**Climate
of the Past
Discussions**

Refining error estimates for a millennial temperature reconstruction

M. N. Juckes

STFC, Rutherford Appleton Lab, Chilton, Didcot, OX11 0QX, UK

Correspondence to: M. N. Juckes (martin.juckes@stfc.ac.uk)

Abstract

The statistical uncertainties in a 1000 year Northern Hemisphere mean temperature
reconstruction obtained from 15 proxy chronologies are examined in detail by analysing
the range of estimates obtained from all possible subsets of the proxy collection with
up to 6 proxies omitted. The study is motivated in part by the large range of recently
published reconstructions in the 15th and 16th centuries. The uncertainty estimates
support the conclusions of the 3rd and 4th Intergovernmental Panel on Climate Change
(IPCC) assessment reports concerning the likelihood that temperatures at the end of
the 20th century were likely (greater than 66% confidence) to have been exceptional.
It is also shown that the last ten years to date have been warmer than any decade of
the past millennium with 95% confidence.

1   Introduction

There have been many recent reconstructions of the temperature of the last millennium
(see Jones et al., 2009; Mann et al., 2008; Juckes et al., 2007; Jansen et al., 2007, for
reviews). All show broadly similar shape, with an initial warm period, slow general
decline after that and finally a steep increase in temperatures. There is, however, less
clarity about the degree of uncertainty in the reconstructions and a need for "more
realistic assessments of reconstruction uncertainty" (Jones et al., 2009). The clearest
statements of uncertainty come from IPCC reports. In the 2001 assessment Folland
et al. (2001, hereafter IPCC2001) conclude 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", while (Jansen et al., 2007, hereafter IPCC2007) in
the 2007 assessment conclude that it is "likely that" 1950 to 2000 "was the warmest
period in the last 1.3 kyr". In both the IPCC reports cited above something is considered
"likely" if it is estimated to have a 2 in 3 chance of occurring or having occurred.

Uncertainties in the estimated temperatures arise from many sources, mainly asso-
ciated with the fact that we do not have access to direct measurements and must work

with indirect evidence. This evidence, the proxy data, reflects environmental variations: but the link between the environment and the proxy data is generally not precisely quantified. For this reason it is usual to use statistical relations between climate and proxy data.

A number of authors (Bürger et al., 2006; Lee et al., 2008; von Storch et al., 2009) have used "virtual reality" tests, in which reconstruction methods are assessed by calculating pseudo-reconstructions based on pseudo-proxy records generated from past climates simulated by models. Christiansen et al. (2009) show that the performance of any given reconstruction method can vary substantially for different realisations of the climate, and call into question the reliability of methodological studies which have used only a limited number of model climate realisations and suggest that this explains the sometimes contradictory results from other studies.

In the present paper error estimates are obtained by analysing in detail the effect of omitting data with the Jackknife technique (e.g. Shao and Tu, 1995), and by comparing two reconstructions made with fully independent proxy records over the period 1750 to 1850. If all local/regional records used in a given reconstruction showed synchronous variations with the same amplitude, the uncertainty estimate given by the Jackknife method would be zero. The uncertainty thus reflects inhomogeneities in the proxy data and the extent to which the true NH-mean temperature variations are captured by the retained data. This method is applied to the reconstruction from Juckes et al. (2007, hereafter JAB2007), which used 13 proxies and to a reconstruction using 3 new proxy data series (described in the next section) together 12 of those from the collection used by JAB2007. The geographical distribution of the proxies is shown in Fig. 1.

The outline of this paper is as follows: Section 2 discusses the data used, Sect. 3 investigates the temporal homogeneity of the data. Section 4 presents the reconstruction of the Northern Hemisphere temperature and Sect. 5 evaluates its uncertainty. Section 6 looks specifically at uncertainties of general statements, such as those quoted above from IPCC Assessment Reports.

2633

2 Data

2.1 Proxy climate records

JAB2007 used 13 proxy climate records covering the period 1000 to 1980 AD, 12 in the Northern Hemisphere and one (the Quelcaya glacier record) at 14° S, taken to be indicative of tropical temperatures in both hemispheres.

Here, both the 13 proxy collection of JAB2007, and an extended series using the following additional 3 proxies will be considered:

- A Mongolian tree ring composite record (D'Arrigo et al., 2001),

- a Chinese temperature reconstruction (Ge et al., 2003),

- and a Donard lake sediment record (Moore et al., 2001).

In the expanded proxy collection, the Arabian sea planktonic foraminifera series previously used is omitted: this series was used in Moberg et al. (2005) and JAB2007 in order to enhance spatial coverage, but is at best an indirect indicator of temperature. This leaves an increase from 13 to 15 series, a modest 15% increase.

In a recent study, Mann et al. (2008) assembled a collection of 95 proxies extending back to 1000 AD, of which 79 are in the Northern Hemisphere: these numbers are reduced to 58 and 45 respectively when tree ring chronologies constructed from fewer than 8 samples are omitted. Out of the 45 Northern Hemisphere unscreened series 19 are tree ring chronologies, 16 in North American and 3 in Eurasia. There are 4 ice core series (all from Greenland); 5 from cave stalactites (2 in China, 2 in Costa Rica and one in Scotland); 2 regional temperature reconstructions (both from China); and 15 sediment series, of which one is marine. Mann et al. (2008) then screen for correlations with observed temperatures, which reduces the number of Northern Hemisphere proxies to 19, comparable with the 15 used here.

From this brief survey it can be seen that the amount of data extending back to 1000 AD remains limited, despite significant additions.

2634

Rather than following Mann et al. (2008) in using correlations with temperature to select proxies, the selection here is based on a priori reasoning. As noted in JAB2007, there are two advantages of this approach: the possibility of including poor proxies because there is insufficient data to screen accurately is avoided and not using temperature measurements in the data selection simplifies the uncertainty estimation which is a key part of this work. The first of these gains is of course offset by the possibility of including poor proxies because of insufficient a priori knowledge.

From the 45 Northern Hemisphere proxy records dating back to 1000 AD which are included in the Mann et al. (2008) collection prior to temperature screening, low latitude sediment series which are expected to reflect precipitation variation are omitted. The 2 Costa Rican speleothem series were supplied to Mann et al. (personal communication, 2008): in the absence of a peer reviewed description of the data origins they will not be used here. Speleothems from the Dongge cave (China) and Scotland used by Mann et al. (2008) are believed to be more precipitation sensitive. No additional Greenland ice core or N. American tree ring data has been used here, as these two regions were already well represented in JAB2007.

Ljungqvist (2009, hereafter LJ2009) reviews 71 proxy climate records. His collection includes a long tree record from Solongotyn Davaa in Mongolia, but here we follow Mann et al. (2008) in using the composite series of D'Arrigo et al. (2001). This composite has a correlation of 0.79, over the period AD 1000–1980, with the Solongotyn Davaa series. LJ2009 includes a Taimyr record from Naurzbaev et al. (2002), but not that from Naurzbaev and Vaganov (1999) which was used by Esper et al. (2002), Juckes et al. (2007) and is used here. The correlation between these two Taimyr series is 0.96. The following series used here are not included in the LJ2009 review: Donard lake, Quelcaya (Peru) $^{18}\delta$O record, China composite of Yang et al. (2002), Crete (Greenland) series, and tree ring series from Col du Zad (Morocco), Boreal (USA), Upper Wright (USA), Northern Fennoscandia, and Northern Urals (Russia) (see Appendix A for further details).

