University of East Anglia

Annually resolved patterns of summer temperature over the Northern Hemisphere since AD 1400 from a tree-ring-density network

T. J. Osborn, K. R. Briffa,
F. H. Schweingruber and P.D. Jones

2004 (published as CRU RP10 in 2012)

Climatic Research Unit
School of Environmental Sciences
University of East Anglia

Climatic Research Unit Research Publication 10

(CRU RP10)

About the Climatic Research Unit

www.cru.uea.ac.uk

The Climatic Research Unit (CRU) is widely recognised as one of the world's leading institutions concerned with the study of natural and anthropogenic climate change. CRU is part of the School of Environmental Sciences at the University of East Anglia in Norwich. The aim of the Climatic Research Unit is to improve scientific understanding in three areas:

- past climate history and its impact on humanity;
- the course and causes of climate change;
- prospects for the future.

About Climatic Research Unit Research Publications

www.cru.uea.ac.uk/publications/

The majority of CRU's research output is published in the peer-reviewed science literature (journals and books), listed in full on our publications webpage. CRU Research Publications (CRU RPs) are occasional research reports that, due to various factors such as length, style or intended audience, are not suitable for publishing in the science literature. The first nine CRU RPs were originally published in hard copy form but are now available in electronic form, while from CRU RP10 onwards they are available only from our website in electronic form.

CRU Research Publications series

www.cru.uea.ac.uk/publications/crurp/

CRU RP1 (1973) Lamb HH
The seasonal progression of the general atmospheric circulation affecting the North Atlantic and Europe.

CRU RP2 (1974) Collected abstracts of the International CLIMAP Conference held in Norwich in May 1973
Mapping the atmospheric and oceanic circulation and other climatic parameters at the time of the Last Glacial Maximum, about 17000 years ago.

CRU RP3 (1974) Lamb HH
The current trend of world climate – a report on the early 1970s and a perspective.

CRU RP4 (1975) Wright PB
An index of the Southern Oscillation.

CRU RP5 (1977) Lamb HH
Understanding climatic change and its relevance to the world food problem.

CRU RP6 (1978) Douglas KS, Lamb HH & Loader C
A meteorological study of July to October 1588: the Spanish Armada storms.

CRU RP6a (1979) Douglas KS & Lamb HH
Weather Observations and a Tentative Meteorological Analysis of the Period May to July 1588.

CRU RP7 (1980) Perry AH & Fearnside T
Northern Hemisphere pentad (5-day) mean sea level pressure values for the period 1951–70 and comparisons with earlier epochs.

CRU RP8 (1985) Jones PD, Ogilvie AEJ & Wigley TML
Riverflow data for the United Kingdom: reconstructed data back to 1844 and historical data back to 1556.

CRU RP9 (1988) Santer BD
Regional validation of General Circulation Models.

CRU RP10 (2004) Osborn TJ, Briffa KR, Schweingruber FH & Jones PD
Annually resolved patterns of summer temperature over the Northern Hemisphere since AD 1400 from a tree-ring-density network.

Annually resolved patterns of summer temperature over the Northern Hemisphere since AD 1400 from a tree-ring-density network

Timothy J. Osborn[1], Keith R. Briffa[1],

Fritz H. Schweingruber[2], Phil D. Jones[1]

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, United Kingdom

[2]Swiss Federal Institute of Forest, Snow and Landscape Research, Birmensdorf, Switzerland

2004

(published as CRU RP10 in 2012)

Correspondence to: Dr. Tim Osborn
Climatic Research Unit
University of East Anglia
Norwich NR4 7TJ
U.K.

tel: +44 1603 592089
fax: +44 1603 507784
e-mail: t.osborn@uea.ac.uk

Abstract

The development of annually-resolved estimates of summer temperature patterns over the mid-to-high latitudes of the Northern Hemisphere land masses is described. The estimates were derived from a calibration of an extensive network of tree-ring density chronologies, with data extending back as far as 1400 (though with greatly reduced spatial coverage prior to 1600). The chronology data were first gridded onto a 5° longitude by 5° latitude grid, followed by a local calibration and verification of each grid box series against observed April-to-September mean temperatures. It was demonstrated that calibration by simple linear regression of individual grid boxes in isolation (i.e., without using any non-local proxy or temperature information) yields a spatial reconstruction with suppressed variance when averaged over large areas. The local calibration was adapted, therefore, to ensure that the magnitude of sub-continental, regional-scale variations is realistically maintained, while still achieving significant skill at the grid-box scale. A second version of the reconstructed grids was also generated, by incorporating additional variations on century time scales that may have been removed by the original standardisation of the tree-ring density chronologies. The spatial extent of both versions (with and without the additional low-frequency variability) was extended using principal component regression, for all locations where a verification correlation of 0.4 or above was achieved. These reconstructions of summer temperature have significant skill, with an overall space-time correlation of around 0.7 for both the calibration and verification against observed temperatures.

Keywords: Dendroclimatology; Tree-ring density; Temperature reconstructions; Climate change.

1. Introduction

A number of quantitative reconstructions of large area-average temperatures have been developed during the last decade (e.g., Bradley and Jones, 1993; Jones *et al.*, 1998; Mann *et al.*, 1999; Crowley and Lowery, 2000; Huang *et al.*, 2000; Briffa *et al.*, 2001; Esper *et al.*, 2002), based on various networks of natural climate proxies and sometimes documentary and early instrumental records. There are fewer spatially-resolved temperature reconstructions, however, that can allow the calculation of arbitrary regional indices or spatial analyses (e.g., a composite pattern of response to

volcanic eruptions). Those that do exist are generally of restricted spatial extent (e.g. Fritts, 1991; Schweingruber *et al.*, 1991; Vaganov *et al.*, 1996; Luterbacher *et al.*, 2004), with the notable exception of the Mann *et al.* (1998) reconstruction of grid-box temperatures that covers 53% of the globe.

The present paper is one of a number (e.g., Briffa *et al.*, 1998a,b, 2001, 2002a,b) that document stages in the analysis and exploitation of a particular network of tree-ring maximum latewood density (MXD) measurements (Figure 1a), covering much of the mid-to-high latitudes of the Northern Hemisphere (Schweingruber and Briffa, 1996). The purpose of the present paper is to describe for the first time the development of *spatially-resolved* reconstructions of past summer temperatures from this data set.

Briffa *et al*. (2002a) demonstrated the strong relationship between the MXD chronologies in this tree-ring network and warm-season temperature, and Briffa *et al.* (1998a,b, 2001, 2002a) used these data (or earlier subsets) to make regional and "hemispheric-wide" area-averaged temperature reconstructions. However, the strength and consistency of the correlation between these tree-ring density data and warm-season temperature, together with the broad coverage of the network (Figure 1a), make them excellent material from which to develop spatially-resolved temperature reconstructions over most of the Northern Hemisphere extra-tropical land masses. This paper describes the development of two such reconstructions; one is a version that is suitable for studies of climate variability on inter-annual to multi-decadal time scales, but with reduced variability on time scales of a century or more, while the second version has additional variability on these longer time scales. Initial results from the first version have been presented by Briffa *et al.* (2002b), and selected analyses of both versions were reported by Briffa *et al.* (2003), but those papers did not document the development of these reconstructions. It is the purpose of the present paper (which is the primary reference for the spatially-resolved reconstructions) to provide these details, including an assessment of reconstruction skill and a justification of the underlying assumptions.

A key principle of this study is to reconstruct the temperature history at each grid box independently (i.e., on a grid-box by grid-box basis). Other approaches (e.g., Briffa *et al.*, 1986; Mann *et al.*, 1998; Luterbacher *et al.*, 2004) often make use of the leading Empirical Orthogonal Function (EOF) patterns obtained from a Principal Component Analysis (PCA) of twentieth century

temperature observations. Although we are *not criticising* such approaches (and indeed we have used such an approach with the MXD data set – Rutherford *et al.*, 2005), the grid-box basis of the present study was selected specifically to provide an alternative that allows the variations in the MXD data themselves to determine the pre-twentieth century spatial covariance structure. The approaches that reconstruct EOF patterns impose, to some extent, fixed patterns of temperature variability throughout the reconstruction (though the relative variances of each pattern may differ over time). For proxy networks that have a heterogeneous response to climate (e.g., a network that combines proxies that respond to different seasons or to different climate variables), an EOF-type approach is clearly necessary to extract the common signals from the diverse data. The relatively homogeneous climate response of the MXD network (to warm-season temperature), however, does allow a grid-box by grid-box approach to be taken. This approach clearly does not make use of the spatial covariance present in the predictor and predictand fields (as EOF-based methods do), but it provides an interesting alternative for future comparison with results from pattern-based reconstruction techniques.

The tree-ring density data, including their spatial and temporal coverage, are described in section 2, together with a description of the instrumental temperature observations used for the calibration of the reconstruction. Section 3 explains how the MXD time series were assembled into a regular grid, and how infilling of some early grid-box time series was undertaken to maximise the spatial coverage for selected time periods. The calibration of the gridded MXD data, to create estimates of warm-season temperature, and its subsequent verification are reported in section 4, together with a discussion of some drawbacks of the grid-box by grid-box approach to calibration. Section 5 describes the extension (using principal component regression) of the temperature reconstruction to provide temperature estimates for grid boxes where there are no tree-ring measurements. The incorporation into these reconstructions of previously-reconstructed long-term (multi-decadal to multi-centennial) temperature variations is described in section 6. Some preliminary analysis of the annually-resolved patterns of summer temperature produced in this study is presented in section 7, though more thorough analysis and comparison with other data will be (or is already) published elsewhere. The study's main results are summarised and discussed in section 8.

2. Tree-ring and climate data

2.1. Tree-ring density data

The data set used as the basis for the reconstructions reported here is a set of 387 tree-ring density chronologies that cover most of the high latitudes, and some of the mid latitudes, of the Northern Hemisphere land masses (Figure 1a). The primary objective for developing this tree-ring network (Schweingruber and Briffa, 1996) has been to provide the basis for reconstructing a multi-century history of temperature variability; to this end, the collection of tree-core samples was focussed at relatively cool and moist sites (usually near to the elevational or latitudinal tree-line).

