

Millennial temperature reconstruction intercomparison and evaluation

M. N. Juckes¹, M. R. Allen², K. R. Briffa³, J. Esper⁴, G. C. Hegerl⁵, A. Moberg⁶, T. J. Osborn³, and S. L. Weber⁷

¹British Atmospheric Data Centre, SSTD, Rutherford Appleton Laboratory Chilton, Didcot, Oxfordshire, OX11 0QX, UK

²University of Oxford, Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, UK

³Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

⁴Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

⁵Div. of Earth and Ocean Sciences, Nicholas School of the Environment and Earth Science, Duke University, Durham, NC, USA

⁶Department of Meteorology and Department of Physical Geography and Quaternary Geology, Stockholm University, 106 91 Stockholm, Sweden

⁷Royal Netherlands Meteorological Institute (KNMI), Postbus 201, 3730 AE De Bilt, The Netherlands

Received: 26 September 2006 – Published in *Clim. Past Discuss.*: 26 October 2006

Revised: 12 July 2007 – Accepted: 21 September 2007 – Published: 5 October 2007

Abstract. There has been considerable recent interest in paleoclimate reconstructions of the temperature history of the last millennium. A wide variety of techniques have been used. The interrelation among the techniques is sometimes unclear, as different studies often use distinct data sources as well as distinct methodologies. Here recent work is reviewed and some new calculations performed with an aim to clarifying the consequences of the different approaches used. A range of proxy data collections introduced by different authors is used to estimate Northern Hemispheric annual mean temperatures with two reconstruction algorithms: (1) inverse regression and, (2) compositing followed by variance matching (CVM). It is found that inverse regression tends to give large weighting to a small number of proxies and that the second approach (CVM) is more robust to varying proxy input. The choice of proxy records is one reason why different reconstructions show different ranges. A reconstruction using 13 proxy records extending back to AD 1000 shows a maximum pre-industrial temperature of 0.25 K (relative to the 1866 to 1970 mean). The standard error on this estimate, based on the residual in the calibration period, is 0.14 K. Instrumental temperatures for two recent years (1998 and 2005) have exceeded the pre-industrial estimated maximum by more than 4 standard deviations of the calibration period residual.

1 Introduction

The climate of the last millennium has been the subject of much debate in recent years, both in the scientific literature and in the popular media. This paper reviews reconstructions of past temperature, on the global, hemispheric, or near-hemispheric scale, by Jones et al. (1998) [JBB1998], Mann et al. (1998) [MBH1998], Mann et al. (1999) [MBH1999], Huang et al. (2000) [HPS2000], Crowley and Lowery (2000) [CL2000], Briffa et al. (2001) [BOS2001], Esper et al. (2002) [ECS2002], Mann and Jones (2003) [MJ2003], Moberg et al. (2005) [MSH2005], Oerlemans (2005) [OER2005], and Hegerl et al. (2007a) [HCA2007].

Climate variability can be partitioned into contributions from (1) internal variability of the climate system and (2) response to forcings, with the forcings being further partitioned into natural and anthropogenic components. The dominant change in forcing in the late 20th century arises from human impact in the form of greenhouse gases (primarily carbon dioxide, methane and chloro-fluoro carbons: Mitchell et al., 2001, [IPCC2001]). The changes in concentration of these gases in the atmosphere are well documented and their radiative properties which reduce, for a given temperature, radiative loss of heat to space from the mid and lower troposphere are beyond dispute (for carbon dioxide, this was first quantified by Arrhenius, 1896).

However, there remains some uncertainty on two issues: firstly, how much of the observed climate change in the 20th century is due to greenhouse forcing as opposed to natural forcing and internal variability; secondly, how significant,

Correspondence to: M. N. Juckes
(m.n.juckes@rl.ac.uk)

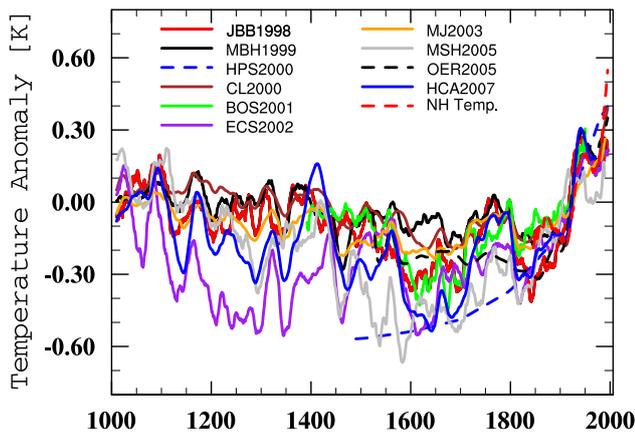


Fig. 1. Proxy based temperature reconstructions from AD1000 to present for various regions on hemispheric to global scales: see text for details. With mean of 1900 to 1960 removed, 21-year running means. NH mean instrumental temperatures are shown for the period AD1866 to 1995.

compared to past natural changes, are the changes which we now observe and expect in the future?

Uncertainty in the answer to the first question remains because of uncertainty about the precise amplitude of climate response to a given forcing and uncertainty about the amplitude of internal variability on centennial time scales (e.g. Hegerl et al., 2007b). The second question reflects the uncertainty in the response of the climate system to a given change in forcing. In the last century both the variations in forcing and the variations in response have been measured with some detail, yet there remains uncertainty about the contribution of natural variability to the observed temperature fluctuations. In both cases, investigation is hampered by the fact that estimates of global mean temperature based on reliable direct measurements are only available from 1856 onwards (Jones et al., 1986).

Climate models are instrumental in addressing both questions, but they are still burdened with some level of uncertainty and there is a need for more detailed knowledge of the behaviour of the actual climate on multi-centennial timescales both in order to evaluate the climate models and in order to address the above questions directly.

The scientific basis for proxy based climate reconstructions may be stated simply as follows: there are a number of physical indicators which contain information about the past environmental variability. As these are not direct measurements, the term proxy is used. Jones and Mann (2004) review proxy evidence for climate change in the past two millennia and conclude that the 20th century has seen the greatest temperature change within any century in this period. The coolest centuries appear to have been the 15th, 17th, and 19th centuries. See also Jansen et al. (2007).

Previous studies have examined the varying response of proxies to temperatures in different seasons. Here we adopt

a simpler approach and seek only to reconstruct the Northern Hemisphere annual mean temperature. We select two methods out of a wide range used in the literature: one which weights all proxies equally and a second which weights them in proportion to their correlation with the annual mean temperature in the calibration period. The Northern Hemisphere temperature record (HadCRUT2v) from Jones and Moberg (2003) is used¹.

This paper reviews different contributions and evaluates the impact of different methods and different data collections used. Section 2 discusses recent contributions, which have developed a range of new methods to address aspects of the problem. Section 3 presents some new results using data collections from 5 recent studies. Section 4 discusses general criticisms of millennial temperature reconstructions raised by McIntyre and McKittrick (2003) (hereafter MM2003).

Appendices provide information on the regression techniques and statistical tests used.

2 A survey of recent reconstructions

This section gives brief reviews of recent contributions, displayed in Fig. 1. Of these, 4 are estimates of the Northern Hemisphere mean temperature (JBB1998, MBH1999, CL2000, MSH2005), 4 of the Northern Hemisphere extra tropical mean temperature (HPS2000, BOS2001, ECS2002, HCA2007) and 2 of the global mean temperature (MJ2003, OER2005). HPS2000 and BOS2001 represent land temperatures only. The ECS2002 curve is based on extratropical data but has here been calibrated to match the variance of the hemispheric mean temperature. There are also differences with respect to the seasons represented: the BOS2001 reconstruction represents growing season temperatures (April to September). The graph labeled HCA2007 is their “CH-blend (Dark ages)” reconstruction representing the mean over land areas north of 30° N. All, except the inherently low resolution reconstructions of HPS2000 and OER2005, have been smoothed with a 21 year running mean. The time series in Fig. 1 have been centred to have zero mean in the AD 1900 to 1960 period. The ECS2002 series was obtained as an uncalibrated index and has here been scaled (by 1.73) to match the variance of Northern Hemisphere temperature in the period 1856 to 1980. These differences in the region and seasons represented are likely to contribute to the difference between the reconstructions. With the exception of HPS2000 and OER2005, the reconstructions use partly overlapping data, so their errors are not statistically independent. In addition, the above works also use a range of techniques. The subsections below cover different scientific themes, ordered according to the date of key publications. Some reconstructions which do

¹downloaded from <http://www.cru.uea.ac.uk/ftpdata/tavenh2v.dat>, the data extends to 2005 using the methodology of Jones and Moberg (2003)

not extend all the way back to AD 1000 are included because of their importance in addressing specific issues.

2.1 High-resolution paleoclimate records

Jones et al. (1998) [JBB1998] present the first annually resolved hemispheric reconstructions of temperatures back to AD 1000, using a composite of 10 standardised proxies for the Northern Hemisphere and 7 for the southern, with variance damped in the early part of the series to account for the lower numbers of proxies present (6 series extend back to AD 1000, 3 in each hemisphere) following Osborn et al. (1997). The composites are scaled by variance matching (Appendix A) against the hemispheric annual mean summer temperatures for 1901–1950. An evaluation of each individual proxy series against instrumental data from 1881 to 1980 shows that tree-rings and documentary reconstructions are more closely related to temperature than those from coral and ice-cores.

With regard to the temperatures of the last millennium, the primary conclusion of JBB1998 is that the twentieth century was the warmest of the millennium. There is clear evidence of a cool period from 1500 to 1900, but no strong “Medieval Warm Period” [MWP] (though the second warmest century in the Northern Hemisphere reconstruction is the 11th). The MWP is discussed further in Sect. 2.4 below.

JBB1998 also raise concerns about the homogeneity of some of the proxies on longer timescales (see Sect. 2.5 below). This is an important issue, since many climate reconstructions (all those reviewed here except HPS2000) rely on a constant relationship between temperature anomalies and the proxy indicators.

2.2 Climate field reconstruction

Mann et al. (1999) published the first reconstruction of the last thousand years northern hemispheric mean temperature which included objective error bars, based on the analysis of the residuals in the calibration period. The authors concluded not only that their estimate of the temperature over the whole period AD 1000 to AD 1860 was colder than the late twentieth century with “moderately high levels of confidence”. The methods they used were presented in MBH1998 which described a reconstruction back to AD 1400.

MBH1998 use a collection of 415 proxy indicators, many more than used in Jones et al. (1998), but many of these are too close geographically to be considered as independent, so they are combined into a smaller number of representative series. The number of proxies also decreases significantly back in time: only 22 independent proxies extend back to AD 1400, and, in MBH1999, 12 extend back to AD 1000 (7 in the Northern Hemisphere). MBH1998 and MBH1999 have been the subject of much debate (see Sect. 4) since the

latter was cited in IPCC2001, though the IPCC conclusions² were based on several reconstructions and were weaker than those of MBH1999 (i.e., they included an element of expert assessment of the robustness of the MBH1999 results).

MBH1998,1999 also differ from Jones et al. (1998) in using spatial patterns of temperature variability rather than a hemispheric mean temperature time series. The aim is to exploit proxies which respond to climate anomalies (e.g. rainfall changes) and so might have an indirect link to temperature changes even when there is no direct link with local temperatures.

Different modes of atmospheric variability are evaluated through an Empirical Orthogonal Function [EOF] analysis of the time period 1902 to 1980, expressing the global field as a sum of spatial patterns (the EOFs) multiplied by Principal Components (PCs – representing the temporal evolution). Earlier instrumental data are too sparse to be used for this purpose: instead they are used in a validation calculation to determine how many EOFs should be included in the reconstruction. Time series for each mode of variability are then reconstructed from the proxy data using a least squares inverse regression followed by variance matching. The skill of the regression of each PC is tested using the 1856 to 1901 validation data. Prior to 1450 AD it is determined that only one PC can be reconstructed with any accuracy.