2.2 Instrumental data

The HADCRUT3 (Brohan et al., 2006) Northern Hemisphere mean data is used to calibrate the new reconstructions, replacing the HadCRUT2v data from Jones and Moberg (2003) used by JAB2007. The new data extends further back (to 1850 instead of 1856) and has improved (that is, more reliable) uncertainty estimates.

3 Temporal homogeneity of the proxy data

Proxy based reconstructions rely on the assumption that the proxy response to temperature variations in past centuries was close to that observed in the last 150 years. Recent detailed modelling of the response of trees to climate variations (Anchukaitis et al., 2006) reinforces the expectation that the response of tree-rings is related to known physical and biological properties of the trees. The modelling approach implies that, in the absence of other factors, the response to temperature variations in the past would match that in the calibration period, but it cannot test for the absence of other factors.

A complementary approach is to look at the level of consistency between different types of proxies. The mean anomaly correlation between tree and non-tree proxies is defined as follows:

$$c_n = \frac{1}{N_p} \sum_{\substack{i \in \text{trees} \\ j \in \text{non}-\text{trees}}} \overline{p'_i p'_j}^{(n)}$$

where N_p is the number of pairs i,j such that p_i is a tree-ring proxy and p_j is a non-treering proxy and $\overline{p'_i p'_j}^{(n)}$ is the time mean over the time interval n. The prime denotes an anomaly relative to the mean of the 50 year period over which the correlation is being calculated.

c_n has been evaluated over 50 year periods at 10 years intervals, 1000–1049, 1010–1059, etc., and is plotted in Fig. 2. There is substantial variability, and periods for which the correlation is negative. However, the spread appears consistent with random variation about the mean and there is no clear trend: the largest value of c_n occurs around 1450, and the values in the 11th century are comparable with those in the calibration period.

Figure 2 also shows the correlations among the tree ring chronologies (green) and among the non-tree ring chronologies (purple). The correlations between the non-tree ring proxies and the tree ring chronologies are higher than the correlations among the non-tree ring proxies.

This approach will not detect systematic changes that affect all the proxies equally, but it does provide an indication of the level of consistency.

This brief analysis of correlations does not produce any evidence of systematic changes in proxy behaviour.

4 Northern Hemisphere temperature

4.1 The influence of new data

As in JAB2007, the reconstruction is generated by centring the proxies on the calibration period and normalising them to unit variance in that period, forming the composite with all proxies equally weighted, and then scaling by a factor γ_{vm} such that the variance of the scaled composite matches the variance of the instrumental temperature record in the calibration period:

$$\gamma_{vm} = \sqrt{\frac{\overline{T_i'^2}}{\overline{C'^2}}} \tag{1}$$

where the over-line is a mean over the calibration period, and the prime indicates a departure from the mean over that period, T_i is the instrumental temperature and C is the proxy composite.

Figure 3 shows the JAB2007 reconstruction and the new reconstruction using the collection of 15 proxy data series (hereafter referred to as "R15") and the updated and extended instrumental record described in Sect. 2. Also shown is a reconstruction using the JAB2007 proxies and the new instrumental temperature record: it can be seen that this change has little impact. The differences between the R15 reconstruction and that of JAB2007 are well within any reasonable uncertainty estimate, but the difference is highly coherent over the first 7 centuries of the reconstruction. This underlines the fact that uncertainties in proxy based reconstructions are likely to be correlated in time.

4.2 Comparison with independent data

Some studies have followed Mann et al. (2008) in using part of the instrumental record for calibration and withholding part of the record in order to validate the reconstruction or to select from a family of reconstructions. However, shortening the length of the calibration period reduces the accuracy of the calibration. Furthermore, this study uses a number of proxies with low temporal resolution, such that it would be impossible to treat the validation period as independent from the calibration period. For these reasons, the full instrumental record has been used for calibration and independent proxy records are used for validation. It was noted in the introduction that the number of proxy records extending back to 1000 AD is limited, but there are many more records of shorter extend. Here, two recent studies extending back to 1751 are used. Because of the greater data volume available, these reconstructions can be expected to have greater accuracy.

Wilson et al. (2006) produce a 1751–1981 reconstruction of tropical SST, using coral data. This tropical reconstruction can be complemented by the extra-tropical reconstruction of Wilson et al. (2007), which extends back to 1750. The Wilson et al. (2007) work uses tree-rings selected to be free of the so called "divergence problem". In both

studies, the proxy data is entirely independent of the data used here. In order to obtain an NH extra-tropical temperature estimate, the Wilson et al. (2007) index is scaled to match the variance of the NH mean temperature north of 20° N, 1856 to 2000. This is then combined with the Wilson et al. (2006) series, weighting each series with the area north and south of 20° N respectively to produce an estimate of Northern Hemisphere mean temperature (hereafter referred to as "Wilson*", Fig. 4a).

Figure 4b shows the period 1751 to 1900, and it can be seen that there is a very close correspondence between the R15 reconstruction and the combined Wilson* data in the period 1751 to 1850. There is agreement in the timing and magnitude of a cooling trend from 1790 to 1810 AD, though the Wilson* series has a somewhat cooler period from 1810 to 1820 AD.

Both the above time series are independent of the R15 reconstruction in terms of the proxy data input, though in both cases the scaling is determined by instrumental temperature.

The root mean square difference between the combined Wilson series and the R15 reconstruction is 0.136 K (0.139 K for the JAB2007 reconstruction). This cannot, however, be used as an error estimate because the Wilson et al. (2006, 2007) reconstructions have used observed temperature in the data selection and calibration.

The correlation between the series is, on the other hand, independent of the the calibration of the series. This composite Wilson series is found to have a correlation

$$R = 0.695$$

with the R15 series. Taking a conservative estimate of 15 degrees of freedom, this correlation is statistically significant at th 99.80% level, so that we can reject with near certainty the null hypothesis that fluctuations in the time series are purely random.

5   Uncertainty estimates

The comparison with a wholly independent set of data in the previous section gives a strong validation of the robustness of the climate signal in the proxy records, but does not provide an uncertainty estimate. This section will investigate and quantify different factors contributing to uncertainty.

5.1   The delete-d Jackknife estimator

The Jackknife method (e.g. Shao and Tu, 1995) will be used to generate ensembles of NH temperature reconstructions. The form of the Jackknife most widely used in atmospheric sciences is based on an ensemble obtained from deleting in turn each element of the input data. Here the "delete-d Jackknife" variants of the method, which are based on ensembles obtained by deleting all possible combinations of d elements of the input data, will be used. The "delete-d" version provides a means of estimating confidence limits as well as variances with reduced reliance on assumptions of normality.