The data used here are the maximum latewood densities – i.e., the maximum density measured within an annual growth ring, usually associated with the cells formed towards the end of the growing season (hence, "latewood"). Each chronology is an annual time series formed from the average of indices derived from density measurements of individual tree cores. Each measurement series was first detrended by taking residuals from a generalised exponential, or "Hugershoff", function fitted to the series. The purpose of detrending (or "standardisation") is to remove the multi-century variance that results from a tendency for an older tree to produce less-dense rings than a younger tree, but it also results in the loss of long time scale variance that might be caused by changing climate (Cook *et al*., 1995; Briffa *et al*., 1996). Hence, the data and reconstructions generated in sections 3, 4 and 5 are only applicable to temperature variability that encompasses inter-annual, decadal and multi-decadal-to-century time scales, but not longer (i.e., multi-century) periods. Briffa *et al*. (2001) have, however, processed the same set of tree core density measurements using a different method that was designed to preserve even the multi-century variability present in the data (while still removing the age-related trends). Section 6 describes how this longer time scale variability is incorporated (at the sub-continental, regional scale used by Briffa *et al.*, 2001) into the reconstructed summer temperature patterns – yielding spatially-resolved data and reconstructions that are representative of a wider spectrum of variability than has hitherto been produced from this tree-ring network.

At least 20 tree cores from one site and for one tree genus are typically available, though the tree cores do not all span the same time period. Not only is the number of chronologies with data

lower in the earlier centuries (Figure 2a), but each chronology also comprises fewer tree cores (Figure 2b). The fraction (f) of constituent cores in each chronology that have data in a particular year, averaged over only those chronologies that have at least some data in that year (i.e., excluding those with $f = 0$ in that year), falls from almost 1.0 in the mid-twentieth century, to 0.2 in 1600, and remains at 0.2 back to at least 1400 (Figure 2b). Although this fraction remains steady before 1600, the number of chronologies with data continues to decrease from 76 in 1600, to 25 in 1500 and eight in 1400. For this reason, it was decided to limit the reconstruction of summer temperature patterns to the period since AD 1400, with the additional caveat that spatial coverage (and possibly reconstruction skill) will be limited prior to 1600.

2.2. Temperature observations

The instrumental observations used for calibration and verification of the reconstructions are taken from the global set of monthly mean anomalies (with respect to 1961–90) available on a 5° latitude by 5° longitude grid-box basis since 1856 (HadCRUT1; Jones *et al*., 1999). These data are air temperatures measured at land stations combined with sea surface temperature observations.

Briffa *et al.* (2002a) have already identified many of the climate signals recorded in the tree-ring data, and concluded that the tree-ring density variability is dominated by the temperature during the growing season. Although the season with optimal sensitivity varied spatially (and possibly with tree genus), if a single seasonal definition is chosen to best represent the temperature signal within the network as a whole, then the April-to-September mean temperature is the most appropriate (Briffa *et al.*, 2002a).

April-to-September mean temperature anomalies were calculated from the monthly grid-box data, requiring at least three monthly values to form a seasonal mean. A small number of missing values were infilled, to maximise the available data for calibration and verification of the reconstructions. This was achieved first by spatial infilling, such that any missing value that had three or more immediately neighbouring grid boxes with non-missing values, was replaced by an unweighted average of those neighbouring values. Temporal infilling was subsequently applied, such that any missing April-September temperature that had non-missing values for both the previous and

subsequent summers, was replaced by an unweighted average of those values. The impact of these infilling procedures is fairly minor, but is nevertheless considered worthwhile to provide the maximum amount of temperature values for the calibration and verification of the tree density data. For the 115 grid boxes (see section 3) where the MXD data are located, and for the 80 years (see section 4.1) over which calibration is undertaken, only 81% of the 9200 values (i.e., 115×80) had complete data. Allowing as few as three monthly anomalies to be used to form a seasonal mean increased the data coverage to 83% of the calibration values. The spatial and temporal infilling further increased this to 87% (almost all due to the spatial infilling).

3. Gridding and infilling the tree-ring density data

The 387 density chronologies were screened according to their correlation with the "local" grid-box April-to-September temperature, already computed by Briffa *et al.* (2002a; their Figure 5a) over the period 1881–1994. The actual period over which these correlations were computed is often shorter than 1881–1994, because either some temperature observations are missing (see section 2.2) or the trees were sampled before 1994 (Figure 2a). Almost half of the chronologies exhibit a correlation of 0.5 or above, and almost 90% exhibit a correlation of 0.22 or above (the latter threshold indicates that we can be 95% confident of a true association between MXD and April-to-September temperature, given a sample of 81 years, which is the average overlap between chronology and temperature data). The 46 chronologies that do not meet this criterion were excluded from further use; Figure 1b distinguishes between the retained (open circles) and discarded (filled circles) chronologies. Some of the discarded chronologies may in fact have a summer temperature signal in them, perhaps being obscured because the selected seasonal-mean is not optimal (e.g., those in eastern Siberia that Briffa *et al.*, 2002a, show respond to a shorter growing-season temperature), or because the nearest grid-box temperature series is not very "local" to the chronology site.

The 341 chronologies that were retained were gridded onto the same 5º latitude by 5º longitude grid that the observed temperature data are on. Each grid-box time series was generated using all tree-ring density chronologies whose sites are located within the grid box, and no others were used. For those grid boxes with more than one chronology, the chronology time series were averaged

together, each weighted by its time series *f* (the fraction of its cores with data in a given year). Thus, a chronology with only a few of its cores having measurements early on would be given less weight in that early period.

The variance of any grid-box time series depends upon the number of chronologies (*n*) being averaged together, and if *n* varies through time then so will the variance of the grid-box time series (Osborn *et al.*, 1997). Such changes in variance are wholly artificial, since they are unrelated to any change in climate variability. For example, if chronologies spanning 1740–1990 and 1600–1986 were located in the same grid box, then the 1600–1990 grid-box time series would be an average of two chronologies during 1740–1986, with lower variance than the 1600–1739 and 1987–1990 parts of the time series which would be obtained from single chronologies. All grid-box time series with a time-dependent number of chronologies were adjusted to remove such artificial changes in variance, using the procedure described by Osborn *et al.* (1997) based on *n* and the mean inter-chronology correlation.

After averaging the chronologies together and adjusting for time-dependent changes in variance, each grid-box time series was "normalised" so that it had zero mean and unit variance over the 1881–1960 period. A total of 115 grid-box time series were formed in this way, at the locations shown in Figure 3a. The number of boxes with data gradually falls further back in time, to 40 in 1600 and to just seven in 1400 (from only eight chronologies), as shown by the grey shading in Figure 2c.

Many multi-variate analysis techniques, such as principal component analysis (PCA), require data sets with no missing values. Although such techniques are not applied here, the reconstructions may be subjected to such techniques elsewhere (e.g., Briffa *et al.*, 2002b, present results of a PCA of the gridded reconstructions described in the present paper). It is not valid to infill the grid-box time series so that all of them have complete data for the full period since 1400, but it is reasonable to attempt to extend relatively short sections of some grid-box time series if doing so will help to produce a useful space–time subset with complete values (whereas not doing so would give a much reduced subset due to a relatively small number of missing values). Such infilling is only accepted if it passes a rigorous verification procedure.

The grid-box MXD time series were extended by using multiple regression models with neighbouring MXD series as predictors, calibrated and verified using data from the 1810–1970 period.

This period is split into quarters, with the first and last quarters used for calibration and the middle two quarters used for verification. The long, 80-year verification period enables a rigorous test to be made of the regression model performance. The regression model is rejected, and no infilling is performed for a particular grid box, unless the correlation coefficient between the prediction and actual data during the verification period exceeds 0.7 (i.e., $r^2 \geq 0.5$). Of the 115 grid boxes with some data, nine do not have complete data for the calibration and verification periods and are not extended (they are left as relatively short time series only) – nor are they used as predictors for extending other grid-box series.

Extension of the remaining series is first attempted back to 1400. The seven grid-box time series (shown black in Figure 3c) that have complete data back to 1400 form the pool of potential predictors for the other 99 boxes. For each of the 99 boxes without complete data back to 1400, four alternative multiple regression models are built using those potential predictors that are within 800, 1600, 2400 and 3000 km of the predictand grid box. The regression model with the highest verification correlation is used to infill the series back to 1400, provided that the best model meets the minimum verification skill criterion. Out of the 99 that were attempted, 19 time series were successfully extended back to 1400 (shown grey in Figure 3c); these 19 are all very close to the seven predictor grid boxes. The coverage in 1400 was, therefore, increased from the seven boxes with real measurements to 26 including those that were infilled.

This procedure was repeated, but only trying an extension back to 1453. There were 15 boxes with complete data back to this year (note that the infilled sections of time series were never used as predictors for other grid boxes). Extension of the 91 boxes that did not have complete data back to 1453 was attempted and was successful for 24 (giving a total of 39 boxes with values back to 1453). Some of the 91 boxes for which infilling to 1453 was tried had already been successfully infilled back to 1400. If the infilling to 1453 failed, then the previous estimates all the way back to 1400 were still used; if the infilling to 1453 was successful (i.e., the regression model passed the verification test), then these values were used from 1453 onwards, on the assumption that the greater pool of potential predictors was likely to yield improved results.

This process was repeated for periods back to 1540, 1583, 1612, 1660, 1697, 1743 and 1777,

times that were chosen because they occurred immediately after notable increases in the spatial coverage of the un-infilled data (thus increasing the number of potential predictors). The number of grid-box time series successfully extended to each of these years was 22 (out of 82), 25 (out of 69), 21 (out of 61), 20 (out of 50), 20 (out of 34), 12 (out of 15) and 3 (out of 5), respectively (Table 1). The increased coverage through time is shown by the difference between the black line and the grey shading in Figure 2c, and the spatial patterns of coverage back to 1583 and 1697 are given in Figures 3e and 3g, respectively. Extensions forward to 1994, 1988 and 1980 were also attempted, but with less success: 2 (out of 101), 4 (out of 47) and 6 (out of 16), respectively. Of the 46,497 values in the extended tree-ring density data set, 15% are estimates from the regression-based infilling procedure (Figure 2c).