The reconstructed temperature evolution (Fig. 1) is rather less variable than that of Jones et al. (1998). The most substantial differences occur in the 17th and 19th centuries, when the MBH1999 reconstruction is about 0.3 K warmer than that of JBB1998. The overall picture is of gradual cooling until the mid 19th century, followed by rapid warming.

2.3 Borehole temperatures

Huang et al. (2000) [HPS2000] estimate Northern Hemisphere temperatures back to 1500 AD using measurements made in 453 boreholes (their paper also presents global and Southern Hemisphere results using an additional 163 Southern Hemisphere boreholes). The reconstruction is included here, even though it does not extend back to AD 1000, because it has the advantage of being completely independent of the other reconstructions shown. Temperature fluctuations at the surface propagate slowly downwards, so that

²Folland et al. (2001) concluded (closely following the wording of MBH1999) that “The 1990s are likely to have been the warmest decade of the millennium in the Northern Hemisphere, and 1998 is likely to have been the warmest year.” Unlike MBH1999, they attached a precise meaning to “likely”: a greater than 66% probability. Since 2001 it has been recognised that there is a need to explicitly distinguish between, on the one hand, the simple citation of results of statistical test and, on the other hand, an expression of confidence, as made by the IPCC in this quote, which should include expert assessment of the robustness of statistical methods employed and remaining uncertainties, in addition to results of statistical tests (Manning et al., 2004).

measurements made in the boreholes at depth contain a record of past surface temperature fluctuations. HPS2000 used measurements down to around 300 m. The diffuse nature of the temperature anomaly means that short time scale fluctuations cannot be resolved (Clow, 1992).

The impact of changes in snow cover has been discussed by Mann et al. (2003c); Pollack and Smerdon (2004); González-Rouco et al. (2003); Huang (2004); Chapman et al. (2004); Rutherford and Mann (2004); Bartlett et al. (2005); this appears to be a minor effect. In Huang (2004) an expanded network of 696 boreholes in the Northern Hemisphere is used.

The overall picture of HPS2000 is of an accelerated warming trend over the last 500 years, with the 16th century being about 0.5 K below the 1900 to 1960 average. The 16th century cooling trend seen in JBB1998 and MBH1999, when smoothed to centennial resolution, is not present in HPS2000. It should be noted that the technique used to generate the borehole estimate (Pollack et al., 1998) assumes a constant temperature prior to AD 1500. The absence of a cooling trend after this date may be influenced by this boundary condition.

2.4 Medieval Warm Period

Despite much discussion (e.g. Hughes and Diaz, 1994; Bradley et al., 2003), there is no clear quantitative understanding of what is meant by the “Medieval Warm Period” (MWP). Crowley and Lowery (2000) [CL2000] discuss the evidence for a global MWP, which they interpret as a period of unusual warmth in the 11th century. All the reconstructions of the 11th century temperature shown in Fig. 1 estimate it to have been warmer than most of the past millennium. However, a question of more practical importance is not whether it was warmer than the 12th to 19th centuries, which is generally accepted, but whether it was a period of comparable warmth to the late 20th century. MBH1999 concluded, with 95% confidence, that this was not so. CL2000 revisit the question using 15 proxy records (7 annually resolved, 3 with decadal scale variability and 5 with only centennial temporal resolution). The low-resolution (decadal and centennial) series and 3 of the high-resolution series used by CL2000 were not used in the studies cited above.

CL2000 draw attention to the spatial localization of the MWP in their proxy series: it is strong in North America, North Atlantic and Western Europe, but not clearly present elsewhere. Periods of unusual warmth do occur in other regions, but these are short and asynchronous.

Their estimate of northern hemispheric temperature over the past millennium (see Fig. 1) is close to that of MBH1999. They conclude that the occurrence of decades of temperatures similar to those of the late 20th century cannot be unequivocally ruled out, but that there is, on the other hand, no evidence to support the claims that such an extended period

of large-scale warmth occurred (see also Osborn and Briffa, 2006).

MJ2003 extend the study period to the last 1800 years using a combination of low and high resolution proxies and compositing followed by variance matching (CVM: see Appendix A) with regional sub-composites and principal components to deal with local oversampling. They conclude that the late 20th century warmth is unprecedented in the last two millennia (the latter thousand years of their global reconstruction is included in Fig. 1).

Soon and Baliunas (2003) carry out an analysis of local climate reconstructions. They determine the number of such reconstructions which show (a) a sustained climate anomaly during AD 800–1300, (b) a sustained climate anomaly during AD 1300–1900 and (c) their most anomalous 50 year period in the 20th century. Their definition of a “sustained climate anomaly” is 50 years of warmth, wetness or dryness for (a) and (c) and 50 years of coolness, wetness or dryness in (b). It should be noted that they do not carry out evaluations which allow direct comparison between the 20th century and earlier times: they compare the number of extremes occurring in the 20th century with the number of anomalies occurring in periods of 3 and 4 centuries in the past. Both the use of sampling periods of differing length and different selection criteria make interpretation of their results problematic. They have also been criticised for interpreting regional extremes which occur at distinct times as being indicative of global climate extremes (Jones and Mann, 2004). This issue is discussed further in Sect. 2.9 below (see also Mann et al., 2003a; Soon et al., 2003; Mann et al., 2003b). Osborn and Briffa (2006) perform a more rigorous and quantitative analysis along the lines of Soon and Baliunas (2003), using a method that by-passes the problem of proxy calibration against instrumental temperatures, and conclude that the proxy records alone show an unprecedented synchronous anomaly in the 20th century.

2.5 Segment length curse

Briffa et al. (2001) and Briffa et al. (2002) discuss the impact of the “segment length curse” (Cook et al., 1995; Briffa et al., 1996; Briffa, 2000) on temperature reconstructions from tree rings. Tree ring chronologies are often made up of composites of many trees of different ages at one site. The width of the annual growth ring depends not only on environmental factors but also on the age of the tree. The age dependency on growth is often removed from the tree ring data for each tree by subtraction of or, more usually, by division by an empirical growth curve (Fritts, 1976; Cook and Peters, 1997). This process can remove environmental trends which span the life of the tree along with age related trends. Briffa et al. (2001) use a more sophisticated method (Age Band Decomposition [ABD], which forms separate chronologies from tree rings in different age bands, and then averages all the age-band chronologies) to construct Northern Hemisphere

temperatures back to AD 1400, and show that a greater degree of long term variability is preserved (they also use density data, rather the tree ring width). The reconstruction lies between those of JBB1998 and MBH1999, showing the cold 17th century of the former, but the relatively mild 19th century of the latter.

The potential impact of the segment length limitations is analysed further by Esper et al. (2002, 2003), using “Regional Curve Standardisation” (RCS) (Briffa et al., 1992). The RCS method uses growth curves (different curves reflecting different categories of growth behaviour) obtained by compositing data from all the trees in a region with respect to age and then applying some smoothing. Whereas ABD circumvents the need to use a growth curve, RCS seeks to evaluate a growth curve which is not contaminated by climate signals. The ECS2002 analysis agrees well with that of MBH1999 (and others) on short time scales, but has greater variability on longer timescales (Esper et al., 2004). ECS2002 suggest that this may be partly due to the lack of tropical proxies in their work, which they suggest should be regarded as an extratropical Northern Hemisphere estimate which would be expected to reflect the greater variability of the extratropics relative to the tropics. However, it should also be noted that among the proxies used by MBH1999 which extend back to AD 1000 (12 in total), only 2 are located in the tropics, both at one location (see Table 1 below).

Cook et al. (2004) study the data used by ECS2002 and pay particular attention to potential loss of quality in the earlier parts of tree-ring chronologies when a relatively small number of tree samples are available. Their analysis suggests that reconstructions solely based on tree-rings should be treated with caution prior to AD 1200.

2.6 Separating timescales

Moberg et al. (2005) [MSH2005] follow BOS2001 and ECS2002 in trying to address the “segment length curse”, but, rather than trying to improve the standardization of tree-ring chronologies, they discard the low frequency components of the tree-ring data and replace them with information from proxies with lower temporal resolution. A wavelet analysis is used to filter different temporal scales.

Each individual proxy series is first scaled to unit variance and then wavelet transformed. Averaging of the wavelet transforms is made separately for tree ring data and the low-resolution data. The average wavelet transform of tree-ring data for timescales less than 80 years is combined with the averaged wavelet transform of the low-resolution data for timescales longer than 80 years to form one single wavelet transform covering all timescales. This composite wavelet transform is inverted to create a dimensionless temperature reconstruction, which is calibrated against the instrumental record of Northern Hemisphere mean temperatures, AD 1856–1979, using a variance matching method.

Unfortunately, the calibration period is too short to independently calibrate the low frequency component. The variance matching thus represents a form of cross-calibration. In all calibrations against instrumental data, the long period (multi-centennial) response is determined by a calibration which is dominated by sub-centennial variance. The MSH2005 approach makes this explicit and shows a level of centennial variability which is much larger than in MBH1999 reconstruction and similar to that in ECS2002 and also in simulations of the past millennium with two different climate models, ECHO-G (von Storch et al., 2004) and NCAR CSM (“Climate System Model”) (Mann et al., 2005). However, Mann et al. (2005) criticize MSH2005 and argue that their method can lead to inflated low-frequency variance at the cost of deflated high-frequency variance. This criticism is not yet responded to.

2.7 Glacial advance and retreat

Oerlemans (2005) provides another independent estimate of the global mean temperature over the last 460 years from an analysis of glacial advance and retreat. As with the borehole based estimate of HPS2000, this work uses a physically based model rather than an empirical calibration (though the physical model is highly simplified and relies heavily on empirically determined coefficients). The resulting curve lies within the range spanned by the high-resolution proxies, roughly midway between the MBH1999 Climate Field Reconstruction and the HPS2000 borehole estimate.

Unlike the borehole estimate, but consistent with most other works presented here, this analysis shows a cooling trend prior to 1850, related to glacial advances over that period.

2.8 Regression techniques

Many of the reconstructions listed above depend on empirical relationships between proxy records and temperature. von Storch et al. (2004) suggest that the regression technique used by MBH1999 under-estimates the centennial variability³ of past climate. This conclusion is drawn after applying a method similar to that of MBH1999, though using detrended data for calibration, to output from a climate model using a set of pseudo-proxies: time series generated from the model output and degraded with noise which is intended to match the noise characteristics of actual proxies. Mann et al. (2005) also test reconstruction techniques against climate model simulation, though using different reconstruction techniques, and found no evidence of systematic under-representation of low-frequency variability. The debate is ongoing (Wahl and Ammann, 2007; Esper et al., 2005; Rutherford et al., 2005; Bürger and Cubasch, 2005; Bürger et al., 2006; von Storch et al., 2006; Mann et al., 2007).

³Note, however, that the MBH1998, 1999 papers provide no formal estimate of variability.