The Jackknife method usually assumes independent of errors between input elements – an assumption which cannot be sustained in the case of proxy data. The impact of error correlations in the input data will be considered below, and a simple correction will be applied to the Jackknife uncertainty estimate.

The spread of reconstructions within ensembles which are generated by omitting up to 6 elements of the dataset of 13 and 15 member proxy collections are shown in Table 1. For each delete-d ensemble, the standard deviation about the ensemble mean, σ_d, is evaluated for each time step. The Jackknife error estimate for each time step (neglecting, for now, the impact of error correlation, which will be discussed below) is then given by:

$$\sigma_{d:\text{jack}} = \sigma_d \sqrt{\frac{N_c - d}{\overline{d}}} \qquad (2)$$

where $N_c = 13, 15$ is the number of proxies in the collection (Shao and Tu, 1995).

Table 1 gives the time-mean values of σ_d and $\sigma_{d:\text{jack}}$. It is clear that $\sigma_{d:\text{jack}}$ is only weakly dependent on the parameter d. However, for reasons mentioned above, the Jackknife estimate cannot be taken as a reliable guide to actual uncertainty without further work.

Figure 5 shows the 5% and 95% levels of the Jackknife distribution for the $d = 4, 5, 6$ ensembles, before and after scaling by the correction factor in Eq. (2). The scaling effectively collapses the results from different values of d onto a single line. There is slightly greater range in the 16th century and significantly lower range in the 18th century. The narrow range of reconstructions in the 18th century is likely to be partially due to the presence of proxies with long correlation time scales. Apart from this, the range does not vary much with time. The following sub-sections will use these ranges as the basis of uncertainty estimates.

The 5 and 95% percentiles of the Jackknife distribution will be used to estimate 5 and 95% confidence limits. This reduces the reliance on the assumption of Gaussian statistics. The Jackknife distribution does, however, have to be scaled in order to provide an estimate of the error distribution and that scaling is based on an analysis of the standard deviation. That is, instead of assuming Gaussian statistics, it is assumed that the shape of the Jackknife distribution scales linearly to the uncertainty distribution.

5.2 Estimating proxy error correlation

As noted above, the Jackknife method assumes that the errors in the proxies are uncorrelated. The impact of error correlations on the confidence limits can be estimated from the simpler problem of evaluating the mean of N_c errors, each of standard deviation σ. The squared variance of the mean, σ_m^2, is given by mean of the N_c^2 elements of the covariance matrix. If the correlations are zero, this gives the familiar result $\sigma_m^2 = N_c^{-1}\sigma^2$. If, however, the $N_c(N_c - 1)$ off-diagonal elements of the covariance matrix have a mean value $c\sigma^2$, we obtain:

$$\sigma_m^2 = N_c^{-1}\sigma^2[1 + (N_c - 1)c].$$

2641

That is, as is well known, the uncertainty in the mean increases with increasing error correlation.

However, it is shown in Appendix A that the presence of a positive correlation actually decreases the Jackknife error estimate given by Eq. (2):

$$\sigma_{d:\text{jack}}^2 = N_c^{-1}\sigma^2[1 - c].$$

This is understandable when one considers that a value of $c = 1$ will result in zero spread in the Jackknife ensemble and hence $\sigma_{d:\text{jack}} = 0$.

To adjust for this a corrected Jackknife estimate is introduced:

$$\sigma_{d:\text{jack:c}}^2 = \sigma_{d:\text{jack}}^2 \left[\frac{1 + (N_c - 1)c}{1 - c} \right]. \tag{3}$$

If the correlation were known, Eq. (3) would provide an improved uncertainty estimate. In practice, the correlation c is not known and we must rely on estimates, three of which are given in Table 2:

(1) $\overline{c}$: The anomaly correlation, with the anomaly calculated relative to the time mean;

(2) $\overline{c^*}$: The anomaly correlation, with the anomaly calculated relative to a reduced composite, omitting the proxy pair being correlated;

(3) $\overline{c^r}$: The correlation of residuals in the calibration period, after removing a fit to temperature in the calibration period.

In order to obtain a conservative estimate of the accuracy of the reconstructions, the largest of the 3 estimates is used in Eq. (3) to evaluate the correction factor in the last column. Although the correlations are not large, the presence of the factor $N_c - 1$ in the numerator leads to a significant correction.

Figure 6 shows the impact of scaling the spread of the Jackknife distribution using the factor 1.66 from Table 2.

2642

5.3 Structural uncertainty

The scaling of the reconstruction depends on the calibration method used. The disagreement between different methods is an indication of uncertainty which is not captured in the Jackknife estimation.

Figure 7 shows the spread of the composite against the Northern Hemisphere temperature record, together with 3 regression lines, representing different calibration methods, namely: least squares regression (red), inverse regression (blue) and variance matching (green). The range of regression slopes suggests that upper and lower reasonable bounds for the slope are 0.285 and 0.135 respectively. For the 13 proxy collection of JAB2007 the range is 0.262 to 0.122.

Suppose that the regression coefficient used to scale the composite series has a probability density function, $f_s(\gamma)$: that is, the probability of γ falling in a narrow interval around γ_1 is given by:

$$P\left[\gamma \in (\gamma_1 - \Delta/2, \gamma_1 + \Delta/2)\right] \approx \Delta f_s(\gamma_1) \tag{4}$$

for small Δ.

The Jackknife ensemble, $f_j(T)$, for a given γ:

$$P\left[T \in (T_1 - \Delta_T/2, T_1 - \Delta_T/2)|\gamma = \gamma_{vm}\right] = \Delta_T f_j(T_1). \tag{5}$$

Equations (4, 5) can then be combined to give a probability distribution of the reconstructed temperatures taking uncertainty in γ into account, making use of the fact that the distribution of temperatures for any value of γ can be obtained by scaling the temperature axis by γ/γ_{vm}. Convolving the two probability distributions then gives:

$$P\left[T \in (T_1 - \Delta_T/2, T_1 - \Delta_T/2)\right] = \Delta_T \int f_j\left(\frac{T_1\gamma_1}{\gamma_{vm}}\right) f_s(\gamma_1)d\gamma_1. \tag{6}$$

As with the error correlations discussed in the previous subsection, we do not have precise information about the probability distribution of γ. Two model distributions will

be considered: a piecewise uniform (almost uniform, but taking into account the asymmetry between values above and below γ_{vm}), and a piecewise Gaussian. These distributions are described in more detail in Appendix C.

Figure 6 shows the impact of the structural uncertainty adjustments, assuming a piecewise uniform distribution, on the 50-year mean. Results obtained with the piecewise Gaussian distribution are very similar and will be discussed further in the next section. The uncertainty in the calibration coefficient has a larger effect when the anomaly is large, but has a smaller effect in the 11th century, when the composite anomaly is small. Note, however, that the structural uncertainty multiplies the sampling uncertainty. Thus, even though the expected anomaly is near zero in the 11th century, the range of possible anomalies implied by the Jackknife ensemble is significant, and this range is magnified by the structural uncertainty.

