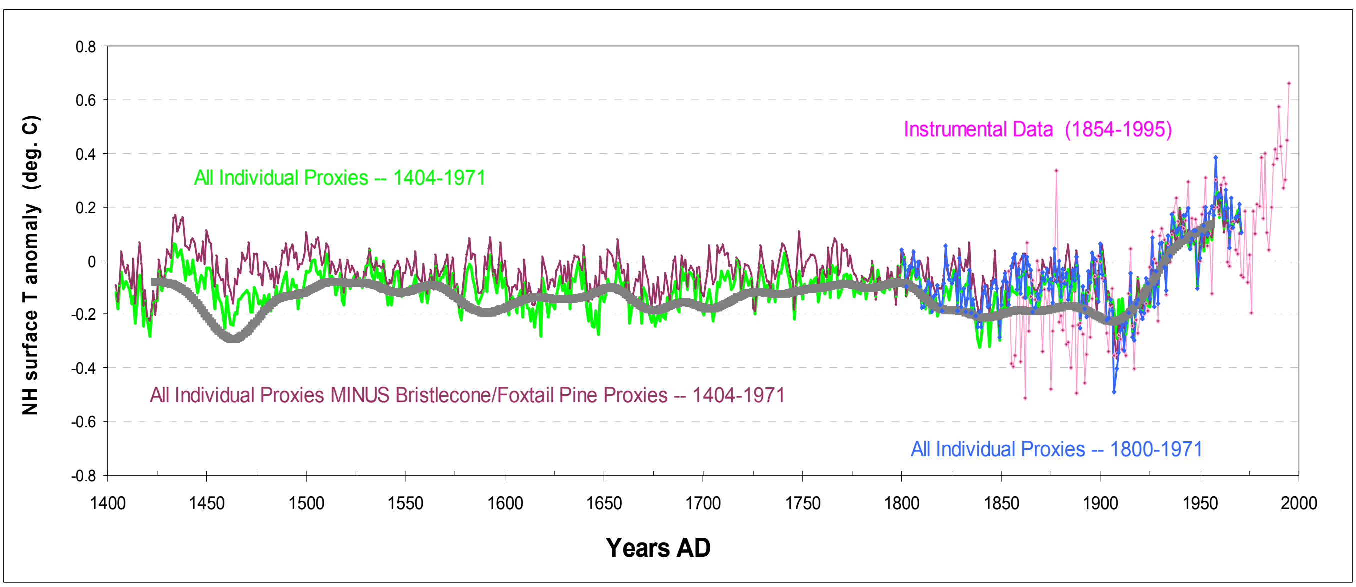

Figure 2

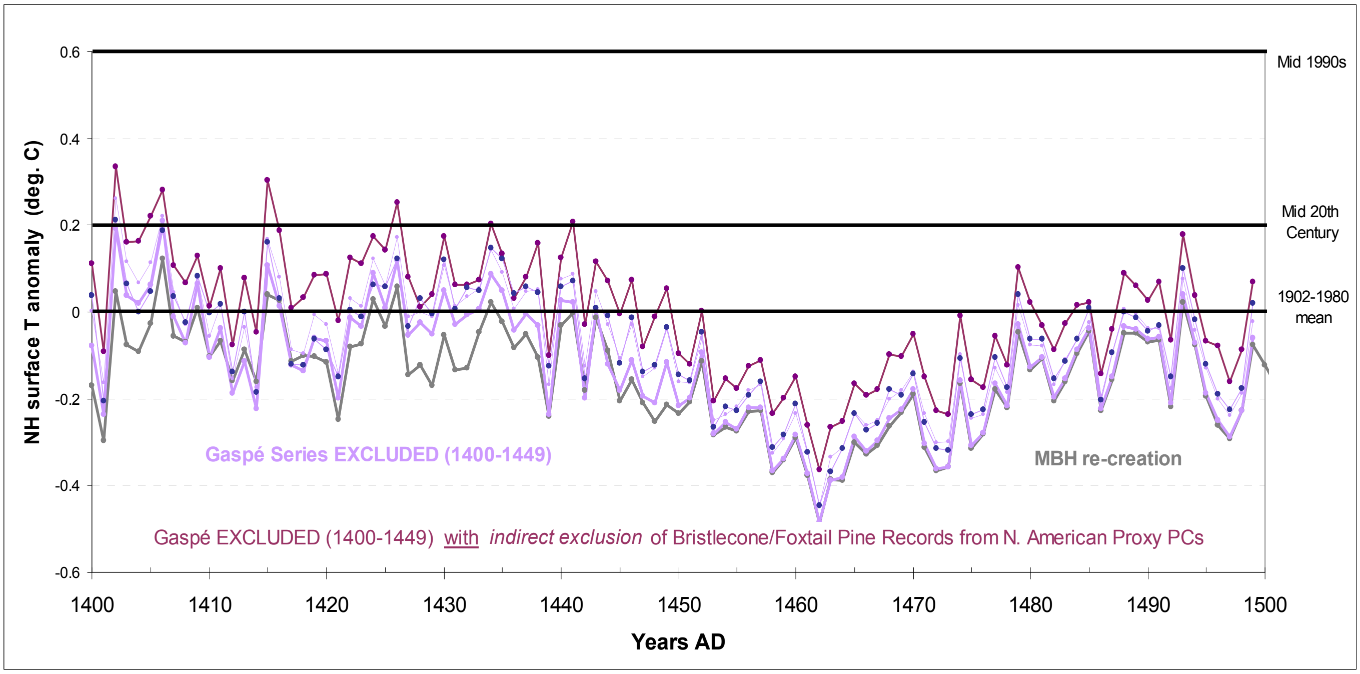