Table 1. A selection of millennial temperature proxy records used in previous studies and used in the analysis of Sect. 4. The primary reference for each data set is indicated by the superscript in the first column as follows: 1: (Dahl-Jensen et al., 1998), 2: Yang et al. (2002), 3: Shiyatov (1993), 4: Grudd et al. (2002), 5: Gupta et al. (2003), 6: see Lloyd and Graumlich (1997), 7: Tan et al. (2003), 8: Thompson (1992), 9: Bartholin and Karlén (1983), 10: Naurzbaev and Vaganov (1999), 11: Briffa et al. (1992), 12: Hantemirov and Shiyatov (2002), 13: Briffa et al. (1995), 14: Lara and Villalba (1993), 15: Fisher et al. (1996), 16: Boninsegna (1992), 17: Cook et al. (1991), 18: Cronin et al. (2003), 19: Chbouki (1992), 20: Ferguson and Graybill (1983), 21: Hughes and Funkhouser (2003), 22: LaMarche (1974), 23: Serre-Bachet et al. (1991), 24: MSH2005. In column 4, three-letter combinations indicate publications in which the data were used (JBB1998, MBH1999, ECS2002, MSH2005 and HCA2005) with the year omitted for brevity. The 13 proxy series marked with a “*” are used in the “Union” reconstruction. The type of proxy is indicated in column 6: tree-ring [TR], tree-ring composite [TR C], tree-ring principle component [TR PC], sediment [SE], ice core [IC], multi-proxy composite [MC] and speleothem [SP]. *R* is the anomaly correlation between the proxy record (interpolated to annual resolution where necessary) and the Northern Hemisphere temperature record over the period AD1856 to 1980 (†: the ITRDB:FRAN010 series was filled using persistence for this calculation).

Name	Lat.	Lon.	Id	<i>R</i>	Type
GRIP: borehole temperature (degC) (Greenland) ¹	73	-38	*,MSH	0.67	[IC]
China: composite (degC) ²	30	105	*,MSH	0.63	[MC]
Taymir (Russia)	72	102	HCA	0.60	[TR C]
Eastern Asia	35	110	HCA	0.58	[TR C]
Polar Urals ³	67	65	ECS	0.51	[TR]
Torneträsk (Sweden) ⁴	68	19	MSH	0.50	[TR]
ITRDB [pc01]	40	-110	MBH	0.49	[TR PC]
Mongolia	50	100	HCA	0.46	[TR C]
Arabian Sea: Globigerina bulloides ⁵	18	58	*,MSH	0.45	[SE]
Western Siberia	60	60	HCA	0.44	[TR C]
Northern Norway	65	15	HCA	0.44	[TR C]
Upper Wright (USA) ⁶	38	-119	*,ECS	0.43	[TR]
Shihua Cave: layer thickness (degC) (China) ⁷	40	116	*,MSH	0.42	[SP]
Western Greenland	75	-45	HCA	0.40	[IC]
Quelccaya 2 [$\delta^{18}\text{O}$] (Peru) ⁸	-14	-71	*,MBH	0.37	[IC]
Boreal (USA) ⁶	35	-118	*,ECS	0.32	[TR]
Torneträsk (Sweden) ⁹	68	19	ECS	0.31	[TR]
Taymir (Russia) ¹⁰	72	102	*,ECS, MSH	0.30	[TR]
Fennoscandia ¹¹	68	19	*,JBB, MBH	0.28	[TR]
Yamal (Russia) ¹²	70	70	MSH	0.28	[TR]
Northern Urals (Russia) ¹³	67	65	*,JBB, MBH	0.27	[TR]
ITRDB [pc02]	42	-108	MBH	0.21	[TR PC]
Lenca (Chile) ¹⁴	-41	-72	JBB	0.18	[TR]
Crete (Greenland) ¹⁵	71	-36	*,JBB	0.16	[IC]
Greenland stack ¹⁵	77	-60	MBH	0.13	[IC]
Morocco, Col du Zad, [ITRDB:MORC014] ¹⁹	33	-5	*,MBH	0.13	[TR]
North Patagonia ¹⁶	-38	-68	MBH	0.08	[TR]
Indian Garden (USA) [ITRDB:NV515] ²⁴	39	-115	MSH	0.04	[TR]
Tasmania ¹⁷	-43	148	MBH	0.04	[TR]
ITRDB [pc03]	44	-105	MBH	-0.03	[TR PC]
Chesapeake Bay: Mg/Ca (degC) (USA) ¹⁸	38	-76	*,MSH	-0.07	[SE]
Quelccaya 2 [accum] (Peru) ⁸	-14	-71	MBH	-0.14	[IC]
France [ITRDB:FRAN010] ²³	44	7	MBH	-0.17 [†]	[TR]

There is some uncertainty about the true nature of the noise, both on the proxies and on the instrumental record. The inverse regression technique of MBH1998 effectively neglects the uncertainties in the proxy data relative to uncertainties in the temperature (Appendix A, Eq. 1). HCA2007 use total least squares regression (Allen and Stott, 2003; Ad-

cock, 1878), which accounts for the uncertainty due to unknown noise in proxy data, subject to the assumption that the instrumental noise is known. HCA2007 show that this approach leads to greater variability in the reconstruction.

Rutherford et al. (2005) compare reconstructions from AD 1400 to present using a regularised expectation

maximisation technique (Schneider, 2001) and the MBH1998 climate field reconstruction method and find only minor differences. Standard regression techniques assume that we have a calibration period, in which both sets of variables are measured, and a reconstruction (or prediction) period in which one variable is estimated, by regression, from the other. The climate reconstruction problem is more complex: there are hundreds of instrumental records which are all of different lengths, and similar numbers of proxy records, also of varying length. The expectation maximisation technique (Little and Rubin, 1987) is well suited to deal with this: instead of imposing an artificial separation between a calibration period and a reconstruction period, it fills in the gaps in a way which exploits all data present. Regularised expectation maximisation is a generalisation developed by Schneider (2001) to deal with ill posed problems. Nevertheless, there is still a simple regression equation at the heart of the technique.

2.9 Natural variability and forcings

Global temperature can fluctuate through internally generated variability of the climate system (as in the El Niño phenomenon), through variability in natural forcings (solar insolation, volcanic aerosols, natural changes to greenhouse gas concentrations) and human changes. Crowley (2000) reconstructed variations in the external forcings for the last millennium which have been widely used, though recent studies have suggested a lower amplitude of low-frequency solar forcing variability (Lean et al., 2002; Foukal et al., 2004).

Crowley (2000) concluded that changes in the reconstructed solar and volcanic forcings can explain the reconstructed long term cooling through most of the millennium and the warming in the late 19th century seen in the temperature reconstructions of MBH1999 and CL2000. In addition, deforestation may have contributed to the relatively cool climate in the second half of the 19th century (Bauer et al., 2003). Hegerl et al. (2003, 2006) analyse the relation between temperature reconstructions (MBH1999, CL2000, BOS2001 and ECS2002 in the first paper, BOS2001, ECS2002, MJ2003 and HCA2007 in the second) and estimated forcings (Crowley, 2000) using multiple regression. In Hegerl et al. (2006) it is concluded that natural forcings, particularly volcanism, explain a substantial fraction of hemispheric temperature variability on decadal and longer timescales. Greenhouse gas forcing is detectable with high significance levels in all reconstructions analysed there. Weber (2005) also analyses the relation between reconstructions (JBB1998, MBH1999, Briffa (2000), CL2000, ECS2002, MJ2003 and MSH2005) and forcings, but a timescale dependent analysis is used rather than multiple regression. It is shown that the timescale dependence of northern hemispheric temperatures on the forcings found from reconstructions is similar to that found in a climate model. The role of solar forcing is found to be larger for

longer timescales, whereas volcanic forcing dominates for decadal timescales. However, in Weber (2005), it is found that the trend component over the period 1000 to 1850 is, in all reconstructions, larger than the trend implied by the forcings.

The methods employed by Hegerl et al. (2006) attribute about a third of the early 20th century warming, sometimes more, in high-variance reconstructions, to greenhouse gas forcing. These results indicate that enhanced variability in the past does not make it more difficult to detect greenhouse warming, since a large fraction of the variability can be attributed to external forcing. Quantifying the influence of external forcing on the proxy records is therefore more relevant to understanding climate variability and its causes than determining if past periods were possibly as warm as the 20th century.

Goosse et al. (2005) investigate the role of internal variability using an ensemble of 25 simulations of the last millennium with a low resolution, intermediate-complexity AOGCM, using forcing estimates from Crowley (2000). They conclude that internal variability dominates local and regional scale temperature anomalies, implying that most of the variations experienced by a region such as Europe over the last millennium could have been caused by internal variability. On the hemispheric and global scale, however, the forcing dominates. This agrees with results from a long solar-forced model simulation (using the same model) by Weber et al. (2004). Goosse et al. (2005) make the new point, that natural variability can lead to regional temperature anomalies peaking at different times to the forcing, so that disagreements in timing between proxy series should not necessarily be interpreted as meaning there is no common forcing.

2.10 The long view

The past sections have drawn attention to the problems of calibrating temperature reconstructions using a relatively short period over which instrumental records are available. For longer reconstructions, with lower temporal resolution, other methods are available. Pollen reconstructions of climate match the ecosystem types with those currently occurring at different latitudes or different climate regimes (i.e., a spatial rather than temporal calibration). The changes in ecosystem can then be mapped to the temperatures at which they now occur (e.g. Bernabo, 1981; Gajewski, 1988; Davis et al., 2003). These reconstructions cannot resolve decadal variability, but they provide an independent estimate of local low-frequency temperature variations. The results of Weber et al. (2004) and Goosse et al. (2005) suggest that such estimates of regional mean temperatures can provide some information about global mean anomalies, as they strongly reflect the external forcings on centennial and longer timescales. The Goosse et al. results also suggest that hemispheric, extra-tropical and global reconstructions should be strongly

correlated on the centennial time-scale. Pollen records were also included in the CL2000 and MSH2005 reconstructions (one each), but there has, as yet, been no detailed intercomparison between the pollen based reconstructions and the higher resolution reconstructions.

3 Varying methods vs. varying data

One factor which complicates the evaluation of the various reconstructions is that different authors have varied both method and data collections. Here we will run a representative set of proxy data collections through two algorithms: inverse regression and scaled composites. These two methods are explained in Appendix A.

Esper et al. (2005) investigated the differing calibration approaches used in the recent literature, including regression and scaling techniques, and concluded that the methodological differences in calibration result in differences in the reconstructed temperature amplitude/variance of about 0.5 K. This magnitude is equivalent to the mean annual temperature change for the Northern Hemisphere reported in the IPCC 2001 report for the 1000–1998 period. Bürger and Cubasch (2005) take another approach and investigate a family of 64 different regression algorithms derived by adjusting 5 binary switches, using pseudo-proxy data (see also Bürger et al., 2006). They show that these choices, which have all been defended in the literature, can lead to a wide variety of different reconstructions given the same data. They also point out that the uncertainty is greater when we attempt to calibrate proxies measurements which are outside the range occurring in the calibration period. In our reconstruction (see below), the proxy composite has its maximum in the calibration period but the minimum, occurring in the 17th century, is well outside the calibration period range.

Several authors have evaluated composites and calibrated those composites against instrumental temperature. Many of the composites contain more samples in later periods, so that the calibration may be dominated by samples which do not extend into the distant past. Here, we will restrict attention to records which span the entire reconstruction period from to AD 1000 to AD 1980 (with some series ending slightly earlier, as discussed below).

Evaluation of past work is further complicated by the use of different versions of proxy series in different studies, even though they may appear to originate from the same locations or to have been produced from the same primary data sets. In Table 1 there are two series referred to as Torneträsk: that used by ECS2002 is based on a subset of ring-width data used by MSH2005, and they were also processed to produce average tree-ring chronologies using different statistical standardisation methods. The “Fennoscandian” data used in both JBB1998 and MBH1999 also represent the Torneträsk region but these studies actually used a previously published temperature reconstruction (i.e., Briffa et al., 1992) that com-

bined ring-width and ring-density predictors from the same trees used by ECS2002. The Polar Urals series used by ECS2002 and MBH1999 and the Northern Urals series used by JBB1998 are all derived from the same site, but ECS2002 reprocessed only the ring-width data, whereas MBH1999 and JBB1998 again used a linear combination of the ring-width and ring-density series from these trees that had been used to produce a temperature reconstruction (Briffa et al., 1995). The Taymir (sometime referred to as Taimyr) data used by HCA2007 is a smoothed version of a chronology produced by ECS2002, whereas the high-pass version of this series was used in MSH2005. The Greenland stack data used by MBH1999 is a composite of data analysed by Fisher et al. (1996), but the precise nature of the composite is not described by Fisher et al. (1996).