5.4 Summary

Figures 8, 9 and 10 show the estimated uncertainty ranges for annual, decadal and 50-year averaged data. The 95% percentile of the uncertainty estimate of the annual reconstruction does not exceed the 1998 temperature at any point, but it would be incorrect to conclude that there is 95% certainty that the 1998 temperature was not exceeded at any point. To illustrate this, consider an idealised case in which every year of a century has a 4% chance of having exceeded some threshold T_{max} and that all years are independent: the probability that at least one year of the century exceeded T_{max} is easily calculated as $1 - (0.96)^{100} = 0.98$; in this idealised case there is only a 2% chance that T_{max} is not exceeded. The calculation for the temperature reconstruction is, however, more complicated because the probability of exceeding T_{max} varies from year to year, and the years are not entirely independent from each other. Section 6 below addresses this problem.

Other studies have shown that annual reconstructions have skill in reproducing the response to volcanic eruptions (e.g. Weber, 2005), but the longer term evolution of the temperature is clearer in the decadal and 50-year averaged fields (Figs. 9 and 10

respectively). The uncertainties here are not smoothed versions of the annual uncertainties: they are evaluated from Jackknife ensembles of smoothed reconstructions with the corrections for error correlations and structural uncertainty i (described in the previous section) applied.

Now there is a significant increase in uncertainty associated with the larger temperature anomalies of the 15th to 17th centuries.

6   Uncertainties in specific statements

The figures discussed in the previous section show that, for instance, the maximum of the estimated 95th percentile of the estimated uncertainty distribution. These figures cannot, however, directly address statements such as those from the IPCC quoted in the introduction which refer to the likelihood of a threshold being exceeded in a specified time period. The Jackknife technique does, however, provide a means to evaluate likelihoods for such statements.

Here three statements will be tested, all applied to the Northern Hemisphere annual mean temperature:

– S1: The 1998 temperature exceeded the maximum 11th century annual temperature,

– S2: The mean from 1990 to 1999 exceeded the maximum 11th century decadal mean temperature,

– S3: The mean from 1950 to 1999 exceeded the maximum 11th century 50-year mean temperature,

– S4: The mean from 1999 to 2008 exceeded the maximum 11th century decadal mean temperature.

2645

The likelihood of these statements being true is estimated by taking the Jackknife ensembles, scaled to account for structural uncertainty and proxy error correlations, and evaluating the proportion of ensemble members which satisfy each statement.

Results are presented in Table 3, together with the standard deviations of the scaled ensembles.

First, it can be seen that the difference between the $d = 5$ and $d = 6$ results is small for all variations in the table.

Secondly, the R15 reconstruction has standard deviations which are around 10% smaller than those of the JAB2007 union reconstruction, as might be expected from the slight increase in data. The estimated standard deviations decrease slightly with the averaging period, but the decrease is modest: the standard deviation for a 50 year average is no more than 20% smaller than that for the annual data.

Thirdly, although the standard deviations are smaller for the R15 reconstructions are reduced relative to the JAB2007 reconstruction the estimated certainties for statements 1 to 3 are reduced: this is a consequence of the upward shift in the central estimate of past temperatures.

A fourth salient feature of the results is the higher uncertainty with respect to S3, and the significantly lower uncertainty with respect to S2. This is a reflection of the fact that the recent anomaly in the 50 year mean is smaller than the recent decadal anomaly, but the uncertainties are only marginally smaller.

7   Conclusions

A new reconstruction of the millennial temperature has been generated from 15 proxies. The reconstruction uses the same simple composite plus scaling method as JAB2007, and differs through the addition of four and omission of one proxy data series. The temperature evolution is, not surprisingly, similar to that of the JAB2007 reconstruction, out shifted upwards consistently by about 0.2 K.

2646

The reconstruction uncertainty for the two series is, for 50 year means, estimated to be marginally under 0.2 K, so that this shift is within the expected uncertainty range.

The uncertainty analysis has been extended to make specific uncertainty estimates for several key questions. However, the value put on the uncertainty depends on a number of subjective assumptions. For this reason, this study presents a range of uncertainty estimates characterised by more or less optimistic assumptions about the nature of noise in the proxy records.

This study supports the conclusions of IPCC2001 about the exceptional nature of the 1998 temperature maximum but falls short of confirming the IPCC2007 conclusions about the exceptional nature of the temperature of the last 50 years of the 20th century. This latter failure is a consequence of the finding that the reduction in uncertainty which results from the 50 year averaging is very slight, whereas the reduction in signal is significant.

The strongest result relates to the temperature of the last decade, which exceeds any decade prior to 1850 with 95% certainty. The increased certainty compared to the 66% certainty expressed by IPCC2001 is primarily a consequence of the continuing high temperatures which have made the last decade 0.24 K warmer than the last decade of the 20th century, a warming greater than one standard deviation of the reconstruction uncertainty.

Appendix A

Proxy climate records

See Table A1.

Appendix B

The effect of error correlation on the Jackknife uncertainty estimate

Suppose we have N random series, with the covariance between the series given by the $N \times N$ matrix $\mathbf{C}$.

In this appendix the Jackknife algorithm is applied to the simple mean of the N random series. Comparing the Jackknife uncertainty estimate with the known exact solution for this problem gives an indication of the impact of data correlations on the Jackknife error estimates.

Let

$$\sigma^2 = \frac{1}{N} \sum_{i=1,N} \sigma_i^2$$

and

$$c = \sigma^{-2} \frac{2}{N(N-1)} \sum_{i=1,N-1} \sum_{j=i+1,N} \mathbf{C}_{ij}$$

be the mean variance and correlation respectively.

The variance of the mean of the N series is easily evaluated as

$$\sigma_m^2 = \frac{\sigma^2}{N} + \frac{N-1}{N} c$$

Now consider the delete-d Jackknife ensemble. Let $S_{d:\alpha}$, $\alpha = 1, N_\alpha$ be all the subsets of indices with d elements, where $N_\alpha = \binom{N}{d}$ is the number of ways to select d elements from N, and $\overline{S}_{d:\alpha}$ be the complement of $S_{d:\alpha}$ (i.e., all subsets of indices with d elements omitted).