4. Calibration of the gridded data

4.1. Selection of calibration period and method

A number of factors were taken into account when selecting the most appropriate periods for calibration and verification of the gridded density data against observed temperatures. The most important factor is the identification by Briffa *et al.* (1998a) of a recent downward trend in the high-latitude tree-ring density data, relative to (and apparently unrelated to) warm-season temperature. This density decline becomes large enough to impair the calibration after about 1960. For this reason, both Briffa *et al.* (2001) and Briffa *et al.* (2002a) used only pre-1961 data for calibration of their sub-continental, regional temperature reconstructions. This is a reasonable choice, provided that it is explicitly stated that this approach assumes the apparent recent density decline is due to some anthropogenic factor and that similar behaviour is assumed, therefore, not to have occurred earlier in the reconstruction period – which would otherwise introduce bias in the reconstructed temperatures. At present, no satisfactory explanation of the relative MXD decline has been identified, and further work must dictate whether this assumption will be supported or rejected (Briffa *et al.*, 1998a, 2003, and Vaganov *et al.*, 1999, discuss and investigate possible causes).

Two factors combine to restrict the starting year of data used for calibration. The first is the availability of instrumental temperature observations. Though early data exist in Europe, it is not until

1851 that moderately widespread coverage begins (Jones *et al.*, 1999). Even by then, however, many of the high latitudes where the tree-ring chronologies are located do not have data, with measurements in some locations not starting until the second half of the twentieth century. The second factor is that some early observations may be inhomogeneous due to changes in exposure or observing practise during the second half of the nineteenth century, and summer temperatures are particularly doubtful (Chenoweth, 1993; Moberg *et al.*, 2003). Annual-mean temperatures may be less influenced by exposure changes and could provide a longer period of reliable data for calibration. It may be tempting, therefore, to calibrate against annual temperatures (e.g., Mann *et al.*, 1998) and such a calibration may appear to be successful because large-scale summer and annual temperatures are highly correlated during most of the instrumental period. Such a calibration may prove to be invalid, however, for proxies such as the MXD data which are clearly responsive to warm-season temperature (Briffa *et al.*, 2002a), because the correlation between summer and annual temperature may break down outside the calibration period (Briffa and Osborn, 2002; Jones *et al.*, 2003).

It is clear, therefore, that it is most appropriate to calibrate the MXD data against warm-season temperature, even though this choice constricts the calibration at both ends – due to sparse and potentially-biased early temperatures and a recent tree-ring density decline. Briffa *et al.* (2001) and Briffa *et al.* (2002a) used an 1881–1960 calibration period because, for the large regional averages used in those studies, it did not matter if some of the grid boxes within the region had no data in the early part of the calibration period. For the grid-box calibrations reported here, however, many grid boxes had no temperature observations prior to about 1911.

The approach selected for calibrating the data was determined by the various constraints discussed above. The calibration period was 1911–1990, though a minimum overlap between temperature observations and MXD of 21 years was allowed where full data were not available. The recent relative decline in MXD was prevented from affecting the calibration by two steps. First, the MXD data were calibrated using only high-frequency variations, by filtering both predictor and predictand data (the decline has virtually no effect on the relationships at inter-annual time scales). The calibration coefficients derived were then applied to the unfiltered MXD data to generate the reconstruction. Osborn and Briffa (2000) demonstrated, however, that the calibration coefficients may

be significantly larger at low frequencies than at high frequencies (implying a genuine timescale-dependent variation in the relationship between tree growth and climate). The second calibration step was introduced to allow for this possibility (i.e., that the use of high-pass filters prior to calibration may have resulted in underestimated calibration coefficients): the MXD data were (temporarily) adjusted to *artificially* remove the decline, then the calibration coefficients were determined using unfiltered data and applied to the unadjusted MXD data to generate the reconstruction. These steps are described in more detail in the following sections.

4.2. Grid-box by grid-box calibration and verification

To calibrate the MXD data against temperature using only the high-frequency variations (thus removing the influence of the low-frequency "decline" feature), each grid-box time series was filtered with a 40-year high-pass filter. Figure 4a shows the high-frequency correlation between gridded MXD and temperature over the calibration period (1911–1990, or the subset of at least 21 years that has both MXD and temperature values), indicating positive correlations everywhere. This is as expected, because the 46 chronologies that do not show a significant correlation with warm-season temperature in the study of Briffa *et al.* (2002a) had already been removed (section 3), though the high-pass filtering used here could have resulted in differences in correlation to those obtained in that study. Correlations above 0.6 occur in 40 out of the 109 grid boxes (a further 6 boxes, in northern Siberia, have insufficient temperature observations to compute a correlation and are left white in Figure 4a).

The first calibration approach that was tried used linear regression of the high-frequency temperature on high-frequency MXD. The regression slope coefficient was then applied to the unfiltered MXD data, to generate a reconstruction with information at all time scales. The mean of the reconstruction over the calibration period was shifted to match the mean of the observed temperature over the same period (because the high-pass filtering had removed the mean from both data sets, and thus the high-frequency calibration gives no information about the mean level). The reconstruction performs reasonably well during the independent verification period – Figure 4b shows the verification correlation, and Figure 4c shows the verification "Reduction of Error" (RE) statistic (Fritts

et al., 1990). Out of the 109 grid boxes with sufficient data for calibration (Figure 4a), there are sufficient early temperature observations to attempt verification of 74 of them (requiring only 10 or more values in the 1856–1910 period); the remainder are left white in Figures 4b and 4c. All of these 74 boxes have positive verification correlations, while 72 are above 0.2 and 50 are above 0.5 (Figure 4b). The RE statistic is a stricter measure, requiring not just good correlation but also good reproduction of the mean and variance (where, for RE, the mean is measured relative to the mean of the calibration period). A perfect reconstruction would have an RE equal to 1; a common assumption is that any RE > 0 is considered to indicate significant skill. In this reconstruction, 68 out of the 74 verified grid boxes achieve an RE > 0, almost half have RE > 0.3 and 20 have RE > 0.4 (Figure 4c). More than a third of the grid boxes actually have a verification RE that exceeds the equivalent statistic computed over the calibration period.

On the basis of the box-by-box calibration and verification, this first attempt at a reconstruction resolves the spatial variations of summer temperature with significant skill, both in statistical and practical terms. A problem arises, however, when spatial averages of the reconstruction are calculated, because these almost invariably have sub-optimal variance (i.e., the comparison of spatially-averaged observed temperatures with spatially-averaged reconstructed temperatures indicates that a better fit could be obtained by multiplying the reconstruction by a factor greater than one). An example is given in Figure 5a, where the average temperature north of 50°N is compared for the observations and the reconstructions (both data sets are masked by the same time-dependent mask, using only those years and grid boxes where both sets have data values). The correlation between these large-region average series is 0.84, but the variability is clearly underestimated in the reconstruction (especially at the decadal scales shown in Figure 5a). This underestimate in the variability is *more than the usual loss of variance due to regression*: an optimal (i.e., least squares) fit of the spatially-averaged reconstruction to the spatially-averaged temperature is actually achieved by scaling the former by 1.7 (and an appropriate shift in the mean level), shown by the dashed curve in Figure 5a. [The actual value of this factor depends upon the period over which the fit is optimised – here it is the 1856–1960 period to avoid the "decline" feature that is clearly apparent after 1960 – but it is always greater than one.]

There are two possible contributors to this underestimate of the variance of spatial averages. First, the box-by-box calibration was undertaken using high-pass filtered data, and it is possible (Osborn and Briffa, 2000) that this leads to an underestimate of the regression slope coefficient compared with unfiltered data. The recent decline feature was artificially removed from the MXD data to allow calibration using unfiltered data (see section 4.3), but the resulting reconstruction still underestimated the variance of spatial averages. The second, and thus more likely, possibility is that the sub-optimal variance is a generic feature of the box-by-box calibration itself, because the greater relative noise level at the smaller spatial scales tends to weaken the regression slope coefficients. This results in the loss of variance that is usual when performing regression, but because no spatial correlation information is used (implicitly or explicitly) this loss of variance is not reduced when forming spatial averages, despite the lower relative noise level at the larger spatial scales.

A simple example can make this clear. Suppose that all points in a region have the same local temperature variance, σ_T^2, and proxies (e.g., tree-ring density chronologies) exist at all locations and all have the same local reconstruction "skill", measured by the correlation $r(X,T)$ between a proxy series X and its corresponding local temperature series T. If all proxies also have the same linear sensitivity to temperature, then each records the temperature variations plus an "error term" according to:

$$X = T + \varepsilon \tag{1}$$

where ε is the error time series. A key assumption made next is that the error term is due to local non-climatic processes and is independent between sites (i.e., spatially uncorrelated). The earlier assumption of equal skill for all proxies means that the variance of the error term is also the same at all sites, so that:

$$r^2(X,T) = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_\varepsilon^2} \tag{2}$$

If all series have zero mean, then the inter-site correlation between two temperature series is:

$$r(T_1,T_2) = \frac{\sum T_1 T_2}{n\sigma_T^2} \tag{3}$$

and the inter-site correlation between two proxy series is:

$$r(X_1, X_2) = \frac{\sum (T_1 + \varepsilon_1)(T_2 + \varepsilon_2)}{n\left(\sigma_T^2 + \sigma_\varepsilon^2\right)} = r(T_1, T_2) r^2(X, T) \tag{4}$$

The spatial correlation of the proxy data is clearly less than the spatial correlation of the temperature data, because the square of the local proxy skill is less than one (this behaviour was verified by estimating the correlation decay length of the gridded temperature data and the gridded tree-ring density data – 1400 km and 750 km, respectively).