3.1 Reconstruction using a union of proxy collections

The following subsection will discuss a range of reconstructions using different data collections. The first 5 of these collections are defined as those proxies used by JBB1998, MBH1999, ECS2002, MSH2005 and HCA2007, respectively, which extend from AD 1000 to 1980 (1960 for HCA2007). These will be referred to below as, respectively, the JBB, MBH, ECS, MSH, HCA composites to distinguish them from the composites used in the published articles, which include, in some cases, additional, shorter, proxy data series. The Indigirka series used by MSH2005 is not used here because it is not available for unrestricted use. The Sargasso Sea series used by MSH2005 is sometimes mistakenly presented as having a 1975 data point representing the 1950–2000 mean, but the actual end point is 1925 and so this series is also omitted.

Finally there is a “Union” composite made using 13 independent Northern Hemisphere proxy series marked with “*” in Table 1. This “Union” collection contains 6 tree-ring series, 3 ice-cores, one speleothem record, two sediment records and a composite record including historical data.

The series in the “Union” composite have been chosen on the basis that they extend to 1980 (the HCA composites and the French tree ring series end earlier). The Southern Hemisphere series have been omitted apart from the $\delta^{18}\text{O}$ series from Quelccaya glacier, Peru, which is included to improve representation of tropical temperatures. The Quelccaya accumulation data series is omitted, as the $\delta^{18}\text{O}$ series is expected, on physical grounds, to have better temperature sensitivity. The MBH1999 North American PCs have been omitted in favour of individual series used in other studies. The Polar Urals data of ECS2002, MBH1999 have been omitted in favour of data from the same site used by JBB1998. The Yamal data, used by MSH2005, have also been omitted because of the proximity of the site to the Northern Urals site of JBB1998. The Indian Garden series is omitted because of inadequate documentation. The Crete ice core series is preferred to the “stack” series used by MBH1999 because it

is better documented. Torneträsk is omitted because of its high correlation (0.89) and close proximity to the JBB1998 Northern Fennoscandia series.

The composite is intended not only to average out regional anomalies but also to average out errors which might be associated with particular proxies or sets of proxies. It is clear that the proxies are affected by factors other than temperature which are not fully understood. We are carrying out a uni-variate analysis which, by construction, treats all factors other than the one predicted variable as noise.

It should be emphasised here that our selection of proxies is based on physically grounded expectations that they have a temperature signal, not on statistical evaluation. We do not need to assume that individual proxies have a signal to noise ratio greater than unity. The standard deviation of the observations in the calibration period is 0.27 K. A collection of 13 proxies with a signal to noise ratio, in this period, of 2 would, if independent, give a reconstruction standard error of $2 \times 0.27 \text{ K} / \sqrt{13} = 0.15 \text{ K}$, which would be good enough to provide scientifically useful information. While it may appear desirable to reject proxies which have a poor correlation with temperature, such data selection will tend to bias the statistical estimates.

3.2 Intercomparison of proxy collections

Figure 2 shows reconstructions back to AD 1000 using composites of proxies and variance matching (CVM: see Appendix A) (for the proxy principal components in the MBH collection the sign is arbitrary: these series have, where necessary, had the sign reversed so that they have a positive correlation with the Northern Hemisphere temperature record). The “Union” lies mostly in the range spanned by the other reconstructions, though towards and sometimes below the bottom of that range. Its temperature range is about as large as in MSH and ECS. The JBB and MBH collections give a smaller range, while HCA takes an intermediate position. The “Union”, however, fits the calibration period data better than any of the sub-collections (Table 2). This analysis reveals that the choice of proxy records is one reason why different reconstructions show different ranges of temperature variability.

The reconstructions shown in Fig. 3 use the same data as in Fig. 2; this time using inverse regression of the proxy records against hemispheric mean data [INVR] (Appendix A). As mentioned earlier, MBH1998, 1999 also used inverse regression, but the method used here differs from that of MBH1998, 1999 in using Northern Hemisphere temperature to calibrate against, having a longer calibration period, and reconstructing only a single variable instead of multiple PCs. MBH1998 also rescaled their final reconstruction using variance matching. The spread of values is substantially increased relative to the CVM reconstruction and all INVR reconstructions show larger temperature ranges than their CVM counterparts.

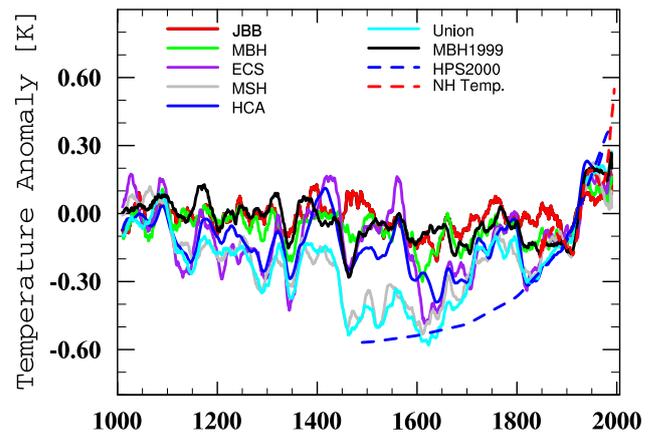


Fig. 2. Reconstruction back to AD 1000, calibrated on 1856 to 1980 Northern Hemisphere temperature, using CVM, for a variety of different data collections. The MBH1999 and HPS2000 NH reconstructions and the Jones and Moberg (2003) instrumental data are shown for comparison. Graphs have been smoothed with a 21-year running mean and centred on 1866 to 1970.

With INVR, only one reconstruction (ECS) shows pre-industrial temperatures warmer than the mid 20th century. The inverse regression technique applies weights to the individual proxies which are proportional to the correlation between the proxies and the calibration temperature signature. For this time series the 5 proxies are weighted as: 1.5 (Boreal); 1.8 (Polar Urals); 1.7 (Taymir); 1.3 (Torneträsk); and 2.3 (Upper Wright) (after standardising the series to unit variance on the whole reconstruction period). Firstly, it should be noted that this collection samples North America and the Eurasian arctic only. The bias towards a small area is strengthened by the weights generated by the inverse regression algorithm, such that nearly half the signal comes from two series located within a few degrees of each other in North America.

The MBH1999 and HPS2000 published reconstructions are shown in Figs. 2, 3 for comparison: the MBH1999 reconstruction often lies near the centre or upper bound of the spread of estimates, while the HPS2000 reconstruction is generally at the lower bound (note, however, that it represents more variable extratropical land temperatures).

Much of the current debate revolves around the level of centennial scale variability in the past. The CVM results for JBB and MBH correspond to a low variance scenario, comparable to MBH1999, while other proxy collections suggest substantially larger variability. The INVR reconstructions have greater variability, as expected on theoretical grounds (Bürger et al., 2006). It should be noted that the MBH1999 inverse regression result use greater volumes of data for recent centuries, so that the difference in Fig. 3 between the dashed black curve and the full green curve in the 17th century may be due to reduced proxy data input in the latter (there is also a difference because MBH1999 used inverse

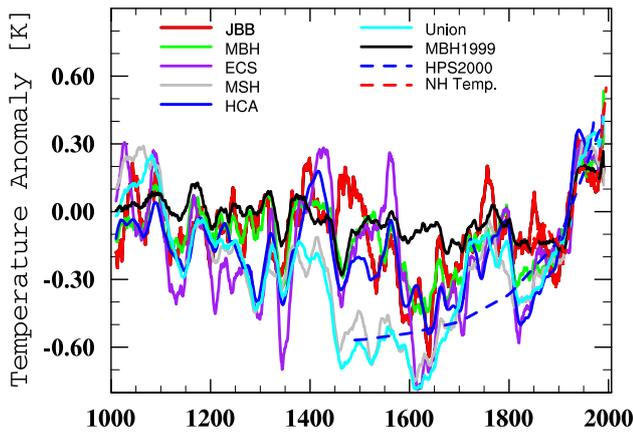


Fig. 3. As Fig. 2, except using inverse regression.

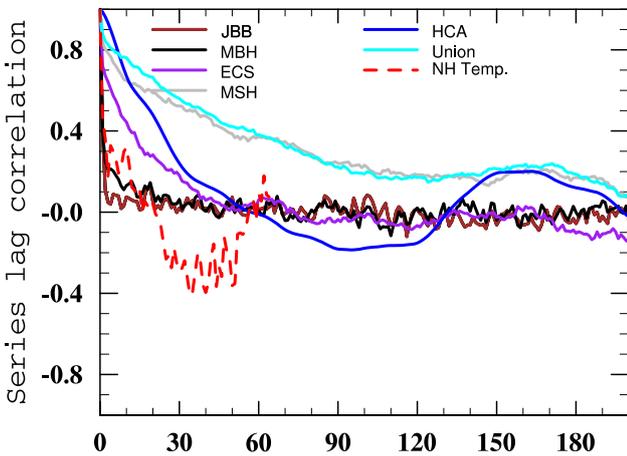


Fig. 4. Auto-correlations for proxy composites and instrumental record (red).

regression against temperature principle components rather than Northern Hemisphere mean temperature as here).

3.3 Validation

Table 3 shows the cross correlations of the reconstructions in Fig. 2, for high pass (upper right) and low pass (lower left) components of the series, with low pass being defined by a 21 year running mean. The low pass components are highly correlated.

The significance of the correlations between these six proxy data samples and the instrumental temperature data during the calibration period (1856–1980) has been evaluated using 10 000 member Monte-Carlo simulations with (1) a first order Markov model (e.g. Grinstead and Snell, 1997) with the same 1-year lag correlation as the data samples and (2) random time series which reproduces the autocorrelation of the data samples (see Appendix B2). Figure 4 shows the autocorrelations of the reconstructions and of the instrumen-

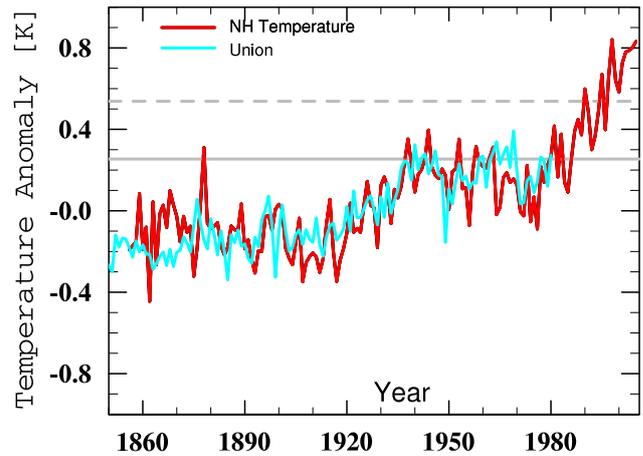


Fig. 5. The “Union” reconstruction (plotted for AD 1850 to 1980), using “composite plus variance scaling” (cyan) and the Northern Hemisphere instrumental temperature (AD 1856 to 2005; red). The pre-industrial maximum of the reconstruction is shown by a gray line, and that value plus 2σ as a gray dashed line.

tal record. The latter has a pronounced anti-correlation on the 40 year time-scale which may be an artifact of the short data record (autocorrelations calculated from reconstructions truncated to the calibration period show a similar anti-correlation around this timescale, supporting this interpretation; see supplementary material: <http://www.clim-past.net/3/591/2007/cp-3-591-2007-supplement.zip>). The MSH, HCA and “Union” composites show multi-centennial correlations which are not present in the other data. The JBB composite have low decadal scale correlations relative to the observational record, while the MSH, HCA and “Union” composites are high.