Now consider the anomalies, e_α, of the Jackknife ensemble elements relative to the mean of the ensemble, which are given by:

$$e_\alpha = \frac{1}{N-d} \sum_{i \in \overline{S}_{d:\alpha}} a_i - \frac{1}{N} \sum_{i=1,N} = \frac{d}{N(N-d)} \sum_{i \in \overline{S}_{d:\alpha}} a_i - \frac{1}{N} \sum_{i \in S_{d:\alpha}a_i} . \qquad \text{(B1)}$$

The expected variance of e_α is given by

$$E[e_\alpha^2] = \left(\frac{d}{N(N-d)} \right)^2 \times \left[\sum_{i \in \overline{S}_{d:\alpha}} \sigma_i^2 + \sum_{i \in \overline{S}_{d:\alpha}} \sum_{j \in \overline{S}_{d:\alpha}, j \neq i} \mathbf{C}_{ij} \right] \qquad \text{(B2)}$$

$$+ \frac{1}{N^2} \left[\sum_{i \in S_{d:\alpha}} \sigma_i^2 + \sum_{i \in S_{d:\alpha}} \sum_{j \in S_{d:\alpha}, j \neq i} \mathbf{C}_{ij} \right] - \frac{2d}{N^2(N-d)} \sum_{i \in \overline{S}_{d:\alpha}} \sum_{j \in S_{d:\alpha}} \mathbf{C}_{ij}$$

If the mean is now taken over all α, the sums of σ_i^2 and $\mathbf{C}_{ij}$ span all possible combinations of i, j without favouring any values: it follows that σ_i^2 can be replaced with the mean value σ^2 and $\mathbf{C}_{ij}$ with its mean value $c\sigma^2$. After a little algebra it follows that the spread of the Jackknife ensemble is given by:

$$\sigma_d^2 = \frac{d}{N(N-d)} \sigma^2 (1-c) \qquad \text{(B3)}$$

2649

Appendix C

Model probability distributions for structural uncertainty.

The to distributions below are used to model the uncertainty in the scaling of the composite reconstructions, based on a central estimate γ_{vm} and upper (γ_{max}) and lower (γ_{min}) extremes which are asymmetrical about the central estimate.

Firstly, a piecewise uniform distribution, with median γ_{vm}:

$$f_{s:u} = \begin{cases} \frac{1}{2(\gamma_{max} - \gamma_{vm})} & \gamma > \gamma_{vm} \\ \frac{1}{2(\gamma_{vm} - \gamma_{min})} & \gamma < \gamma_{vm} \end{cases}$$

Secondly, a piecewise Gaussian distribution, again with median γ_{vm}:

$$f_{s:n} = \frac{1}{\sqrt{2\pi}\sigma_\pm} \exp \left[-\frac{(\gamma - \gamma_{vm})^2}{2\sigma_\pm^2} \right]$$

where

$$\sigma_\pm = \begin{cases} (\gamma_{max} - \gamma_{vm})/2 & \gamma > \gamma_{vm} \\ (\gamma_{vm} - \gamma_{min})/2 & \gamma < \gamma_{vm} \end{cases} .$$

In both cases, the discontinuity at γ_{vm} in the model distribution is used as a convenient way of obtaining the desired asymmetry between positive and negative anomalies.

2650

References

Anchukaitis, K., Evans, M., Kaplan, A., Vaganov, E., Hughes, M., Grissino-Mayer, H., and Cane, M.: Forward modeling of regional-scale tree-ring patterns in the southeastern United States and the recent emergence of summer drought stress, Geophys. Res. Lett., 33(4), L04705, doi:10.1029/2005GL025050, 2006. 2636

Anderson, D. M., Overpeck, J. T., and Gupta, A. K.: Increase in the Asian southwest monsoon during the past four centuries, Science, 297, 596–599, 2002. 2658

Briffa, K. R., Bartholin, T. S., Eckstein, D., Jones, P. D., Karlen, W., Schweingruber, F. H., and Zetterberg, P.: 1,400-year tree-ring record of summer temperatures in Fennoscandia, Nature, 346, 434–439, doi:10.1038/346434a0, 1990. 2658

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. 2658

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, J. Geophys. Res., 111, D12106, doi:10.1029/2005JD006548, 2006. 2636, 2661, 2662

Bürger, G., Fast, I., and Cubasch, U.: Climate reconstruction by regression – 32 variations on a theme, Tellus A, 58(2), 227–235, 2006. 2633

Chbouki, N.: Spatio-temporal characteristics of drought as inferred from tree-ring data in Morocco, Ph.D. thesis, University of Arizona, Tucson, Arizona, 1992. 2658

Christiansen, B., Schmith, T., and Thejll, P.: A Surrogate Ensemble Study of Climate Reconstruction Methods: Stochasticity and Robustness, J. Climate, 22, 951–976, doi:10.1175/2008JCLI2301.1, 2009. 2633

Cronin, T. M., Dwyer, G. S., Kamiya, T., Schwede, S., and Willard, D. A.: Medieval Warm Period, Global Planet. Change, 36, 17–29, 2003. 2658

Dahl-Jensen, D., Mosegaard, K., Gundestrup, N., Clow, G. D., Johnsen, S. J., Hansen, A. W., and Balling, N.: Past temperatures directly from the Greenland Ice Sheet, Science, 282, 268–271, 1998. 2658

D'Arrigo, R., Jacoby, G., Frank, D., Pederson, N., Cook, E., Buckley, B., Nachin, B., Mijiddorj, R., and Dugarjav, C.: 1738 Years of Mongolian temperature variability inferred from a tree-ring width chronology of Siberian pine, Geophys. Res. Lett., 28, 543–546, 2001. 2634, 2635, 2658, 2659

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. 2635

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, Ser. I, vol. 41, edited by: Jones, P. D., Bradley, R. S., and Jouzel, J., Springer-Verlag, New York, 297–328, 1996. 2658

Folland, C. K., Karl, T. R., Christy, J. R., Clarke, R. A., Gruza, G. V., Jouzel, J., Mann, M. E., Oerlemans, J., Salinger, M. J., and Wang, S. W.: Observed Climate Variability and Change, in: Climate Change 2001: The Scientific Basis, edited by: Houghton J. T., et al., Cambridge Univ. Press, Cambridge, 99–181, 2001. 2632

Ge, Q., Zheng, J., Fang, X., Man, Z., Zhang, X., Zhang, P., and Wang, W.-C.: Winter half-year temperature reconstruction for the middle and lower reaches of the Yellow River and Yangtze River, China, during the past 2000 years, The Holocene, 13(6), 933–940, 2003. 2634, 2658

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. 2658

Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, 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, UK and New York, NY, USA, 2007. 2632

Jones, P. D. and Moberg, A.: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, J. Clim., 16, 206–223, 2003. 2636

Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., van Ommen, T. D., Vinther, B. M., Lutherbacher, J., Wahl, E. R., Zwiers, F. W., Mann, M. E., Schmidt, G. A., Ammann, C. M., Buckley, B. M., Cobb, K. M., Esper, J., Goosse, H., Graham, N., Jansen, E., Kiefer, T., Kull, C., Kuettel, M., Mosley-Thompson, E., Overpeck, J. T., Riedwyl, N., Schulz, M., Tudhope, A. W., Villalba, R., Wanner, H., Wolff, E., and Xoplaki, E.: High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects, Holocene, 19, 3–49,

doi:10.1177/0959683608098952, 2009. 2632

Juckes, M. N., Allen, M. R., Briffa, K. R., Esper, J., Hegerl, G. C., Moberg, A., Osborn, T. J., and Weber, S. L.: Millennial temperature reconstruction intercomparison and evaluation, Clim. Past, 3, 591–609, 2007,
    http://www.clim-past.net/3/591/2007/. 2632, 2633, 2635