Each proxy record is then calibrated using linear regression against its local temperature, to produce a local temperature reconstruction:

$$\hat{T} = bX \tag{5}$$

According to earlier work on the averaging of spatially-correlated time series (e.g., Kagan, 1966; Yevjevich, 1972; Wigley *et al.*, 1984; Osborn *et al.*, 1997), the average of the temperature series from *n* locations has a variance ($\sigma^2_{\Sigma T/n}$) given by:

$$\frac{\sigma^2_{\Sigma T/n}}{\sigma_T^2} = \frac{1 + (n-1)\bar{r}_T}{n} \quad \rightarrow \bar{r}_T \quad as \quad n \rightarrow \infty \tag{6}$$

where $\bar{r}_T$ is the average inter-site correlation between the *n* temperature series. A similar equation is obtained for the variance of the average of the reconstructed series from *n* locations:

$$\frac{\sigma^2_{\Sigma bX/n}}{\sigma_{bX}^2} = \frac{1 + (n-1)\bar{r}_{bX}}{n} \quad \rightarrow \bar{r}_{bX} \quad as \quad n \rightarrow \infty \tag{7}$$

where $\bar{r}_{bX}$ is the average inter-site correlation between the *n* reconstructed series. Given that calibration by linear regression does not alter the correlation, $\bar{r}_{bX}$ is equal to the average inter-site correlation between the *n* raw proxy series. From equation (4), each inter-site correlation between a pair of proxy series is smaller than the inter-site correlation between the corresponding pair of temperature series by a factor $r^2(X,T)$, and so equations (6) and (7) combine to give, as $n \rightarrow \infty$:

$$\left[\frac{\sigma^2_{\Sigma bX/n}}{\sigma^2_{\Sigma T/n}}\right] \Big/ \left[\frac{\sigma^2_{bX}}{\sigma_T^2}\right] = r^2(X,T) \quad \leq 1 \tag{8}$$

Thus, the variance of the regional average of many locally-calibrated proxy records (relative to the

variance of the regionally-averaged temperature) is lower than the variance of the individual records (relative to the variance of the individual temperature series). Given that the signal-to-noise ratio of the proxy data does, in fact, *increase* following regional averaging, it might be expected that the relative variance explained by the reconstruction at regional scales would be greater than the relative variance explained at individual locations. The fact that it is not (and is actually less) is the direct result of the calibration at local scales, on a site-by-site basis, and the loss of variance is greater for lower proxy skill and for larger and/or less coherent regions [because, for $n < \infty$, both n and $\bar{r}_T$ would appear in equation (8)].

This analysis, confirmed by the results shown in Figure 5a and also by forming other regional averages (not shown) from the locally-calibrated reconstruction and comparing them against the existing regional reconstructions of Briffa *et al.* (2002a), demonstrates a clear drawback of the box-by-box local calibration of the reconstruction: it is not possible to optimise the reconstruction at the grid-box spatial scale *and* the regional (or hemispheric) average scale, at the same time. Other methods may not suffer from this problem – e.g., those that reconstruct patterns, such as EOFs, rather than individual boxes, or those that take the box-by-box approach but use non-local information in the calibration.

To continue with the approach selected here requires a choice to be made between optimising the reconstruction at either local or regional-average scales (or some intermediate scale). The results reported in the remainder of this paper relax the requirement of optimising the fit for each grid box (which the local linear regression achieved) by instead scaling each grid-box MXD series so that its variance matches the variance of the observed temperature series for that grid box. A "variance-matching" approach has been used by other studies, typically at a hemispheric scale (e.g., Jones *et al.*, 1998; Crowley and Lowery, 2000). Though the variance matching does not give the optimal fit at the grid-box scale (i.e., the root-mean-squared error is not minimised), it yields a spatially-resolved reconstruction that, when regionally-averaged, matches the regionally-averaged temperature much more closely.

The method reported above for the linear regression was repeated but using variance

matching. Thus the temperature series were high-pass filtered before calculating their variance, and the normalised density series were high-pass filtered before estimating the scaling factors required to make their high-frequency variance match that of the observed temperatures. These scaling factors (Figure 6a) were then applied to the unfiltered density series, and the mean was shifted to match the mean of the observed temperature over the calibration period. The correlations during calibration (Figure 4a) and verification (Figure 4b) are unchanged using this approach, but the RE values shown earlier in Figure 4c are not valid because the variance matching does not minimise the root-mean-squared error. The new verification RE values are generally, of course, lower than those obtained by linear regression, though 50 out of 74 are positive and might still be considered significant.

For the six boxes in northern and north-eastern Siberia that had insufficient temperature observations, calibration coefficients (i.e., the scaling factor and the mean value) were estimated by a distance-weighted average of neighbouring values. Figure 6a shows the calibration scaling factors, which reflect the pattern of inter-annual summer temperature variance, being largest in north-central Canada and Eurasia, and smallest at lower latitudes.

There is now closer agreement between the reconstruction and observed temperature when spatial averages are calculated. Figure 5b shows the equivalent results as Figure 5a (i.e., the average over all boxes north of 50°N that have both observed and reconstructed temperatures), but for the "variance-matching" local calibration. The reconstruction variance is greater than in Figure 5a, and is near to optimal. Using linear regression to estimate the best fit of the spatially-averaged reconstruction to the spatially-averaged observations, over the 1856–1960 period, results in a scaling factor of less than 1.1 (i.e., the variance is almost optimal for this particular spatial averaging). The optimal fit is shown in Figure 5b by the dashed curve, and involves a shift in the mean (due to the 1856–1960 regression period, compared with the 1911–1990 period used for local grid-box calibration) but virtually no change in the variance. Similarly improved results are obtained when other regional averages of the gridded reconstructions are calculated and compared with those regional reconstructions published by Briffa *et al.* (2002a).

4.3. Artificial removal of the recent decline in tree-ring density

Despite calibrating the gridded reconstructions with high-pass filtered data, in an attempt to avoid problems that would otherwise be introduced by the recent decline in high-latitude tree-ring density, there are still two issues that must be addressed. The first is that Osborn and Briffa (2000) identified that the tree-density response to temperature changes might be time-scale dependent, and that calibration coefficients obtained using only the high frequencies may be lower than those applicable at multi-decadal (and possibly longer) time scales. The second is that Figure 5b shows clearly that the presence of the decline makes the mean level of the reconstructed temperature very sensitive to the choice of calibration period (compare the thick and dashed lines).

To overcome these problems, the decline is artificially removed from the calibrated tree-ring density series, for the purpose of making a final calibration. The removal is only temporary, because the final calibration is then applied to the unadjusted data set (i.e., without the decline artificially removed). Though this is rather an *ad hoc* approach, it does allow us to test the sensitivity of the calibration to time scale, and it also yields a reconstruction whose mean level is much less sensitive to the choice of calibration period.

Briffa *et al.* (1998a, 2002a) demonstrated that the decline in tree-ring density in this data set, relative to summer temperature, is a strong signal only at large spatial and temporal scales, and influences mainly the high latitudes. The difference (Figure 5c) between the observed (thick curve in Figure 5b) and reconstructed (thin curve in Figure 5b) temperatures averaged over all areas north of 50°N is, therefore, an appropriate measure of the decline. To focus on the longer temporal scales, a function is fit to the difference curve, consisting of a constant line before 1930 and a quadratic in time after 1930. This function is used as a measure of the density decline through time (Figure 5c), and was extended at a constant level back to 1400 (the constant level is thus the "normal" level, and is only non-zero because of the period used to calibrate the gridded density data).

The difference between observed and reconstructed temperature was also computed for every grid box and year for which both have values. Each grid-box difference series was then smoothed with a 10-year filter (again to emphasize the longer time scales) and correlated and regressed against the smoothed measure of the density decline (Figure 5c) over as much of the 1856–1994 period as

possible. Correlations (not shown) are in the range 0.5 to 0.9 over many of the high latitudes (central and eastern Canada, and northern Eurasia from central Scandinavia to eastern Siberia), and are near zero in the far west of North America and in north-west and south-east Europe. The pattern is quite similar to that identified by Briffa *et al.* (1998a) and shown in their Figure 1a. The pattern of regression coefficients is also quite similar. The smoothed measure of density decline (Figure 5c) was then multiplied by the pattern of regression coefficients, to represent the density decline in space and time, and this decline signal was subtracted from the calibrated density fields. This procedure simply (and entirely artificially) brings the calibrated grid-box density data closer to the grid-box temperatures, according to the fitted function in Figure 5c and by an amount that depends on the similarity between that fitted function and the density–temperature difference history in each box.

4.4. Final calibration and verification

The artificially-adjusted, calibrated density data were then re-calibrated against the observed temperature data, again by scaling and shifting each density time series to match the variance and mean of the observed temperature data over the 1911–1990 calibration period. This time, however, neither data set was high-pass filtered. Because the density data were already calibrated, most of the re-calibration scaling factors are close to one (Figure 6b). In fact, out of the 109 grid boxes where calibration was possible, 107 factors were between 0.8 and 1.2, and 97 were between 0.9 and 1.1. The main region where they were consistently above 1.1 is southern Europe, a region that Osborn and Briffa (2000) identified as having the most significant time-scale dependence in the calibration of this tree-ring density data against warm-season temperatures. Thus the use of unfiltered data in the calibration, made possible by artificially removing the decline, prevents the calibration being biased to the high-frequency response only. Overall, however, the fact that the re-calibration scaling factors are close to one means that the reconstructed temperatures are relatively insensitive to the secondary calibration and to the artificial removal of the recent decline in density. The magnitude of the reconstructed variability is very similar to that obtained with the simpler calibration against the high-pass filtered data.

The mean levels of the reconstructed temperatures, however, do change in the re-calibration,

and that turns out to be the main reason for selecting the final calibration rather than the earlier calibration. The mean level is shifted downwards in those regions where the decline is most apparent (i.e., the high latitudes) by around 0.5 K. The mean level is also much more robust to the choice of the calibration period (e.g., changing it to 1856–1960 alters the mean level by only 0.03 K when averaged over all the high latitudes).