The 10 000 R values obtained by correlating the reconstructions obtained from random data with the Northern Hemisphere mean temperature are then used to estimate the significance of that obtained from the data.

Results from the significance calculations are shown in Table 2. If the full autocorrelation of the data were known, it would be true, as argued by McIntyre and McKittrick (2005a), that the first order Markov approach generally leads to an overestimate of significance. Here, however, we have only an estimated correlation structure based on a small sample. Using this finite sample correlation is likely to overestimate the amplitude of long-term correlations and hence lead to an underestimate of significance. Nevertheless, results are presented here to provide a cautious estimate of significance. For the JBB composite, which has a short autocorrelation, the difference between the two methods is minimal. For other composites there is a substantial difference. In all cases the R values for CVM reconstructions exceed the 98% significance level. When detrended data are used the R values are generally lower, but still above the 96% level. The Hegerl et al. data has only decadal resolution, so the lower significance

Table 2. Calibration period residual standard deviation (σ ; columns 2 and 9), R values (columns 3, 6, 10, 13) and significance estimates (columns 4, 5, 7, 8, 11, 12, 14, 15) for proxy collections using CVM (columns 2 to 8) and INVR (columns 9 to 15). All R values are evaluated against the Northern Hemisphere mean temperature (1856 to 1980), together with significance estimates. Significance estimates are evaluated using Monte Carlo simulations with 10,000 realisations, using the Hosking method (Appendix B2). Columns 3, 4, 5, 10, 11 and 12 are evaluated using the full time series, columns 6, 7, 8, 13, 14 and 15 are evaluated from detrended time series.

	CVM							INVR						
	σ [K]	With trend			Detrended			σ [K]	With trend			Detrended		
		R	S_h	S_f	R_{detr}	S_h	S_f		R	S_h	S_f	R_{detr}	S_h	S_f
MBH	0.174	0.535	99.93	99.99	0.306	99.08	99.79	0.229	0.661	99.90	99.99	0.421	94.29	98.04
MSH	0.158	0.610	98.68	99.84	0.379	98.70	98.04	0.305	0.676	93.12	99.88	0.412	85.05	93.39
ECS	0.161	0.599	99.47	99.99	0.389	98.50	99.79	0.302	0.593	95.45	99.93	0.363	84.92	95.59
JBB	0.203	0.367	99.61	99.92	0.263	99.17	99.22	0.392	0.433	95.73	99.53	0.274	82.43	88.42
HCA	0.151	0.649	98.12	97.81	0.406	96.85	99.78	0.265	0.640	62.84	56.99	0.382	40.70	48.85
Union	0.142	0.688	99.91	99.99	0.435	99.59	99.35	0.185	0.698	94.35	99.67	0.485	93.22	97.97

in high frequency variability is to be expected. Many of the INVR reconstructions, however, fail to pass the 95% significance level. The CVM Union reconstruction exceeds 99% significance both for full and detrended R values and for both variants of the Monte Carlo test.

Figure 5 plots the Union reconstruction, with the instrumental data in the calibration period. The composite tracks the changes in Northern Hemisphere temperature well, capturing the steep rise between 1910 and 1950 and much of the decadal scale variability. This is reflected in the significance scores (Table 2) which are high both for the full series and for the detrended series.

Concerns have also been raised about the Bristlecone Pines (which have an anomalously large positive growth anomaly in the 20th century, possibly due to CO₂ fertilisation (see Sect. 4) – e.g. the Indian Garden series in Table 1) and the high latitude Eurasian trees (which have an anomalously low growth anomaly in the late 20th century – e.g. Torneträsk, Fennoscandia, Yamal, Northern Urals in Table 1: Briffa et al., 1998; Wilmking et al., 2005; D’Arrigo et al., 2006). As noted earlier, the presence of factors other than temperature affecting proxy indicators is unavoidable. We need, however, to ensure that the reconstruction is not sensitive to omission of any one series, as such sensitivity would make the reconstruction strongly dependent on factors affecting that series. The robustness of the Union reconstruction has been tested by creating a family of 13 reconstructions each omitting one member of the proxy collection. The standard deviation of the maximum pre-industrial temperature in this ensemble of 13 reconstructions is 0.027 K (the standard deviation of individual years is larger, 0.048 K)⁴. The two

⁴That is, the standard deviation of $\{T_i^{\max} : i=1, 13\}$ is 0.027 K, where T_i^{\max} is the maximum pre-industrial temperature of the i th reconstruction, and the standard deviation of $\{T_{i;k} : i=1, 13; k=1000, 1850\}$ is 0.048 K, where $T_{i;k}$ is the temperature of the i th reconstruction in year k .

Table 3. Cross correlations (times 100) between CVM reconstructions from different proxy data bases (see Sect. 4.1). Lower left block correspond to low pass filtered series (21 day mean), upper right to high pass filtered. Note that the HCA proxies are smoothed composites, so the high-pass signal in the reconstruction is reduced.

	MBH	MSH	ECS	JBB	HCA	Union
MBH	–	14	28	63	22	57
MSH	62	–	28	12	7	31
ECS	71	58	–	13	34	61
JBB	67	46	47	–	13	36
HCA	76	69	85	54	–	26
Union	68	96	63	49	76	–

series whose omission has the largest impact are the Grip borehole data and the Arabian Sea sediment record, but in both cases the impact is well within the uncertainty range estimated from the calibration period residual (0.14 K).

In order to verify that the standard deviation is a fair measure of variations among the 13 members of this family we have also applied a Pearson’s chi-squared test to the distribution of the 13 pre-industrial maxima and found no significant departure from a Gaussian distribution.⁵

As a further check on the proxies, the correlation between each of the 13 series used in the “union” and the composite of the remaining 12 was evaluated. These are shown in Table 4. In each case the significance of the correlation was also evaluated against an ensemble of random curves with the same autocorrelation as the proxies (see Appendix B). 7 of the 13 tests exceed 95% significance and a further 3 exceed

⁵Pearson’s chi-squared was applied using 4 cells defined by the quartiles of a Gaussian with matching standard deviation. This gave a value of 3.3 which is exceeded in a chi-squared distribution with 3 degrees of freedom with a probability of 35%.

Table 4. Correlation between each proxy in the Union composite and the composite of the remaining 12 (column 2), the significance of that correlation (column 3) and the maximum pre-industrial temperature in the reconstruction omitting that proxy minus the maximum pre-industrial temperature in the union reconstruction.

Name	<i>R</i>	Significance	Max – Max _U
GRIP: borehole temperature (degC) (Greenland)	0.3424	84.0%	–0.065 K
Chesapeake Bay: Mg/Ca (degC) (USA)	0.2168	94.7%	0.022 K
Shihua Cave: layer thickness (degC) (China)	0.4936	99.6%	0.009 K
China: composite (degC)	0.6340	99.8%	0.002 K
Arabian Sea: Globigerina bulloides	0.5314	98.9%	0.058 K
Boreal (USA)	0.3117	97.3%	–0.010 K
Taymir (Russia)	0.2891	99.9%	–0.007 K
Upper Wright (USA)	0.4007	99.9%	–0.026 K
Northern Fennoscandia	0.2087	99.6%	0.011 K
Northern Urals (Russia)	0.1557	94.8%	–0.006 K
Crete (Greenland)	0.0485	88.9%	0.010 K
Morocco	0.0315	78.3%	0.002 K
Quelccaya 2 [do18] (Peru)	0.1492	92.1%	0.000 K

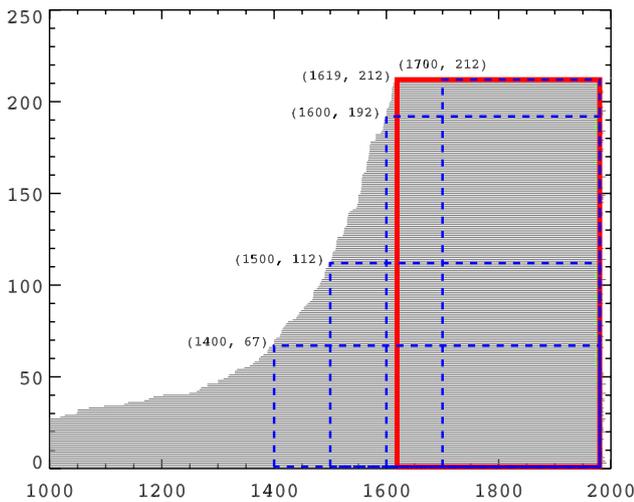


Fig. 6. Data blocks for PC calculation of North American tree ring sites by MBH1998. Each of the 212 data series is shown as a horizontal line over the time period covered. The dashed blue rectangles indicate some of the blocks of data used by MBH1998 for their proxy principal component calculation, using fewer series for longer time periods. The red rectangle indicates the single block used by MM2003, neglecting all data prior to 1619.

90%. Thus, it is clear that there is a common signal in this disparate collection of time-series.

3.4 Discussion

The significance tests clearly show that the CVM Union reconstruction has a better statistical relationship to the calibration temperature than the other data collections and that the CVM method generally gives better results than the INVR method.

The CVM Union reconstruction has a cold period from AD 1450 to AD 1700, close to the borehole estimates of HPS2000. As noted in Sect. 3.2, the difference between the Union reconstruction and the MBH1999 reconstruction in this period could be influenced by the greater number of proxies used by MBH1999 for this period of their reconstruction.

The highest temperature in the reconstructed data, relative to the 1866–1970 mean is 0.25 K in the 11th century. This temperature was first exceeded in the instrumental record in 1878, again in 1938 and frequently thereafter. The instrumental record has not gone below this level since 1986. Taking $\sigma=0.14$ K, the root-mean-square residual in the calibration period, 1990 is the first year when the reconstructed pre-industrial maximum was exceeded by 2σ . This happened again in 1995 and every year since 1997. 1998 and every year since 2001 have exceeded the pre-industrial maximum by 3σ . Two recent years (1998 and 2005) have exceeded the pre-industrial estimated maximum by more than 4σ .

4 The debate on the methodology of millennial temperature reconstructions

This section addresses the criticism of the Mann et al. (1998, 1999) work, which have dominated the public debate since the appearance of the IPCC 2001 report which featured Mann et al. (1999) in its Summary. Although many of the issues are not directly relevant to the results presented here, there are suggestions about the soundness of the whole methodology of large-scale temperature reconstructions which need to be addressed. Some of the critics identify the IPCC consensus with the MBH1998 work. However, this is not the case: IPCC2001 contains a broader body of work on this subject,

and its conclusions are less strong than those of MBH1998 (Sect. 2.2).

McIntyre and McKittrick (2003) [MM2003] criticize MBH1998 on many counts, some related to deficiencies in the description of the data used and possible irregularities in the data themselves. These issues have been largely resolved in Mann et al. (2004). In addition, they report that they are unable to reproduce the MBH1998 reconstruction. Indeed, the stepwise reconstruction method was first fully described in Mann et al. (2004). There are 4 main stages of the algorithm: (1) sub-sampling of regions with disproportionate numbers of proxies, (2) EOF decomposition of the instrumental global temperature record, (3) regression, (4) validation and uncertainty estimates. For example, Fig. 6 illustrates how the stepwise approach applies to the North American tree ring network. Of the total of 212 chronologies, only 66 extend back beyond 1400 AD. MM2003 only calculate principal components for the period when all chronologies are present (see also supplement: <http://www.clim-past.net/3/591/2007/cp-3-591-2007-supplement.zip>, Sect. 3). Also, data from South West Mexico was omitted in MM2003 (discussed further in the SM). McIntyre and McKittrick (2005b) [MM2005b] revisit the MM2003 work using the stepwise calculated principal components. They claim to reproduce the results of MM2003 by making only minor changes to the MBH1998 proxy data base. However, examination of the relevant figures shows that the MM2005b predictions for the 15th century are 0.3 K warmer than the MBH1998 result as opposed to 0.5 K in MM2003. This is still significant, but, unlike the discredited MM2003 result, it would not make the 15th century the warmest on record.