Lee, T. C. K., Zwiers, F. W., and Tsao, M.: Evaluation of proxy-based millennial reconstruction methods, Clim. Dynam., 31, 263–281, doi:10.1007/s00382-007-0351-9, 2008. 2633

Ljungqvist, F. C.: Temperature proxy records covering the last two millennia: a tabular and visual overview, Geogr. Ann. A., 91A, 11–29, 2009. 2635

Lloyd, A. H. and Graumlich, L. J.: Holocene dynamics of treeline forests in the Sierra Nevada, Ecology, 78, 1199–1210, 1997. 2658

Mann, M., Zhang, Z., Hughes, M., Bradley, R., Miller, S., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, P. Natl. Acad. Sci. USA, 105, 13252–13257, doi:10.1073/pnas.0805721105, 2008. 2632, 2634, 2635, 2638

Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlén, W.: Highly variable Northern Hemisphere temperatures reconstructed from low- and high-reslution proxy data, Nature, 433, 613–617, 2005. 2634

Moore, J., Hughen, K., Miller, G., and Overpeck, J.: Little Ice Age Recorded in Summer Temperatures from Varved Sediments of Donard Lake, Baffin Island, Canada, J. Paleolimnol., 25, 503–517, 2001. 2634, 2658

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. 2635

Naurzbaev, M. M., Vaganov, E. A., Sidorova, O. V., and Schweingruber, F. H.: Summer temperatures in eastern Taimyr inferred from a 2427-year late-Holocene tree-ring chronology and earlier floating series, The Holocene, 12, 727–737, 2002. 2635, 2658

Shao, J. and Tu, D.: The Jackknife and Bootstrap, Springer Verlag, New York, NY, USA, 540 pp., 1995. 2633, 2640

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:10.1029/2003GL017 352, 2003. 2658

Thompson, L. G.: Ice core evidence from Peru and China, in Climate Since A.D. 1500, edited by: Bradley, R. S. and Jones, P. D., New York, 517–548, 1992. 2658

von Storch, H., Zorita, E., and Gonzalez-Rouco, F.: Assessment of three temperature reconstruction methods in the virtual reality of a climate simulation, Int. J. Earth Sci., 98, 67–82, doi:10.1007/s00531-008-0349-5, 2009. 2633

Weber, S. L.: A timescale analysis of the Northern Hemisphere temperature response to volcanic and solar forcing, Clim. Past, 1, 9–17, 2005,
    http://www.clim-past.net/1/9/2005/. 2644

Wilson, R., Tudhope, A., Brohan, P., Briffa, K., Osborn, T., and Tett, S.: Two-hundred-fifty years of reconstructed and modeled tropical temperatures, J. Geophys. Res., 111, C10007, doi:10.1029/2005JC003188, 2006. 2638, 2639, 2662

Wilson, R., D'Arrigo, R., Buckley, B., Büntgen, U., Esper, J., Frank, D., Luckman, B., Payette, S., Vose, R., , and Youngblut, D.: A matter of divergence: Tracking recent warming at hemispheric scales using tree ring data, J. Geophys. Res., 112, D17103, doi:10.1029/2006JD008318, 2007. 2638, 2639, 2662

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:10.1029/2001GL014 485, 2002. 2658

Table 1. Properties of the delete-d Jackknife ensembles. Column 1: number of proxies omitted from collection; column 2, 5: number of ensemble members; column 3, 6: time-mean σ_d ; column 4, 7: time-mean uncertainty estimate $\sigma_{d:\text{jack}}$ (see Eq. 2). Columns 2–4 are for the 13 proxy JAB2007 collection, columns 5–7 for the R15 proxy collection.

	$N_c = 13$			$N_c = 15$		
d	N_d	$\overline{\sigma_d}$	$\overline{\sigma_{d:\text{jack}}}$	N_d	$\overline{\sigma_d}$	$\overline{\sigma_{d:\text{jack}}}$
6	1716	0.140	0.151	5005	0.113	0.139
5	1287	0.123	0.156	3003	0.100	0.141
4	715	0.106	0.160	1365	0.086	0.143
3	286	0.089	0.163	455	0.072	0.144
2	78	0.071	0.166	105	0.057	0.145
1	13	0.048	0.168	15	0.039	0.147

2655

Table 2. Different estimates of the proxy error correlation: see text for details.

N_c	$\overline{c}$	$\overline{c^*}$	$\overline{c^r}$	$\sqrt{\frac{1+(N-1)c}{1-c}}$
13	0.122	0.082	0.116	1.68
15	0.070	0.095	0.105	1.66

Table 3. Estimated likelihood that statements 1 to 4 are untrue (columns 5 to 8). Column 2 indicates the proxy collection used, column 3 gives the number of elements omitted in the Jackknife ensemble, column 4 gives the form used to model the structural uncertainty. The last three columns give the estimated standard error of the annual, decadal and 50 year averaged reconstructions.

Label	Col.	d	Struc	S1	S2	S3	S4	σ_a	σ_{10}	σ_{50}
1	R15	5	U	8.1%	12.0%	32.1%	0.9%	0.23 K	0.20 K	0.18 K
2	R15	5	G	16.2%	17.8%	34.4%	3.4%	0.26 K	0.22 K	0.20 K
3	R15	6	U	7.0%	11.1%	31.2%	0.8%	0.23 K	0.19 K	0.18 K
4	R15	6	G	14.8%	16.7%	33.5%	3.0%	0.26 K	0.22 K	0.20 K
5	JAB2007	5	U	4.7%	4.0%	18.1%	0.7%	0.27 K	0.24 K	0.23 K
6	JAB2007	5	G	10.5%	7.5%	20.9%	2.2%	0.31 K	0.28 K	0.26 K
7	JAB2007	6	U	3.7%	3.5%	17.6%	0.4%	0.26 K	0.23 K	0.22 K
8	JAB2007	6	G	9.0%	6.6%	20.0%	1.9%	0.30 K	0.27 K	0.25 K

Table A1. Proxy climate records.

| Name (country)	Position	Type	Citation
GRIP: borehole temperature (Greenland)	73° N, 38° W	Ice core	Dahl-Jensen et al. (1998)
Composite (China)	30° N, 105° E	Composite	Yang et al. (2002)
Globigerina bulloides [frequency] (Arabian Sea)	18° N, 58° E	Foraminifera	Anderson et al. (2002) Gupta et al. (2003)
Upper Wright (USA)	38° N, 119° W	Tree ring	Lloyd and Graumlich (1997)
Shihua Cave: layer thickness (China)	40° N, 116° E	Stalactite	Tan et al. (2003)
Quelcaya 2 [δ^{18}O] (Peru)	14° S, 71° W	Ice core	Thompson (1992)
Boreal (USA)	35° N, 118° W	Tree ring	Lloyd and Graumlich (1997)
Taymir (Russia)	72° N, 102° E	Tree ring	Naurzbaev et al. (2002)
Fennoscandia	68° N, 23° E	Tree ring	Briffa et al. (1990)
Northern Urals (Russia)	66° N, 65° E	Tree ring	Briffa et al. (1995)
Crete (Greenland)	71° N, 36° W	Ice core	Fisher et al. (1996)
Col du Zad (Morocco)	33° N, 5° W	Tree ring	Chbouki (1992)
Chesapeake Bay: Mg/Ca (USA)	38° N, 76° W	Sediment	Cronin et al. (2003)
Composite (Mongolia)	48° N, 99° E	Tree ring	D'Arrigo et al. (2001)
Temperature reconstruction (China)	40° N, 116° E	Phenological records	Ge et al. (2003)
Donard lake (Canada)	67° N, 61° W	sediments	Moore et al. (2001)