The re-calibration coefficients were then applied to both versions of the calibrated, gridded MXD data, with and without the artificial removal of the recent decline. The version without this artificial adjustment is the "preferred" summer temperature reconstruction that should normally be analysed and presented. The version with the artificial adjustment, however, is useful if the data are to be used for further calibration exercises, provided that the artificial adjustment of the data is clearly acknowledged (e.g., section 5 of this paper that uses the data to predict past temperatures over a larger region, or Briffa *et al.*, 2002b, where the data are used to predict past variations in atmospheric circulation).

The final temperature reconstructions have the same verification correlations as those presented in Figure 4a, because the changes introduced by using a "variance-matching" approach and the derivation of re-calibration coefficients using the artificially-adjusted data only affect the mean and variance of the reconstruction, not the temporal sequence or relative magnitude of individual yearly values. The verification RE statistics are different, however, and are shown in Figure 4d. There is a clear improvement following the final calibration, with 60 out of 74 boxes achieving an RE > 0 (following the matching of the high-frequency variance stage, only 50 had a positive RE). There is also some improvement in the skill, as measured by the RE, compared with the linear regression against the high-frequency variations (Figure 4c). The variance matching does not optimise the local fit, which is why 14 boxes have RE < 0 in Figure 4d, whereas only 6 have RE < 0 in Figure 4c. But the final calibration does improve the mean level of the reconstruction, and this is reflected by the number of boxes with higher RE values: 43 are above 0.3 (previously 32), 30 are above 0.4 (previously 20) and 20 are above 0.5 (previously 7). Though the variance-matching approach does not optimise the local fit, the results are clearly a significant improvement over simple climatology, even at the local scale.

5. Extending the spatial coverage by principal component regression

Warm-season temperature reconstructions with extended spatial coverage have also been developed, making use of the spatial correlation evident in temperature variability to predict past temperatures even in grid boxes without any tree-ring density data. The calibration was undertaken on a box-by-box basis, and each grid-box temperature series was predicted using multiple linear regression against the leading principal components (PCs) of the calibrated, gridded reconstructions described in section 4.4. The PCs were computed from the correlation matrix of the reconstructions, so the calibration was in effect removed and similar results would have been obtained if the PCs of the raw, gridded density data had been used instead. The only difference is that the calibrated data *with the artificial removal of the recent decline* were used for the PCA. Using the adjusted data avoids the problems otherwise introduced by the existence of the decline (see section 4), though all reconstructions after 1930 will be artificially closer to the real temperatures because of the adjustment (the adjustment is quite small until about 1960 – Figure 5c). Tests with the unadjusted data show that none of the spatial patterns associated with the leading PCs are affected by the adjustment, and the only PC time series that is affected is the leading PC and then only during the post-1930 period. In other words, the adjustment pattern is very similar to the leading EOF pattern, and orthogonal to the others, and thus only influences the first PC time series.

Due to time-varying spatial coverage of the reconstructions (Figures 2c and 3a,c,e,g), different sets of PC time series were estimated for seven different periods (Table 1). The lower spatial coverage earlier in the reconstruction also means that less PCs are required to capture the dominant patterns of variability in the data. The number of PCs used as predictors in each period was, therefore, varied according to the number of grid boxes with chronologies in them (indicated by the shading in Figure 2c) divided by ten, plus one. This resulted in two predictors for the early period (back to 1400), and 12 predictors for the 1822–1976 period (Table 1). The eight PC time series (and associated patterns) obtained for the period 1697–1976 were shown by Briffa *et al.* (2002b) in their Figures 3–10.

For each period in turn, reconstructions were attempted for every grid box in the Northern Hemisphere. The overlap period between the observed warm-season temperature and the tree-density principal components (which all end in 1976; Table 1) was split into two (provided it contained at

least 40 yearly values) and a multiple linear regression model was built using data from the first half, then verified on the second half. The regression model was then calibrated on the second half and tested on the first half. The regression was *not* a screening regression – i.e., all tree-density PCs allowed for the period were used as predictors. If the average of the two verification correlations was less than 0.4 then no reconstruction was attempted for that grid box. If the average correlation did exceed 0.4 then a reconstruction was generated by re-calibrating the multiple regression over the full overlap period, and applying it to predict the grid-box temperature time series for the entire period.

The different periods were combined in the following way. Any reconstructions that were generated for the longest (i.e., 1400–1976) period formed the starting point. For those that were also successfully reconstructed using the regression model from the 1453–1976 period, the values for 1453–1976 were taken from the latter reconstruction (as were any that were successful for 1453–1976, but had been unsuccessful 1400–1976). This was repeated for all periods through to the shortest (Table 1), always favouring reconstructed values that were generated from the greater coverage of the more recent periods. A handful of isolated grid boxes (especially if they were over the ocean) were removed from the reconstruction, in case their successful verification was simply due to chance. Finally, any grid box time series that had already been reconstructed by the local calibration method described in section 4 were used in preference to the estimates based on principal component regression (and they were also used for those grid boxes where the PC regression failed). Thus, the summer temperature reconstruction for the region with tree-ring density chronologies in the grid boxes is the same in both cases, and the PC regression is limited to the extension beyond this core region. The coverage achieved (distinguishing both the core region and the extension) is shown in Figures 3b,d,f,h.

The calibrated reconstructions of warm-season (April-to-September) temperature anomalies obtained from this combination of local calibration and extension using principal component regression are presented for every summer from 1600 to 1887 by Briffa *et al.* (2002b; their Figure 2). Maps and data for the full period from 1400 to 1960 are available from the Climatic Research Unit website and will be submitted to the World Data Center for Paleoclimatology (see section 8 for details). As discussed in section 2.1, these reconstructions, based on "Hugershoff" standardised tree-

ring density data are suitable for studying climate variability on inter-annual to multi-decadal time scales.

6. Incorporating long time scale variations

6.1. Adjustment of the gridded data

The gridded reconstructions described in sections 4 and 5 are likely to be deficient in their representation of low-frequency temperature variability (centennial and longer periods) because the "Hugershoff" standardisation that removed the dependence of a ring's density on the age of the tree when it grew also removed long-term density changes that might have been caused by temperature variability. Briffa *et al.* (2001) described a new technique, that they named "age-band decomposition" (ABD), intended to retain all time scales of variability while preventing changes in mean tree age throughout the length of chronologies from biasing the results. The ABD technique achieves this by initially combining only information from trees of similar age; this is done in absolute density units and values can thus be compared over all time scales (for which trees of this age range are available in the data set). Different bands (or ranges) of tree ages are subsequently normalised and averaged together, for each site and species. Briffa *et al.* (2001) applied the ABD technique to the same data set used in the present study, but only at the regionally-averaged scale.

The availability of the Briffa *et al.* (2001) low-frequency temperature reconstructions on regional-mean scales only, is not a severe limitation given that climate variability on these time scales is likely to be coherent on regional spatial scales (e.g., Jones *et al.*, 1997). For each of the nine regions defined in Figure 1a, the average of the calibrated, gridded temperature reconstructions developed here is compared with the regional-mean reconstruction published by Briffa *et al.* (2001) using the ABD technique to retain low-frequency variations. To assess whether the differences between the ABD reconstructions and the average of the gridded reconstructions are really due to the missing low-frequency variations, rather than (e.g.) a failure of the gridded calibration to correctly express regional variations (section 4.2), a comparison is also made with the regional temperature reconstructions of Briffa *et al.* (2002a), based on the same Hugershoff-standardised density data used for the gridded reconstructions. The average of the gridded reconstructions matches the Hugershoff regional series

quite closely for NEUR (Figure 7a), NSIB (Figure 7b), SEUR, ESIB and WNA (not shown, but see Figure 1a for regional definitions). The match is also close for the more recent centuries of NWNA (Figure 7c) and ECCA (not shown), but deteriorates when there are very few grid boxes in the early centuries. The match is also poor for CAS and TIBP, which comprise very few grid boxes throughout. With fewer grid boxes, the area-means still correlate highly (they are, after all, the same data), but the magnitude of the fluctuations is different (e.g., Figure 7c) because of the variance-matching approach used for grid-box series calibration.

The "missing" low-frequency temperature variability is identified, for each region, as the difference between the 30-year smoothed ABD regional time series and the 30-year smoothed regional average of the gridded reconstructions. Where this difference is not available in the earlier centuries (i.e., before the start of the ABD series, which are generally shorter given the minimum data requirements imposed by Briffa *et al.*, 2001), it is infilled by linearly interpolating between the earliest value and an assumed difference of 0 K at a time 50 years before the earliest value. A 0 K difference is assumed from that point back to 1400. Given the disagreement between gridded and regional data described above for CAS and TIBP, the difference series were not computed for those two regions.

It was assumed that each "difference series" is representative of a location at the centroid of each region (defined from the coverage of boxes containing data, Figure 3a). The additional low-frequency variability added to each of the 115 locally-calibrated grid-box temperature reconstructions was computed as a weighted average of the regional "difference series", with the weighting determined by a Guassian-shaped function (with "standard deviation" of 1200 km) of the distance between the centre of the grid box and the centroid of each region. Because all data sets used in this process were calibrated (i.e., all are temperature anomalies in Kelvin, relative to the 1961–1990 mean of the observed temperature data), the resulting data with additional low-frequency variability is automatically in those units too, and no further calibration is required.

The regional averages of the low-frequency gridded temperature reconstructions (Figure 7) are now much closer to the regional reconstructions of Briffa *et al.* (2001). The largest difference, out of all the regions, is that shown for NSIB in Figure 7b; the reason why the adjustment has not produced an exact match between the gridded temperature and the regional reconstructions is because the

adjustment applied was a weighted average of the NSIB adjustment and the surrounding (smaller) adjustments for NEUR and ESIB.