Two more fundamental methodological issues were raised in the literature, one relates to the impact of the standardisation of the proxy records prior to principal component calculation and the other to the validity of specific proxy data (bristlecone pines) as indicators of past temperature variations. The first issue, which was raised by McIntyre and McKittrick (2005a), is discussed extensively by von Storch and Zorita (2005); McIntyre and McKittrick (2005c); Huybers (2005); McIntyre and McKittrick (2005d); Wahl and Ammann (2007). In Fig. 7 we illustrate this point using a reconstruction back to AD 1000 based on the MBH1999 data and the INVR method as described in Sect. 3. Curve (1) in Fig. 7 follows MBH1999, including an empirical adjustment of the first North American PC to compensate for a hypothetical CO₂ fertilisation effect. The warmth of the MBH1999 published reconstruction after AD1500 relative to the curve (1) could result from the use of greater volumes of data for the more recent periods. Standardisation on the whole series rather than just the calibration period is seen to have little impact on the reconstruction (curve (4) versus curve (3)), confirming earlier results based on model data (von Storch and Zorita, 2005) and the full MBH1998 data collection (Wahl and Ammann, 2007). The empirical correction of the CO₂ fertilisation effect does, however, have a major effect (curve

(3) versus curves (1,2)). Further issues raised by MM2005b are discussed in the supplement: <http://www.clim-past.net/3/591/2007/cp-3-591-2007-supplement.zip>, Sect. 4. Indeed, it is found that impact of the empirical adjustment of the leading PC (reported in MBH1999) is considerably greater than impact of the changes suggested by MM2005b. It is shown that these issues are not directly relevant to the “Union” reconstruction presented in the previous section.

Rising CO₂ levels may have contributed significantly to the 19th and 20th century increase in growth rate in some trees, particularly the bristlecone pines (Briffa and Osborn (1999) and MM2005b). Such a CO₂ fertilisation has been measured in saplings and strip-bark orange trees (which were well watered and fertilised) (Graybill and Idso, 1993, and references therein). However, efforts to reproduce the effect in controlled experiments with mature forest trees in natural conditions (Körner et al., 2005) have not produced positive results. The cause of the dramatic growth in the MBH1998 “North American proxy PC1” is therefore not clear. MM2005b suggest that this PC1 is a statistical outlier as far as its correlation to Northern Hemispheric temperature T_{nh} is concerned, but this is in fact not the case, as shown in Table 1. MM2005c and Wahl and Ammann (2007) both find that excluding the north American bristlecone pine data from the proxy data base removes the skill from their 15th century reconstructions. Note, however, that reconstructions from different data collections which have no bristlecone pine data, discussed in Sect. 3, do have skill.

5 Conclusions

There is general agreement that the warmest pre-industrial temperatures of the last thousand years occurred at the start of millennium, and the coolest at some point during the 16th to 19th centuries. There is also general consensus that the warmest pre-industrial temperatures were close to the mean of the 20th century. There remains, however, a range of views about the strength of the negative anomaly in the 16th to 19th centuries: estimates of the Northern Hemisphere mean temperature in this period range from no more than 0.2 K cooler than the 1900 to 1960 mean to as much as 0.6 K cooler.

The IPCC2001 conclusion that temperatures of the past millennium are unlikely to have been as warm, at any time prior to the 20th century, as the last decades of the 20th century is supported by subsequent research and by the results obtained here. We have also reviewed and, in some cases, tested with new analysis, papers (in particular Soon and Balinas, 2003, MM2003 and MM2005b) which claim to refute that IPCC2001 conclusion and found that those claims were not well supported. The IPCC 2007 conclusion that “It is very likely that average NH temperatures during the second half of the 20th century were warmer than any other 50-year period in the last 500 years and likely the warmest in at least

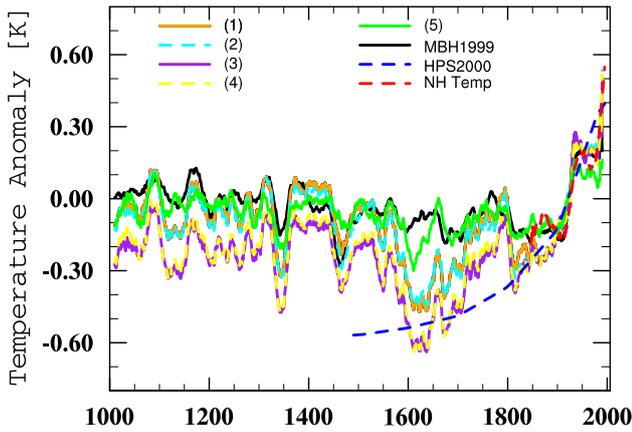


Fig. 7. (1) to (5): Reconstructions back to 1000, calibrated on 1856 to 1980 Northern Hemisphere temperature, using the MBH1999 proxy data collection. Also shown are: (6) the MBH1999 NH reconstruction, (7) the HPS2000 reconstruction and (8) the the Jones and Moberg (2003) instrumental data. All data have been smoothed with a 21-year running mean and centred on 1866 to 1970 (the time period covered by both the smoothed instrumental record and the smoothed reconstructions). (1) to (4) are constructed using the INVR technique and differ in the details of the proxies: (1) uses the proxy series as used by MBH1999, including an adjustment to the first Principal Component of the North American tree-rings and filling, with persistence, of the last 7 years of the French tree-ring series (ITRDB:FRAN010), (2) omits the French tree-ring series, (3) as (2), but also using unadjusted PCs, (4) as (2), using recalculated PCs based on 25 series out of the 28 used by MBH1999, omitting 3 that required filling, and standardising on the whole series rather than the calibration period. (5) uses the same data as (3) with the CVM procedure.

the past 1300 years” (Solomon et al., 2007) is also supported by our analysis.

A major area of uncertainty concerns the accuracy of the long time-scale variability in the reconstructions. This is particularly so for timescales of a century and longer. There are fewer concerns related to the ability of the proxy records to capture more rapid changes on a 10 to 50 year time scale, such as we have experienced in recent decades.

Using two different reconstruction methods on a range of proxy data collections, we have found that inverse regression local proxies against hemispheric mean temperature tends to give large weighting to a small number of proxies and that the relatively simple approach of compositing all the series and using variance matching to calibrate the result gives more robust estimates. The choice of proxy records is one reason why different reconstructions show different temperature ranges.

A new reconstruction made with a composite of 13 proxies extending back to AD 1000 fits the instrumental record to within a standard deviation of 0.14 K. The maximum pre-industrial temperature in the reconstruction is 0.25 K, in

the 11th century, relative to the AD 1866 to 1970 mean. The maximum temperature from the instrumental record is 0.84 K in AD 1998, over 4 standard deviations larger.

The reconstructions evaluated in this study show considerable disagreement during the 16th century. The new 13 proxy reconstruction implies 21-year mean temperatures close to 0.6K below the AD 1866 to 1970 mean. As this reconstruction only used data extending back to AD 1000, there is a considerable volume of 16th century data which has not been used. This will be a focus in future research. Further information can be found on mitrie.badc.rl.ac.uk/.

Appendix A

Calibration methods

Ideally, the statistical analysis method would be determined by the known characteristics of the problem. Unfortunately, the error characteristics of the proxy data are not sufficiently well quantified to make the choice clear. This appendix describes two methods used in this study.

A1 Inverse regression [INVR]

Suppose x_{ik} , $i=1, N_{pr}$, $k=1, L$ is a set of N_{pr} standardised proxy records of length L and that we are trying to obtain an estimate \hat{y}_k of a quantity y_k which is known only in a calibration period (C).

Several “optimal” estimates of y_i can be obtained, depending on the hypothesised relation between the proxies and y .

Inverse regression follows from the model

$$\beta_i y_k + \mathcal{N}_k = x_{ik}, \quad (\text{A1})$$

where \mathcal{N} is a noise process, independent between different k and independent of y_k (and hence, by construction, correlated with x_{ik}). This contrasts with the forward regression which would follow from an assumption that \mathcal{N} is independent of x_{ik} . It is also assumed that \mathcal{N} has the same variance and autocorrelation for each proxy. It follows that optimal estimate for the coefficients β_i are:

$$\hat{\beta}_i = \frac{\sum_{k \in C} x_{ik} y_k}{\sum_{k \in C} y_k^2}. \quad (\text{A2})$$

Given these coefficients, the optimal estimate (minimising the expected error variance) of the y_k outside the calibration period is:

$$\hat{y}_k = \frac{\sum_i \hat{\beta}_i x_{ik}}{\sum_i \hat{\beta}_i^2}. \quad (\text{A3})$$

A2 Composite plus variance matching [CVM]

This method starts out from the hypothesis that different proxies represent different parts of the globe. A proxy for

the global mean is then obtained as a simple average of the proxies:

$$\bar{x}_k = N_{pr}^{-1} \sum_i x_{ik}. \quad (\text{A4})$$

The forward regression technique damps variance (Bürger et al., 2006), while the inverse regression model amplifies variance. A compromise between these two, for uni-variate regression (provided $\overline{x'y'} > 0$), is the variance matching approach:

$$\hat{y}_k = \bar{x}_k \left(\frac{\sigma_y}{\sigma_x} \right). \quad (\text{A5})$$

Appendix B

Statistical tests

B1 Tests for linear relationships

The simplest test for a linear relationship is the anomaly correlation (also known as: Pearson Correlation, Pearson's product moment correlation, R , product mean test):

$$R = \frac{\overline{y'x'}}{\sqrt{\overline{y'^2} \overline{x'^2}}} \quad (\text{B1})$$

where the over-bar represents a mean over the data that the test is being applied to, and a prime a departure from the mean (Pearson, 1896).

B2 Significance tests

Significance calculations depend on having an accurate estimate of the number of degrees of freedom, n , in a sample. Ideally, if the noise affecting all the x and y values is independent, n is simply the number of measurements. This is unlikely to be the case, so an estimate of n is needed. The Monte-Carlo approach is more flexible: a large sample of random sequences with specified auto-correlations is created, and the frequency with which the specified R coefficient is exceeded can then be used to estimate its significance.

In this study two sets of Monte-Carlo simulations have been carried out: firstly, using a first order Markov approximation to the auto-correlation. In the second, following Hosking (1984), random time series with the specified auto-correlations of the proxies or composite, as appropriate, are obtained from the partial correlation coefficients, which are generated using Levinson-Durbin regression.

It is, however, not possible to generate a sequence matching an arbitrarily specified correlation structure and there is no guarantee that an estimate of the correlation structure obtained from a small sample will be realizable. It is found that the Levinson-Durbin regression diverges when run with the auto-correlation from the HCA composite. This is resolved by truncating the regression after 100 years for this

series. The sample autocorrelation coefficients are, in any case, unreliable beyond this point. Truncating the regression results in a random sequence with an autocorrelation fitting that specified up to the truncation point and then decaying.

The random time series thus generated are passed through precisely the same algorithms used to generate the reconstructions in order to test the significance of those reconstructions.

Appendix C

Acronyms and abbreviations

Table C1 shows a list of acronyms and abbreviations used in this paper.

Table C1. Acronyms and abbreviations used in the text.