Figure 3

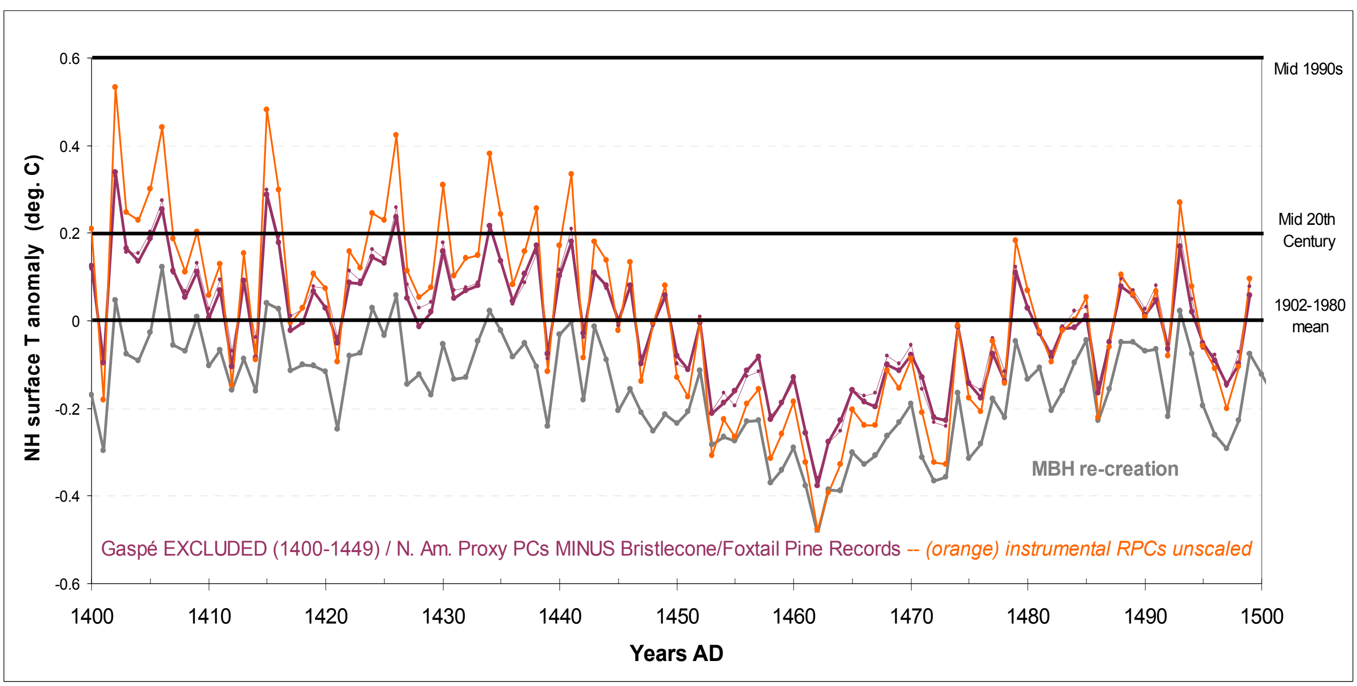

Figure 4