6.2. Extending the spatial coverage by principal component regression

The same procedure as described in section 5 was used to extend the spatial coverage of the gridded reconstructions that now have additional low-frequency variability. The principal component regression again used a calibration of the leading PCs of the gridded reconstructions, to attempt to estimate past temperatures in grid boxes without any tree-ring density data. As before, the recent decline in the tree-ring density would hamper this calibration, so the additional low-frequency variability was also added to the version of the reconstructions that had been adjusted to artificially remove the recent decline, and it was the PCs of this data set that were used as predictors in the principal component regression. Thus, the additional low-frequency variability was *not* explicitly added to all locations in the extended coverage reconstruction; instead it was added only to the 115 boxes of the core data set that have tree chronologies in them. The low-frequency variability then appears in the PCs of this core data set, and thus also appears in the extended coverage reconstructions that are predicted from those PCs. The exact spatial coverage of the extended reconstructions differs very slightly between the versions with and without additional low-frequency variability, because the verification criterion (see section 5) that was used to determine where a reconstruction would be generated was met at a slightly different set of boxes in each case.

7. Visualisation and comparison of reconstructions

Detailed analysis of the reconstructions documented in this paper, and comparison with other reconstructions, will be (or is already) published elsewhere. For example, Briffa *et al.* (2002b) performed a principal component analysis of the gridded (but not spatially extended) calibrated reconstructions (without the additional low-frequency variations), and related each mode to patterns of temperature and atmospheric circulation variability. Rutherford *et al.* (2005) compare the gridded (but not spatially extended) calibrated reconstructions (including additional low-frequency), with further reconstructions obtained using an alternative calibration method, and also with reconstructions derived from the Mann *et al.* (1998) multi-proxy network. In the present paper, analysis is limited to two

comparisons that are included to further assess the reliability of the reconstructions and to gauge the importance of the additional low-frequency variability derived from the work of Briffa *et al.* (2001).

Figure 8 compares the observed warm-season temperature with that reconstructed, for every year between 1856 and 1960, for every longitude, and averaged over all latitudes north of 45°N. The same mask is applied to both data sets, such that only grid boxes and years where both sets have values are included in the averaging. There is a clear correspondence between the sequencing and longitudinal location of many warm and cold anomalies (e.g., the European warmth of the late 1930s, among many other examples). There are also some discrepancies, such as the extent of the North American warmth around 1900. Overall, the longitude–time fields in Figure 8 have a pattern correlation of 0.72, rising to 0.77 if the series at each longitude is first smoothed with a 10-year filter (not shown), and falling only to 0.68 if the comparison is limited to the 1856–1910 verification period.

Figure 9 compares the longitude–time evolution of warm-season temperature during the full reconstruction period (1400–1960) for the reconstruction without (Figure 9a) and with (Figure 9b) additional low-frequency variability. The results are sensitive to the choice of baseline period and to the structure of climate variations within that period if a relatively short, 30-year baseline is used (e.g., the 1961–1990 period used for Figure 8). For example, the temperature anomalies in Figure 8 (both observed and reconstructed) between 70°E and 80°E are cooler than those immediately to the east and west for most of 1880–1960. Is it true though to say that an "anomalous" zonal temperature structure persisted for these 80 summers, or alternatively that the 70°E to 80°E temperature was anomalously warm for the shorter 1961–1990 baseline? There is no answer, because the "normal" climate is unknown (nor, indeed, is such a concept valid if the climate is continuously varying). In both panels of Figure 9, however, the occurrence of longitudinal banding is reduced by expressing the data as anomalies from a longer baseline mean, 1801–1900.

The intra-centennial variations in Figure 9a are dominated by a number of cool and warm fluctuations that influence a wide span of longitudes. Briffa *et al.* (1998b) attribute some of the cool periods to explosive volcanic eruptions, though it is interesting to note that their influence is strong even at the decadal time scales shown in Figure 9a. Decadal-scale cool periods evident across most of the reconstructed longitudes occur in the 1450s, the first decade of the 1600s, the 1640s, 1690s and

1810s (also the 1830s and 1880s, though with more limited extent). All these periods are associated with explosive volcanic eruptions (Briffa *et al.*, 1998b).

The additional low-frequency fluctuations in Figure 9b are most obvious during the extended cool period over Eurasia from 1580 to 1710 (also slightly cooler over North America). Smaller differences are the strengthened 15th century warmth over eastern Europe and western Russia, with enhanced coldness over eastern Russia in the same period, and slightly strengthened 20th century warming throughout.

8. Summary and discussion

The maximum latewood densities from a network of mid-to-high latitude tree-ring chronologies have proved to be a valuable source of information for estimating past variations in summer temperature (specifically, April-to-September temperature). One possible calibration approach has been documented in this paper, for generating spatially-resolved temperature fields for each summer that are also consistent with previously published large area-average reconstructions based on the same proxy data. Many alternative calibration procedures could have been used (e.g., Rutherford *et al.*, 2005), and differences between the results obtained from each method are an additional source of reconstruction uncertainty.

The tree-ring density chronologies were first assembled onto a regular grid, and then each grid box was calibrated and verified against local temperature. Using simple linear regression for the local calibration was apparently successful at the individual grid-box scale. However, the loss of variance (evident at the grid-box scale) that is usual for any regression-based reconstruction is *increased* when spatial-mean time series are computed from the gridded reconstruction. This contrasts with a *reduction* in lost variance that was expected on the basis of the lower relative noise level at the larger spatial scales. Area averages of the gridded reconstructions thus had lower than optimal variance and did not reproduce previously-published regional temperature reconstructions (Briffa *et al.*, 2001, 2002a), even though they were based on the same density data. A theoretical example was developed to demonstrate that the sub-optimal variance of regional averages was a result of the box-by-box calibration using regression against only local information (i.e., spatial covariance information was not

used).

Each grid box has, instead, been calibrated by scaling its variance to match the variance of the observed summer temperature. This produces a non-optimal fit at the grid-box scale (though more than half of all verification correlation and "reduction of error" statistics were still above 0.5 and 0.3, respectively). The fit was near-optimal at the regional-average scale, however, achieving a correlation of 0.84 with temperature when averaged over all boxes north of 50°N, for example.

The issue of how to optimise the calibration at both grid-box and regional-average scales deserves further investigation, though the method used here performs reasonably well as a compromise between the two. Methods that use non-local information when generating a grid-box temperature reconstruction may reduce this problem, though optimising the reconstruction skill over all possible spatial scales may still be difficult. Methods that do not calibrate at individual grid boxes, such as those that, instead, attempt to reconstruct the past variability of temperature principal component time series (and thus of a set of associated temperature patterns, which can be summed to give a spatially-resolved temperature reconstruction), may not suffer from this problem at all (e.g., Mann *et al.*, 1998; Luterbacher *et al.*, 2004).

The second key issue that arose during the calibration procedure is more specific to the tree-ring density data set used here, because it relates to the decline (relative to that expected on the basis of observed summer temperatures) in density over recent decades at the high latitudes (Briffa *et al.*, 1998). It is extremely important to try to identify the cause of this decline, though investigations are currently hampered by the early sampling of many of the sites and thus the lack of widespread data since the mid-1980s (Briffa *et al.*, 2003). Without a satisfactory explanation, we make the untested assumption that the decline is due to an anthropogenic factor that did not occur earlier in the reconstruction period. Nevertheless, additional uncertainty must surely be associated with the reconstructions because of this assumption, particularly for earlier warm periods. The decline also complicates the reconstruction method. To prevent it from unduly influencing the calibration, the grid-box calibration was first undertaken using high-pass filtered data (the decline was removed by the filtering). The tree-ring density data were also adjusted, in an artificial way, to temporarily remove the decline; this made only a small difference in comparison with the filtered calibration, although it

reduced the sensitivity of the reconstruction's mean level to the choice of calibration period and thus proved a useful, if *ad hoc*, way of dealing with the decline during calibration.

In addition to the local calibration of each grid box, the coverage of the temperature reconstructions was extended by using principal component regression (again, applied to predict each grid box separately). Coverage was more than doubled by this method, even though only those boxes for which the regression had an independent verification correlation of over 0.4 were included. The longitude–time pattern of reconstructed temperature variability, averaged over mid-to-high latitudes, has a correlation with observations of 0.72 overall, and 0.68 over the independent verification period.

The issue of century time scale variability has also been addressed. Variations on these long time scales were reduced in the reconstructions developed in sections 4 and 5, because they were removed during the standardisation of the tree-ring chronologies. Spatially-resolved summer temperature reconstructions, developed in section 6, include variability on longer time scales, due to the simple addition of the low-frequency (i.e., longer than 30-year) differences between the reconstructed temperatures developed here and the regional reconstructions of Briffa *et al.* (2001). The additional signal has been added such that it varies smoothly in space, with a scale determined by the size of the regions used by Briffa *et al.* (2001).

Quantification of reconstruction uncertainty (i.e., error) is very important, especially if the data are to be used for comparison with climate model simulations (Collins *et al.*, 2002). Error estimation was not attempted for the gridded fields described here, however, because the reconstructions were not optimised at the grid-box scale (and thus the mean-squared errors were not minimised). Thus individual grid-box values in the reconstruction should be treated with caution, but the values in each grid box are independently generated (for the core region where local calibration was undertaken rather than principal component regression) and so spatially coherent clusters of similar anomalies are likely to provide a reliable indication of the location and relative magnitude of significant temperature anomalies.

Previous analysis (Briffa *et al.*, 1998b) of these data had shown the importance of volcanic activity on summer temperature for extreme years; analysis of the reconstructions described here demonstrates that the impact of volcanic events is apparent even at the decadal time scale. The

reconstructions will, however, be useful for much more detailed analyses of past climate variability. The maximum latewood density data sets thus provide an excellent resource and are available from the Climatic Research Unit website (www.cru.uea.ac.uk/~timo/datapages/mxdtrw.htm).

Acknowledgements. This study was supported by the Commission of the European Communities (SOAP, EVK2-CT-2002-00160), the UK Natural Environment Research Council (Rapid Climate Change, NER/T/S/2002/00440) and the UK Met Office (PB/B3539).

References

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

Briffa, K.R., Osborn, T.J., 2002. Blowing hot and cold. Science 295, 2227-2228.

Briffa, K.R., Jones, P.D., Wigley, T.M.L., Pilcher, J.R., Baillie, M.G.L., 1986. Climate reconstruction from tree rings: part 2, spatial reconstruction of summer mean sea-level pressure patterns over Great Britain. Journal of Climatology 6, 1-15.