ABD	Age Band Decomposition tree ring standardisation method
AOGCM	coupled Atmosphere-Ocean General Circulation Model
CSM	Climate System Model: A coupled ocean-atmosphere climate model produced by NCAR, http://www.cgd.ucar.edu/csm/
CFM	Climate Field Reconstruction: method for reconstructing spatial structures of past climate variables using proxy data
CVM	Composite plus Variance Matching reconstruction method
ECHO-G	Hamburg coupled ocean-atmosphere climate model
EOF	Empirical Orthogonal Function
INVR	Inverse Regression reconstruction method
IPCC	The Intergovernmental Panel on Climate Change, established by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) to assess scientific, technical and socio-economic information relevant for the understanding of climate change, its potential impacts and options for adaptation and mitigation. It is open to all Members of the UN and of WMO.
ITRDB	International Tree-Ring Data Bank, maintained by the NOAA Paleoclimatology Program and World Data Center for Paleoclimatology (www.ncdc.noaa.gov/paleo)
MWP	Medieval Warm Period
NCAR	National Center for Atmospheric Research (http://www.ncar.ucar.edu)
PC	Principal Component
RCS	Regional Curve Standardisation tree ring standardisation method

Acknowledgements. This work was funded by the Netherlands Environment Assessment Agency (RIVM) as part of the Dutch Scientific Assessment and Policy Analysis (WAB) programme. Additional funding was provided as follows: from the UK Natural Environment Research Council for M. N. Juckes, K. R. Briffa and T. J. Osborn; from the Swedish Research Council for A. Moberg. We are grateful to E. Zorita for extensive discussion and criticism during the preparation of this paper and we thank G. Bürger and two anonymous referees for detailed constructive criticism of this paper.

Edited by: H. Goosse

References

- Adcock, R. J.: A problem in least squares, *The Analyst*, 5, 53–54, 1878.
- Allen, M. R. and Stott, P. A.: Estimating signal amplitudes in optimal fingerprinting, Part I: Theory, *Clim. Dynam.*, 21, 477–491, 2003.
- Arrhenius, S.: On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground, *Philosophical Magazine*, 41, 237–276, 1896.
- Bartholin, T. and Karlén, W.: Dendrokronologi i Lapland, *Dendrokronologiska Sällskapets Meddelanden*, 4, 3–16, 1983.
- Bartlett, M., Chapman, D. S., and Harris, R. N.: Snow effect on North American ground temperatures 1950–2002, *J. Geophys. Res.*, 110, F03 008, doi:10.1029/2005JF000293, 2005.
- Bauer, E., Claussen, M., Brovkin, V., and Huenerbein, A.: Assessing climate forcings of the Earth system for the past millennium, *Geophys. Res. Lett.*, 30(6), 1276, doi:10.1029/2002GL016639, 2003.
- Bernabo, J. C.: Quantitative estimates of temperature changes over the last 2700 years in Michigan based on pollen data, *Quat. Res.*, 15, 143–159, 1981.
- Boninsegna, J. A.: South American dendroclimatological records, *Climate Since A.D. 1500*, edited by: Bradley, R. S. and Jones, P. D., Routledge, 446–462, 1992.
- Bradley, R. S., Hughes, M. K., and Diaz, H. F.: Climate in medieval time, *Climate in medieval time*, *Science*, 302, 404–405, 2003.
- Briffa, K. R.: Annual climate variability in the Holocene: interpreting the message of ancient trees, *Quart. Sci. Rev.*, 19, 87–105, 2000.
- Briffa, K. R. and Osborn, T. J.: Seeing the wood from the trees, *Science*, 284, 926–927, 1999.
- Briffa, K. R., Jones, P. D., Bartholin, T. S., Eckstein, D., Schweingruber, F. H., Karlén, W., Zetterberg, P., and Eronen, M.: Fennoscandian summers from A.D. 500: temperature changes on short and long time scales, *Clim. Dynam.*, 7, 111–119, 1992.
- Briffa, K. R., Jones, P. D., Schweingruber, F. H., Shiyatov, S. G., and Cook, E. R.: Unusual twentieth-century summer warmth in a 1,000-year temperature record from Siberia, *Nature*, 376, 156–159, 1995.
- Briffa, K. R., Jones, P. D., Schweingruber, F. H., Karlen, W., and Shiyatov, S. G.: Tree-ring variables as proxy-climate indicators: Problems with low-frequency signals, *NATO ASI Ser. Ser. I*, vol. 41, edited by: Jones, P. D., Bradley, R. S., and Jouzel, J., pp. 9–41, Springer-Verlag, New York, 1996.
- Briffa, K. R., Schweingruber, F. H., Jones, P. D., Osborn, T. J., Shiyatov, S. G., and Vaganov, E. A.: Reduced sensitivity of recent tree-growth to temperature at high northern latitudes, *Nature*, 391, 678–682, 1998.
- Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: Low-frequency temperature variations from a northern tree ring density network, *J. Geophys. Res.*, 106(D3), 2929–2941, doi:2000JD900617, 2001.
- Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Jones, P. D., Shiyatov, S. G., and Vaganov, E. A.: Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals, *The Holocene*, 12, 737–757, 2002.
- Bürger, G. and Cubasch, U.: Are multiproxy climate reconstructions robust?, *Geophys. Res. Lett.*, 32, L23 711, doi:10.1029/2005GL024155, 2005.
- Bürger, G., Fast, I., and Cubasch, U.: Climate reconstruction by regression - 32 variations on a theme, *Tellus A*, 58(2), 227–235, 2006.
- Chapman, D. S., Bartlett, M. G., and Harris, R. N.: Comment on “Ground vs. surface air temperature trends: Implications for borehole surface temperature reconstructions”, by Mann, M. and Schmidt, G., *Geophys. Res. Lett.*, 31(7), L07205, doi:10.1029/2003GL019054, 2004.
- Chbouki, N.: Spatio-temporal characteristics of drought as inferred from tree-ring data in Morocco, Ph.D. thesis, University of Arizona, Tucson, Arizona, 1992.
- Clow, G.: Temporal resolution of surface-temperature histories inferred from borehole temperature measurements, *Global Planet. Change*, 98, 81–86, 1992.
- Cook, E. and Peters, K.: Calculating unbiased tree-ring indices for the study of climatic and environmental change, *The Holocene*, 7, 361–370, 1997.
- Cook, E. R., Bird, T., Peterson, M., Barbetti, M., Buckley, B., D’Arrigo, R., Francey, R., and Tans, P.: Climatic change in Tasmania inferred from a 1089-year tree-ring chronology of sub-alpine huon pine, *Science*, 253, 1266–1268, 1991.
- Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. S., and Funkhouser, G.: The ‘segment length curse’ in long tree-ring chronology development for paleoclimatic studies, *The Holocene*, 5(2), 229–237, 1995.
- Cook, E. R., Esper, J., and D’Arrigo, R. D.: Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years, *Quat. Sci. Rev.*, 23, 2063–2074, 2004.
- Cronin, T. M., Dwyer, G. S., Kamiya, T., Schwede, S., and Willard, D. A.: Medieval Warm Period, *Global Planet. Change*, 36, 17–29, 2003.
- Crowley, T. J.: Causes of climate change over the past 1000 years, *Science*, 289, 270–277, 2000.
- Crowley, T. J. and Lowery, T.: How Warm Was the Medieval Warm Period? A comment on “Man-made versus natural climate change”, *Ambio*, 29, 51–54, 2000.
- Dahl-Jensen, D., Mosegaard, K., Gundestrup, N., Clow, G. D., Johnsen, S. J., Hansen, A. W., and Baling, N.: Past temperatures directly from the Greenland Ice Sheet, *Science*, 282, 268–271, 1998.
- D’Arrigo, R., Wilson, R., and Jacoby, G.: On the long-term context for late twentieth century warming, *J. Geophys. Res.*, 111, D03103, doi:10.1029/2005JD006352, 2006.
- Davis, B., Brewer, S., Stevenson, A., and Guiot, J.: The temperature of Europe during the Holocene reconstructed from pollen data,

- Quat. Sci. Rev., 22, 1701–1716, 2003.
- Esper, J., Cook, E. R., and Schweingruber, F. H.: Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, *Science*, 295, 2250–2253, 2002.
- Esper, J., Cook, E. R., Krusic, P. J., Peters, K., and Schweingruber, F. H.: Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies, *Tree-Ring Res.*, 59, 81–98, 2003.
- Esper, J., Frank, D. C., and Wilson, R. J. S.: Climate reconstructions – low frequency ambition and high frequency ratification, *EOS*, 85, 113, 120, 2004.
- Esper, J., Frank, D. C., Wilson, R. J. S., and Briffa, K. R.: Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophys. Res. Lett.*, 32(7), L07711, doi:10.1029/2004GL021236, 2005.
- Ferguson, C. W. and Graybill, D. A.: Dendrochronology of bristlecone pine: a progress report, *Radiocarbon*, 25, 287–288, 1983.
- Fisher, D. A., Koerner, R. M., Kuivinen, K., Clausen, H. B., Johnsen, S. J., Steffensen, J. P., Gundestrup, N., and Hammer, C. U.: Inter-comparison of ice core and precipitation records from sites in Canada and Greenland over the last 3500 years and over the last few centuries in detail using EOF techniques, *NATO ASI Ser. I*, vol. 41, edited by: Jones, P. D., Bradley, R. S., and Jouzel, J., pp. 297–328, Springer-Verlag, New York, 1996.
- Folland, C., Karl, T., Christy, J., Clarke, R., Gruza, G., Jouzel, J., Mann, M., Oerlemans, J., Salinger, M., and Wang, S.-W.: Observed Climate Variability and Change, in: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Houghton, J. T., Ding, Y., Griggs, D., Noguer, M., van der Linden, P. J., and Xiaosu, D., chap. 6, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2001.
- Foukal, P., North, G., and Wigley, T.: A stellar view on solar variations and climate, *Science*, 306, 68–69, 2004.
- Fritts, H.: *Tree rings and climate*, Academic Press, New York, 1976.
- Gajewski, K.: Late Holocene climate changes in eastern North America estimated from pollen data, *Quat. Res.*, 29, 255–262, 1988.
- González-Rouco, F., von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, *Geophys. Res. Lett.*, 30(21), 2116, doi:10.1029/2003GL018264, 2003.
- Goosse, H., Renssen, H., Timmermann, A., and Bradley, R. S.: Internal and forced climate variability during the last millennium: a model-data comparison using ensemble simulations, *Quat. Sci. Rev.*, 24, 1345–1360, 2005.
- Graybill, D. A. and Idso, S. B.: Detecting the aerial fertilization effect of atmospheric CO₂ enrichment in tree-ring chronologies, *Global Biogeochem. Cycles*, 7, 81–95, 1993.
- Grinstead, C. and Snell, J.: *Introduction to Probability*, American Mathematical Society, Providence, Rhode Island, second revised edn., 1997.
- Grudd, H., Briffa, K. R., Karlén, W., Bartholin, T. S., Jones, P. D., and Kromer, B.: A 7400-year tree-ring chronology in northern Swedish Lapland: natural climatic variability expressed on annual to millennial timescales, *The Holocene*, 12, 657–666, 2002.
- Gupta, A. K., Anderson, D. M., and Overpeck, J. T.: Abrupt changes in the Asian southwest monsoon during the Holocene and their links to the North Atlantic Ocean, *Nature*, 421, 354–357, 2003.
- Hantemirov, R. M. and Shiyatov, S. G. A.: A continuous multi-millennial ring-width chronology in Yamal, northwestern Russia, *The Holocene*, 12, 717–726, 2002.
- Hegerl, G., Crowley, T., Hyde, W., and Frame, D.: Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, 440, 1029–1032, 2006.
- Hegerl, G., Crowley, T., Allen, M., Hyde, W., Pollack, H., Smerdon, J., and Zorita, E.: Detection of human influence on a new, validated, 1500 year temperature reconstruction, *J. Climate*, 20, 650–666, 2007a.
- Hegerl, G., Zwiers, F. W., Braconnot, P., Gillett, N., Luo, Y., Orsini, J. M., Nicholls, N., Penner, J., and Stott, P.: Understanding and Attributing Climate Change, in: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., chap. 10, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007b.
- Hosking, J. R. M.: Modeling persistence in hydrological time series using fractional differencing, *Water Resour. Res.*, 20(12), 1898–1908, 1984.
- Huang, S.: Merging information from different resources for new insight into climate change in the past and future, *Geophys. Res. Lett.*, 31, L13205, doi:10.1029/2004GL019781, 2004.
- Huang, S., Pollack, H. N., and Shen, P. Y.: Temperature trends over the past five centuries reconstructed from borehole temperatures, *Nature*, 403, 756–758, 2000.
- Hughes, M. K. and Diaz, H. F.: Was there a ‘Medieval Warm Period’ and if so, where and when?, *Clim. Change*, 26, 109–142, 1994.
- Hughes, M. K. and Funkhouser, G.: Frequency-dependent climate signal in upper and lower forest border trees in the mountains of the Great Basin, *Clim. Change*, 59, 233–244, 2003.
- Huybers, P.: comment on “Hockey sticks, principal components, and spurious significance” by McIntyre and McKittrick [2005], *Geophys. Res. Lett.*, 32, L20705, doi:10.1029/2005GL023395, 2005.
- Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliessner, B., Peltier, W., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D.: Palaeoclimate, in: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., chap. 6, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Jones, P. D. and Mann, M. E.: Climate over past millennia, *Rev. Geophys.*, 42, RG2002, doi:10.1029/2003RG000143, 2004.
- Jones, P. D. and Moberg, A.: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Climate*, 16, 206–223, 2003.
- Jones, P. D., Wigley, T. M. L., and Wright, P. B.: Global temperature variations between 1861 and 1984, *Nature*, 332, 430–434, 1986.
- Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B.: High-Resolution Paleoclimatic Records for the Last Millennium: In-