Proxy positions
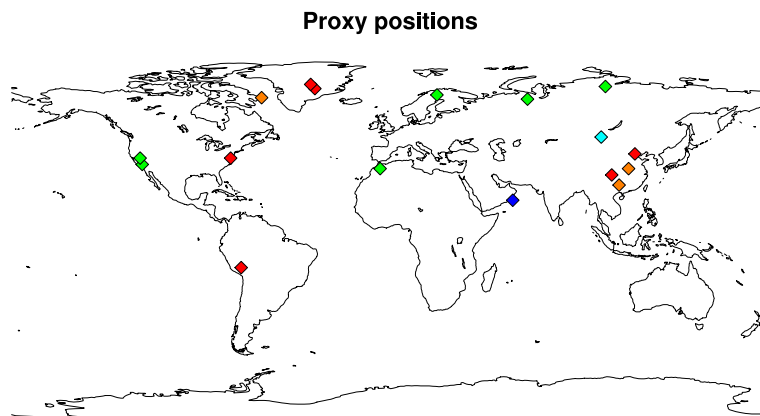
Fig. 1. Map showing the location of climate proxies: green and red show, respectively, tree-ring and non-tree ring proxies used in both JAB2007 and this study, dark blue is the Globigerina bulloides series used in JAB2007, light blue is the Mongolian composite ring of D'Arrigo et al. (2001), and two further Chinese series are in orange. Tree-ring data marked in green.

2659
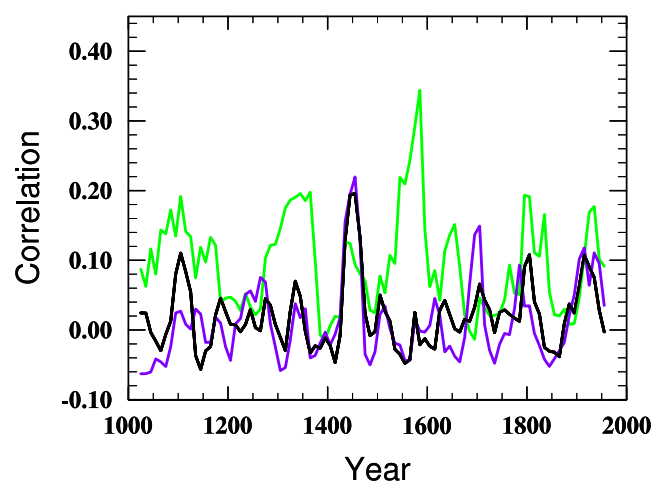
Fig. 2. Proxy correlations as a function of time, c_n, is shown in black. Also shown are the correlations among the tree ring chronologies (green) and among the non-tree ring chronologies (purple).
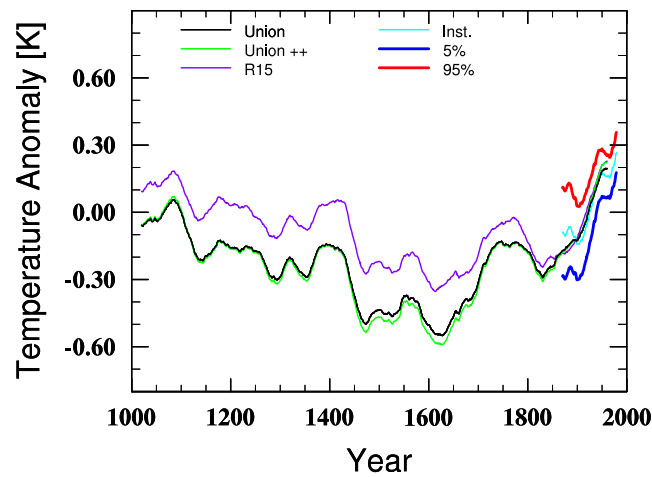
Fig. 3. Reconstructions of the Northern Hemisphere mean temperature: union reconstruction from JAB2007 (black); union reconstruction scaled to new instrumental record (green); R15 reconstruction (blue); and the Northern Hemisphere temperature record from Brohan et al. (2006) with 5% and 95% confidence limits.

2661
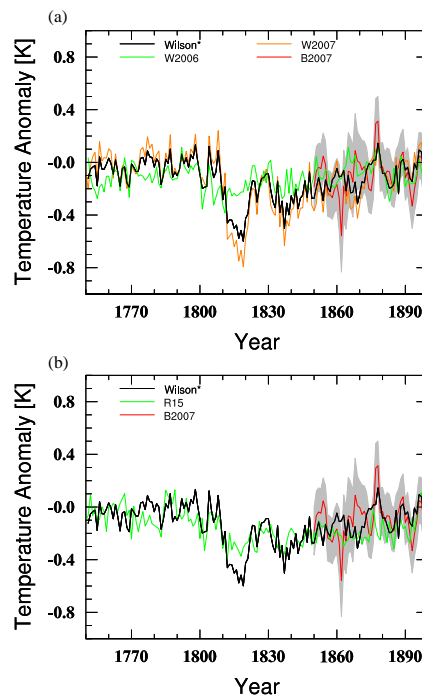
Fig. 4. (a) Tropical temperature reconstruction from Wilson et al. (2006) (W2006, green), extra-tropical temperature reconstruction from Wilson et al. (2007) (W2007, orange), instrumental temperature record from Brohan et al. (2006) (with its 5–95% confidence interval shaded – B2007, red) and the weighted mean of W2006 and W2007 (Wilson*, black). **(b)** The instrumental series and the combined (Wilson*) series as in (a), with the R15 reconstruction plotted in green.
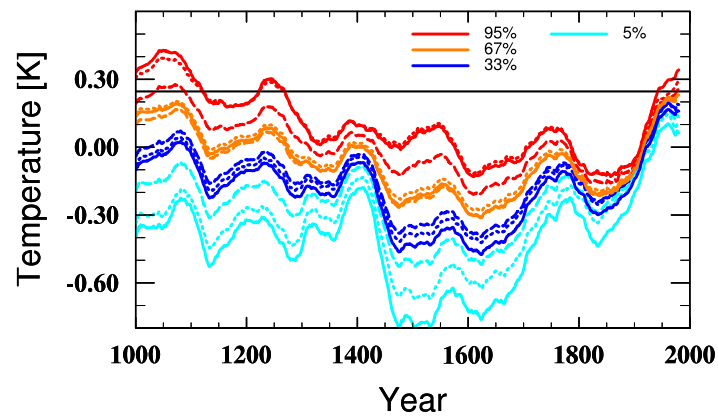
2662
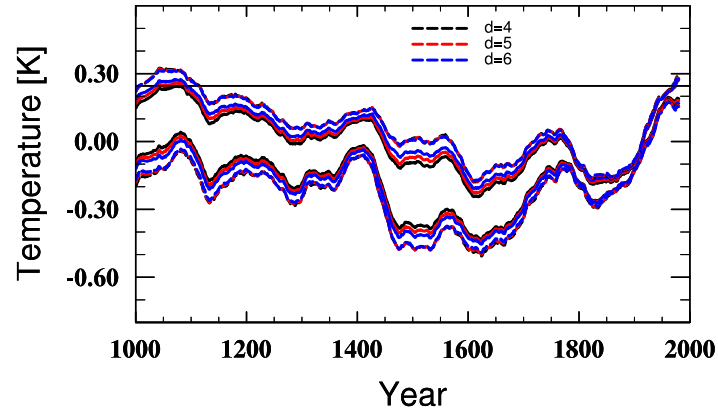
Fig. 5. 5th and 95th percentiles for the delete-d ensembles with $d = 4,5,6$ after smoothing by 50 year running mean. Before scaling (solid lines) and after scaling by Eq. (2) (dashed lines). A horizontal line shows the mean 1951–2000 AD Northern Hemisphere temperature.