Briffa, K.R., Jones, P.D., Schweingruber, F.H., Karlén, W., Shiyatov, S.G., 1996. Tree-ring variables as proxy-indicators: problems with low-frequency signals. In: Jones, P.D., Bradley, R.S., Jouzel, J. (Eds.), Climatic variations and forcing mechanisms of the last 2000 years. NATO ASI series 141, Springer-Verlag, Berlin, pp. 9-41.

Briffa, K.R., Schweingruber, F.H., Jones, P.D., Osborn, T.J., Shiyatov, S.G., Vaganov, E.A., 1998a. Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. Nature 391, 678-682.

Briffa, K.R., Jones, P.D., Schweingruber, F.H., Osborn, T.J., 1998b. Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. Nature 393, 450-455.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., Harris, I.C., Jones, P.D., Shiyatov, S.G., Vaganov, E.A., 2001. Low-frequency temperature variations from a northern tree ring density network. Journal of Geophysical Research 106, 2929-2942.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., Jones, P.D., Shiyatov, S.G., Vaganov, E.A., 2002a. Tree-ring width and density data around the Northern Hemisphere: part 1, local and regional climate signals. The Holocene 12, 737-757.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., Jones, P.D., Shiyatov, S.G., Vaganov, E.A., 2002b. Tree-ring width and density data around the Northern Hemisphere: part 2, spatio-temporal variability and associated climate patterns. The Holocene 12, 759-789.

Briffa, K.R., Osborn, T.J., Schweingruber, F.H., 2003. Large-scale temperature inferences from tree rings: a review. Global and planetary change 40, 11-26 (doi:10.1016/S0921-8181(03)00095-X).

Chenoweth, M., 1993. Nonstandard thermometer exposures at United-States cooperative weather stations during the late-19th-century. Journal of Climate 6, 1787-1797.

Collins, M., Osborn, T.J., Tett, S.F.B., Briffa, K.R., Schweingruber, F.H., 2002. A comparison of the variability of a climate model with palaeo-temperature estimates from a network of tree-ring densities. Journal of Climate 15, 1497-1515.

Cook, E.R., Briffa, K.R., Meko, D.M., Graybill, D.S., Funkhouser, G., 1995. The 'segment length curse' in long tree-ring chronology development for paleoclimatic studies. The Holocene 5, 229-237.

Crowley, T.J., Lowery, T.S., 2000. How warm was the Medieval Warm Period? Ambio 29, 51-54.

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

Fritts, H.C., 1991. Reconstructing large-scale climatic patterns from tree-ring data. University of Arizona Press, Tucson, 286pp.

Fritts, H.C., Guiot, J., Gordon, G.A., 1990. Verification. In: Cook, E.R., Kairiukstis, L.A. (Eds.), Methods of Dendrochronology. IIASA/Kluwer, Dordrecht, pp. 178-185.

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

Jones, P.D., Osborn, T.J., Briffa, K.R., 1997. Estimating sampling errors in large-scale temperature averages. Journal of Climate 10, 2548-2568.

Jones, P.D., Briffa, K.R., Barnett, T.P., Tett, S.F.B., 1998. High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control run temperatures. The Holocene 8, 455-471.

Jones, P.D., New, M., Parker, D.E., Martin, S., Rigor, I.G., 1999. Surface air temperature and its variations over the last 150 years. Reviews of Geophysics 37, 173-199.

Jones, P.D., Osborn, T.J., Briffa, K.R., 2003. Changes in the Northern Hemisphere annual cycle: implications for paleoclimatology? Journal of Geophysical Research 108, 4588 (doi:10.1029/2003JD003695).

Kagan, R.L., 1966. An evaluation of the representativeness of precipitation data (in Russian), Gidrometeoizdat, 191pp.

Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., Wanner, H., 2004. European seasonal and annual temperature variability, trends, and extremes since 1500. Science 303, 1499-1503.

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

Mann, M.E., Bradley, R.S., Hughes, M.K., 1999. Northern Hemisphere temperatures during the last millennium: inferences, uncertainties, and limitations. Geophysical Research Letters 26, 759-762.

Moberg, A., Alexandersson, H., Bergstrom, H., Jones, P.D., 2003. Were southern Swedish summer temperatures before 1860 as warm as measured? International Journal of Climatology 23, 1495-1521.

Osborn, T.J., Briffa, K.R., 2000. Revisiting timescale-dependent reconstruction of climate from tree-ring chronologies. Dendrochronologia 18, 9-26.

Osborn, T.J., Briffa, K.R., Jones, P.D., 1997. Adjusting variance for sample-size in tree-ring chronologies and other regional-mean timeseries. Dendrochronologia 15, 89-99.

Rutherford, S.D., Mann, M.E., Osborn, T.J., Bradley, R.S., Briffa, K.R., Hughes, M.K., Jones, P.D., 2005. Proxy-based Northern Hemisphere surface temperature reconstructions: sensitivity of method, predictor network, target season and target domain. Journal of Climate 18, 2308-2329.

Schweingruber, F.H., Briffa, K.R., 1996. Tree-ring density networks for climate reconstruction. In: Jones, P.D., Bradley, R.S., Jouzel, J. (Eds.), Climatic variations and forcing mechanisms of the last 2000 years. NATO ASI series 141, Springer-Verlag, Berlin, pp. 43-66.

Schweingruber, F.H., Briffa, K.R., Jones, P.D., 1991. Yearly maps of summer temperatures in western Europe from AD 1750 to 1975 and western North America from 1600 to 1982: Results of a radiodensitometrical study on tree rings. Vegetatio 92, 5-71.

Vaganov, E.A., Shiyatov, S.G., Mazepa, V.S., 1996. Dendroclimatic study in Ural–Siberian subarctic (in Russian). Nauka, Novosibirsk, Russia, 245pp.

Vaganov, E.A., Hughes, M.K., Kirdyanov, A.V., Schweingruber, F.H., Silkin, P.P., 1999. Influence of snowfall and melt timing on tree growth in subarctic Eurasia. Nature 400, 149-151.

Wigley, T.M.L., Briffa, K.R., Jones, P.D., 1984. On the average value of correlated time series with applications in dendroclimatology and hydrometeorology. Journal of Climate and Applied Meteorology 23, 201-213.

Yevjevich, V., 1972. Probability and statistics in hydrology, Water Resources Publications, 302pp.

Tables

Table 1: Details of the principal component regressions for seven reconstruction periods, for the "Hugershoff" standardised reconstructions.

Period	Nbox with trees[1]	Nbox infilled[2]	Number of PCs	Nbox successful[3]	Nbox removed[4]	Nbox extra[5]	Nbox total[6]
1822–1976	109	109	12	197	11	150	259
1743–1976	91	103	10	180	8	139	242
1697–1976	72	92	8	166	6	125	217
1660–1976	56	76	6	124	8	92	168
1583–1976	37	62	5	98	2	80	142
1453–1976	15	39	3	67	4	63	102
1400–1976	7	26	2	73	2	59	85

[1]Number of grid boxes with tree-ring chronologies at the start of the period

[2]Number of grid boxes after infilling of the MXD data (see section 3)

[3]Number of grid boxes for which the PC regression was successfully verified

[4]Number of successful grid boxes that were removed because they were isolated ocean boxes

[5]Number of grid boxes for which the PC regression was successful, and which did not already have a locally-calibrated reconstruction already available

[6]Number of grid boxes reconstructed by combining PC regression and local calibration coverage (i.e., the sum of "Nbox infilled" and "Nbox extra")

Figure captions

Figure 1: (a) Location (circles) of the 387 sites with tree-ring density chronologies, together with the boundaries of nine regions: NEUR = northern Europe, SEUR = southern Europe, NSIB = northern Siberia, ESIB = eastern Siberia, CAS = central Asia, TIBP = Tibetan Plateau, WNA = western North America, NWNA = northwestern North America, and ECCA = eastern and central Canada. (b) Locations of the 341 chronologies that were retained (open circles) and the 46 that were discarded (filled circles).

Figure 2: (a) Number of tree-ring density chronologies in the network that have values in each of the years from 1400 to 1994. (b) The fraction of tree cores within each chronology (f) that have measurements in each year, averaged over only those chronologies that have some data in that year (i.e., excluding those with $f = 0$). (c) Number of grid boxes with tree-ring density values in each of the years from 1400 to 1994; grey shading shows the number with real measurements, black line shows the number after extension of some series by regression against their neighbours.

Figure 3: Spatial coverage of the gridded data set, with real density measurements (black boxes in a, c, e, and g), with density data extended by regression against neighbours (grey boxes in c, e and g), with temperature estimates obtained by local calibration of density data (black boxes in b, d, f and h) and temperature estimates obtained using principal component regression (grey boxes in b, d, f and h). Coverage is shown for four periods: (a, b) full coverage 1890–1976; (c, d) data back to 1400; (e, f) data back to 1583; and (g, h) data back to 1697.

Figure 4: (a) Correlations between high-pass filtered density and temperature over the correlation period, 1911–1990. (b) Correlations between unfiltered density and temperature over the verification period, 1856–1910. (c) RE statistic computed over the 1856–1910 verification period for the reconstructions generated by linear regression between high-pass filtered density and temperature. (d) As (c), except for the reconstructions generated by matching the variance between the decline-adjusted density and temperature. The black outline marks the region with tree density data, and white boxes within this outline are those for which there was insufficient temperature data for either calibration or verification.

Figure 5: (a, b) Observed (thick line) and reconstructed (thin line) temperature averaged over all boxes north of 50°N where both observations and reconstructions have data, and the 1856–1960 least-squares fit (dashed line) of the thin curve on the thick curve. Reconstructions are generated by (a) linear regression, and (b) variance matching, between high-pass filtered density and temperature series. (c) The difference between the observed and reconstructed series in (b), together with a smooth function fitted to the difference series. All curves have been smoothed with a 10-year filter.