- terpretation, Integration and Comparison with Circulation Model Control-Run Temperatures, *Holocene*, 8, 455–471, 1998.
- Körner, C., Asshoff, R., Bignucolo, O., Hättenschwiler, S., Keel, S. G., Peláez-Riedl, S., Pepin, S., Siegwolf, R. T. W., and Zotz, G.: Carbon Flux and Growth in Mature Deciduous Forest Trees Exposed to Elevated CO₂, *Science*, 309, 1360–1362, 2005.
- LaMarche, V. C.: Frequency-dependent relationships between tree-ring series along an ecological gradient and some dendroclimatic implications, *Tree-Ring Bull.*, 34, 1–20, 1974.
- Lara, A. and Villalba, R.: A 3620-year temperature record from *Fitzroya cupressoides* tree rings in southern South America, *Science*, 260, 1104–1106, 1993.
- Lean, J. L., Wang, Y. M., and Sheeley Jr., N. R.: The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate, *Geophys. Res. Lett.*, 29(24), 2224, doi:10.1029/2002GL015880, 2002.
- Little, R. J. A. and Rubin, D. B.: *Statistical analysis with missing data*, John Wiley and Sons, New York, pp. 278, 1987.
- Lloyd, A. H. and Graumlich, L. J.: Holocene dynamics of treeline forests in the Sierra Nevada, *Ecology*, 78, 1199–1210, 1997.
- Mann, M. E. and Jones, P. D.: Global surface temperatures over the past two millennia, *Geophys. Res. Lett.*, 30(15), 1820, doi:10.1029/2003GL017814, 2003.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, 1998.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Northern Hemisphere Temperatures During the Past Millennium: Inferences, Uncertainties, and Limitations, *Geophys. Res. Lett.*, 26(6), 759–762, 1999.
- Mann, M. E., Ammann, C., Bradley, R., Briffa, K., Jones, P., Osborn, T., Crowley, T., Hughes, M., Oppenheimer, M., Overpeck, J., Rutherford, S., Trenberth, K., and Wigley, T.: On past temperatures and anomalous late-20th century warmth, *Eos Trans. AGU*, 84(27), 256–258, 2003a.
- Mann, M. E., Ammann, C., Bradley, R., Briffa, K., Jones, P., Osborn, T., Crowley, T., Hughes, M., Oppenheimer, M., Overpeck, J., Rutherford, S., Trenberth, K., and Wigley, T.: Response [to comments on “On past temperatures and anomalous late-20th century warmth”], *Eos Trans. AGU*, 84(44), 473–474, 2003b.
- Mann, M. E., Rutherford, S., Bradley, R. S., Hughes, M. K., and Keimig, F. T.: Optimal surface temperature reconstructions using terrestrial borehole data, *J. Geophys. Res.*, 108(D7), 4203, doi:10.1029/2002JD002532, 2003c.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Corrigendum: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 430, 105, doi:10.1038/nature02478, 2004.
- Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Testing the fidelity of methods used in proxy-based reconstructions of past climate, *J. Climate*, 18, 4097–4107, 2005.
- Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C.: Robustness of proxy-based climate field reconstruction methods, *J. Geophys. Res.*, 112, D12 109, doi:10.1029/2006JD008272, 2007.
- Manning, M. M., Petit, M., Easterling, D., Murphy, J., Patwardhan, A., Rogner, H.-H., Swart, R., and Yohe, G.: IPCC Workshop on Describing Scientific Uncertainties in Climate Change to Support Analysis of Risk and of Options: Workshop report, Intergovernmental Panel on Climate Change (IPCC), Geneva, 2004.
- McIntyre, S. and McKittrick, R.: Corrections to the Mann et al (1998) Proxy Data Based and Northern Hemispheric Average Temperature Series, *Energy and Environment*, 14, 751–771, 2003.
- McIntyre, S. and McKittrick, R.: Hockey sticks, principal components, and spurious significance, *Geophys. Res. Lett.*, 32, L03 710, doi:10.1029/2004GL02, 2005a.
- McIntyre, S. and McKittrick, R.: The M&M critique of the MBH1998 northern hemisphere climate index: update and implications, *Energy and Environment*, 16, 69–100, 2005b.
- McIntyre, S. and McKittrick, R.: Reply to comment by Huybers on “Hockey sticks, principal components, and spurious significance”, *Geophys. Res. Lett.*, 32, L20 713, doi:10.1029/2005GL023586, 2005c.
- McIntyre, S. and McKittrick, R.: Reply to comment by von Storch and Zorita on ‘Hockey sticks, principal components, and spurious significance’, *Geophys. Res. Lett.*, 32, L20 714, doi:10.1029/2005GL023089, 2005d.
- Mitchell, J. F. B., Karoly, D. J., Hegerl, G. C., Zwiers, F. W., Allen, M. R., and Marengo, J.: Detection of climate change and attribution of causes, in: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Houghton, J. T., Ding, Y., Griggs, D., Noguer, M., van der Linden, P. J., and Xiaosu, D., chap. 6, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2001.
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, 433, 613–617, 2005.
- Naurzbaev, M. M. and Vaganov, E. A.: 1957-year tree-ring chronology of eastern part of Taymir, *Siberian Journal of Ecology*, 6, 159–168, 1999.
- Oerlemans, J. H.: Extracting a Climate Signal from 169 Glacier Records, *Science*, 308, 675–677, 2005.
- Osborn, T. J. and Briffa, K. R.: The spatial extent of 20th-Century warmth in the context of the past 1200 years, *Science*, 313, 841–844, 2006.
- Osborn, T. J., Briffa, K. R., and Jones, P. D.: Adjusting variance for sample-size in tree-ring chronologies and other regional mean timeseries, *Dendrochronologia*, 15, 89–99, 1997.
- Pearson, K.: *Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia*, *Philosophical Transactions of the Royal Society A*, 187, 253–318, 1896.
- Pollack, H. N. and Smerdon, J. E.: Borehole climate reconstructions: Spatial structure and hemispheric averages, *J. Geophys. Res.*, 109(D), 11 106, doi:10.1029/2003JD004163, 2004.
- Pollack, H. N., Huang, S., and Shen, P. Y.: Climate change record in subsurface temperatures: A global perspective, *Science*, 282, 279–281, 1998.
- Rutherford, S. and Mann, M. E.: Correction to “Optimal surface temperature reconstructions using terrestrial borehole data” by Mann et al., *J. Geophys. Res.*, 109, D11107, doi:10.1029/2003JD004290, 2004.
- Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K., and Jones, P. D.: Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, *J. Climate*, 18, 2308–

- 2329, 2005.
- Schneider, T.: Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values, *J. Climate*, 14, 853–871, 2001.
- Serre-Bachet, F., Martinelli, N., Pignatelli, O., Guiot, J., and Tessier, L.: Evolution des températures du Nord-Est de L'Italie depuis 1500 AD. Reconstruction d'après les cernes des arbres, *Dendrochronologia*, 9, 213–219, 1991.
- Shiyatov, S. G.: The upper timberline dynamics during the last 1100 years in the Polar Ural mountains, in: *Oscillations of the Alpine and Polar K Tree Limits in the Holocene*, edited by: Frenzel, B., Eronen, M., Vorren, K., and Glaser, B., pp. 195–203, Fischer, Stuttgart, Germany, 1993.
- Solomon, S., Qin, D., Manning, M., Alley, R., Berntsen, T., Bindoff, N., Chen, Z., Chidthaisong, A., Gregory, J., Hegerl, G., Heimann, M., Hewitson, B., Hoskins, B., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Stocker, T., Whetton, P., Wood, R., and Wratt, D.: Technical Summary, in: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Soon, W. and Baliunas, S.: Proxy climatic and environmental changes over the past 1000 years, *Clim. Res.*, 23, 89–110, 2003.
- Soon, W., Baliunas, S., and Legates, D. R.: Comments on “On past temperatures and anomalous late-20th century warmth”, *Eos Trans. AGU*, 84(44), 473–476, 2003.
- Tan, M., Liu, T., Hou, J., Qin, X., Zhang, H., and Tieying, L.: Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature, *Geophys. Res. Lett.*, 30, 1617, doi:10.1029/2003GL017352, 2003.
- Thompson, L. G.: Ice core evidence from Peru and China, in *Climate Since A.D. 1500*, edited by: R. S. Bradley and P. D. Jones, pp. 517–548, New York, 1992.
- von Storch, H. and Zorita, E.: Comment on ‘Hockey sticks, principal components, and spurious significance’ by S. McIntyre and R. McKittrick, *Geophys. Res. Lett.*, 32, L20701, doi:10.1029/2005GL022753, 2005.
- von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., and Tett, S. F. B.: Reconstructing Past Climate from Noisy Data, *Science*, 2004.
- von Storch, H., Zorita, E., Jones, J. M., González-Rouco, J. F., and Tett, S. F. B.: Response to comment on “Reconstructing past climate from noisy data”, *Science*, 312, 5782, doi:10.1126/science.1121571, 2006.
- Wahl, E. R. and Ammann, C. M.: Robustness of the Mann, Bradley, Hughes Reconstruction of Surface Temperatures: Examination of Criticisms Based on the Nature and Processing of Proxy Climate Evidence, *Climatic Change*, in press, 2007.
- Weber, S. L.: A timescale analysis of the NH temperature response to volcanic and solar forcing in the past millenium, *Clim. Past*, 1, 9–17, 2005, <http://www.clim-past.net/1/9/2005/>.
- Weber, S. L., Crowley, T. J., and der Schrier, G. V.: Solar irradiance forcing of centennial climate variability during the Holocene, *Clim. Dynam.*, 22, 539–553, 2004.
- Wilmking, M., D’Arrigo, R., Jacoby, G. C., and Juday, G. P.: Increased temperature sensitivity and divergent growth trends in circumpolar boreal forests, *Geophys. Res. Lett.*, 32, L15715, doi:10.1029/2005GL023331, 2005.
- Yang, B., Bräuning, A., Johnson, K. R., and Yafeng, S.: General characteristics of temperature variation in China during the last two millennia, *Geophys. Res. Lett.*, 29(9), 1324, doi:10.1029/2001GL014485, 2002.