Fig. 6. 5th, 33rd, 67th, and 95th percentiles. Different colours correspond to different percentiles as indicated in top rigth corner. Different line styles indicate different stages of the analysis: Dashed: un-scaled; longer dashes: scaled as in Eq. (2); alternating dashes: scaled for correlation correction; solid: adjusted for structural uncertainty. The horizontal line is as in Fig. 5.
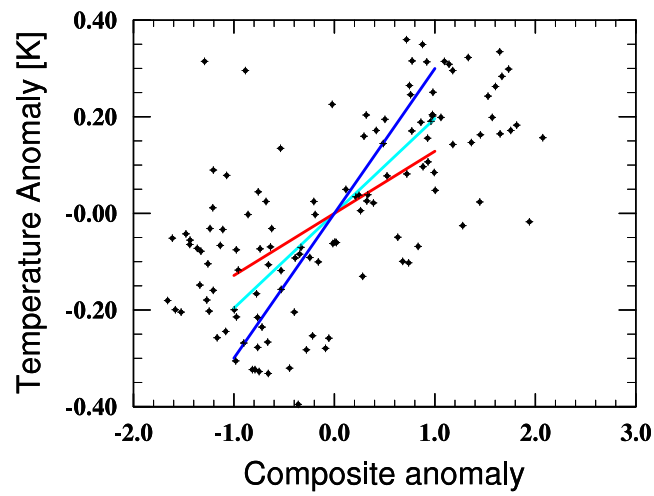
Fig. 7. Scatter of the observed temperature anomaly (y-axis) against the composite anomaly (x-axis). Each star marks one year. Anomalies relative to the 1856 to 1980 mean. Slopes are: 0.18, 0.12 and 0.26 K (the proxy composite is dimensionless).

2665
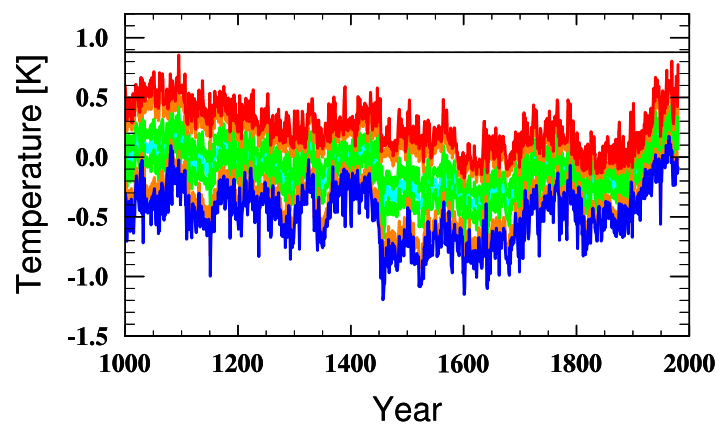
Fig. 8. Uncertainty estimates for annual reconstructed temperature: The percentiles are coloured as follows: 5%: blue; 95%: red; 10% and 90%: orange; 33% and 67%: green; median: cyan. A horizontal plack line indicates the observed Northern Hemisphere mean temperature for 1998.

2666
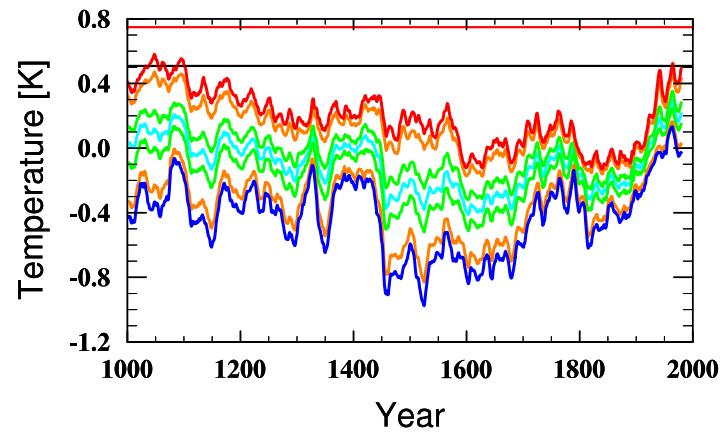
Fig. 9. As Fig. 8, except showing uncertainties for decadally averaged data, and lines indicating the mean of 1990–1999 (black) and 1999–2008 (red).
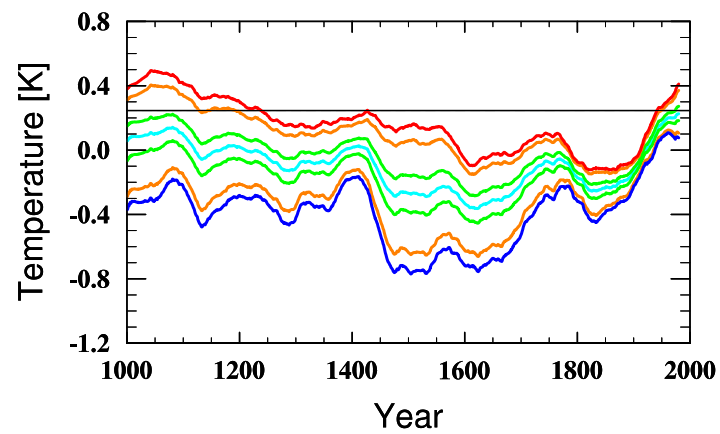
Fig. 10. As Fig. 8, except showing uncertainties for 50 year averaged data, and a line indicating the mean of 1950–1999.