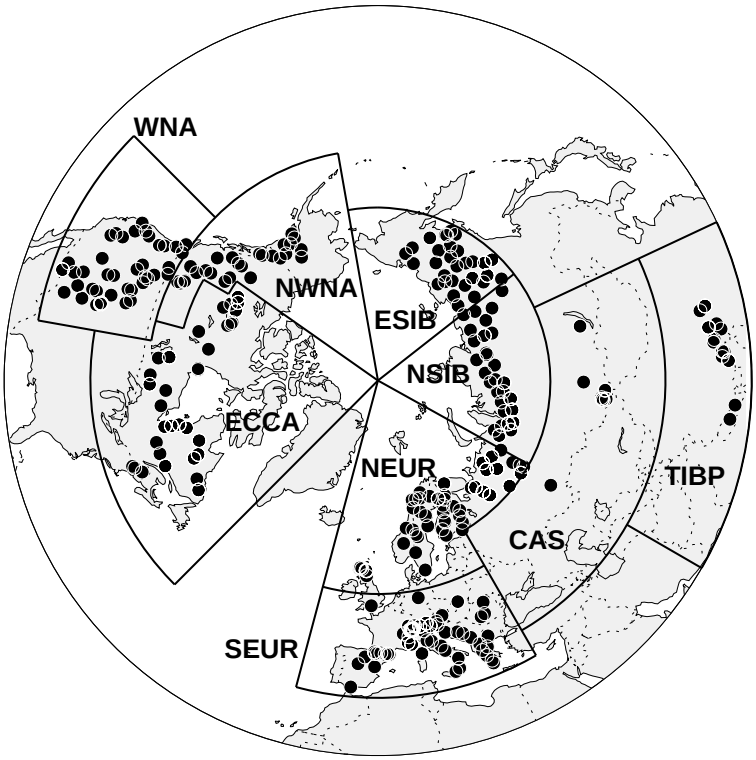
Figure 6: Scaling factors applied to the normalised tree density data to calibrate them into temperature reconstructions: (a) factors required to make the high-frequency variance match the observed temperature high-frequency variance; and (b) factors that are subsequently applied based on the variance of the unfiltered "decline-adjusted" calibrated density and the variance of the unfiltered observed temperature.

Figure 7: Regional-average temperature reconstructions for (a) northern Europe; (b) northern Siberia; and (c) northwestern North America. Continuous lines are the regional reconstructions published by Briffa *et al.* (2002a) using Hugershoff-standardised tree density (thin lines) and by Briffa *et al.* (2001) using ABD-processed tree density (thick lines). Dashed lines are the average of the gridded temperature reconstructions developed in this paper, using those grid boxes that lie in each region, based on Hugershoff-standardised tree density (thin dashed lines) and with additional low-frequency variations included (thick dashed lines). All series have been smoothed with a 30-year low-pass filter.

Figure 8: Longitude versus time variations in April-to-September temperature (K), averaged over the region north of 45°N and expressed as anomalies with respect to the 1961–1990 mean: (a) observed; and (b) reconstructed from the Hugershoff-standardised tree-ring density network, with coverage extended using principal component regression. Both panels were masked to limit their coverage to only those boxes and years where both data sets have values. The data were not smoothed.

Figure 9: As Figure 8, but smoothed in time with a 10-year filter and expressed as anomalies with respect to the 1801–1900 mean, and (a) reconstructed from the Hugershoff-standardised tree-ring density network, with coverage extended using principal component regression; and (b) as (a), except additional low-frequency variability has been added (see section 6).

(a)



(b)

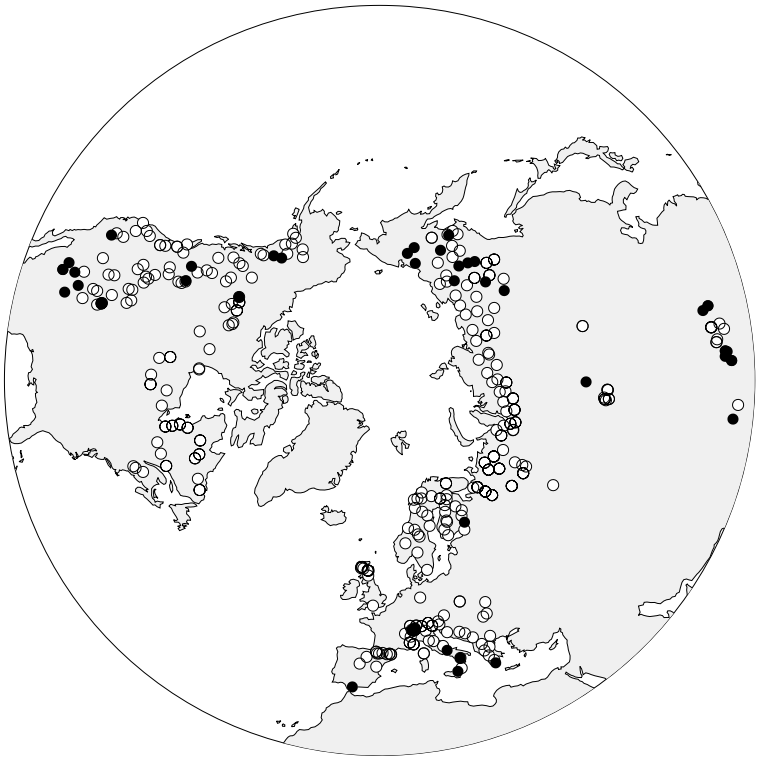


Figure 1: (a) Location (circles) of the 387 sites with tree-ring density chronologies, together with the boundaries of nine regions: NEUR = northern Europe, SEUR = southern Europe, NSIB = northern Siberia, ESIB = eastern Siberia, CAS = central Asia, TIBP = Tibetan Plateau, WNA = western North America, NWNA = northwestern North America, and ECCA = eastern and central Canada. (b) Locations of the 341 chronologies that were retained (open circles) and the 46 that were discarded (filled circles).

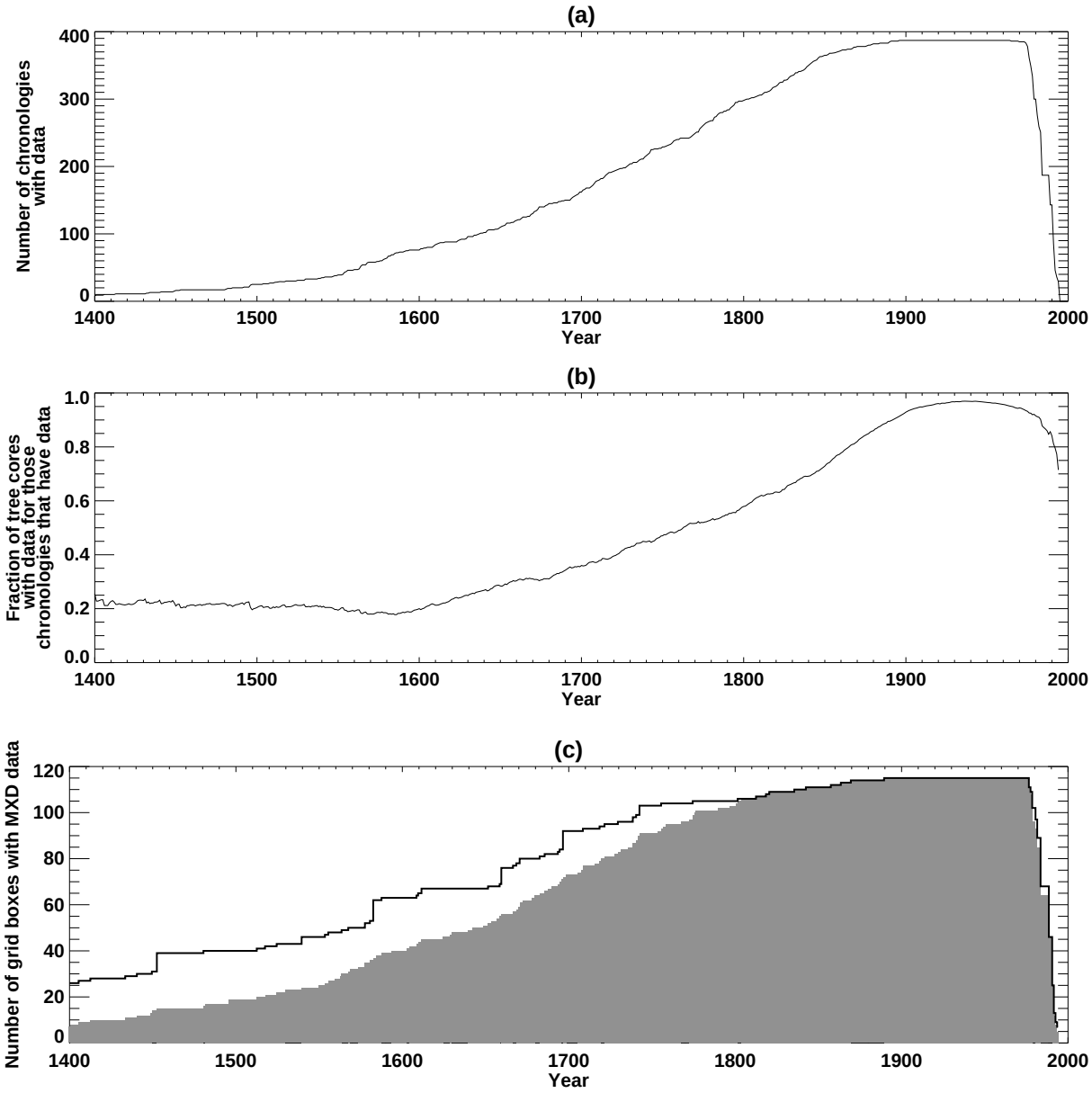


Figure 2: (a) Number of tree-ring density chronologies in the network that have values in each of the years from 1400 to 1994. (b) The fraction of tree cores within each chronology (*f*) that have measurements in each year, averaged over only those chronologies that have some data in that year (i.e., excluding those with $f = 0$). (c) Number of grid boxes with tree-ring density values in each of the years from 1400 to 1994; grey shading shows the number with real measurements, black line shows the number after extension of some series by regression against their neighbours.

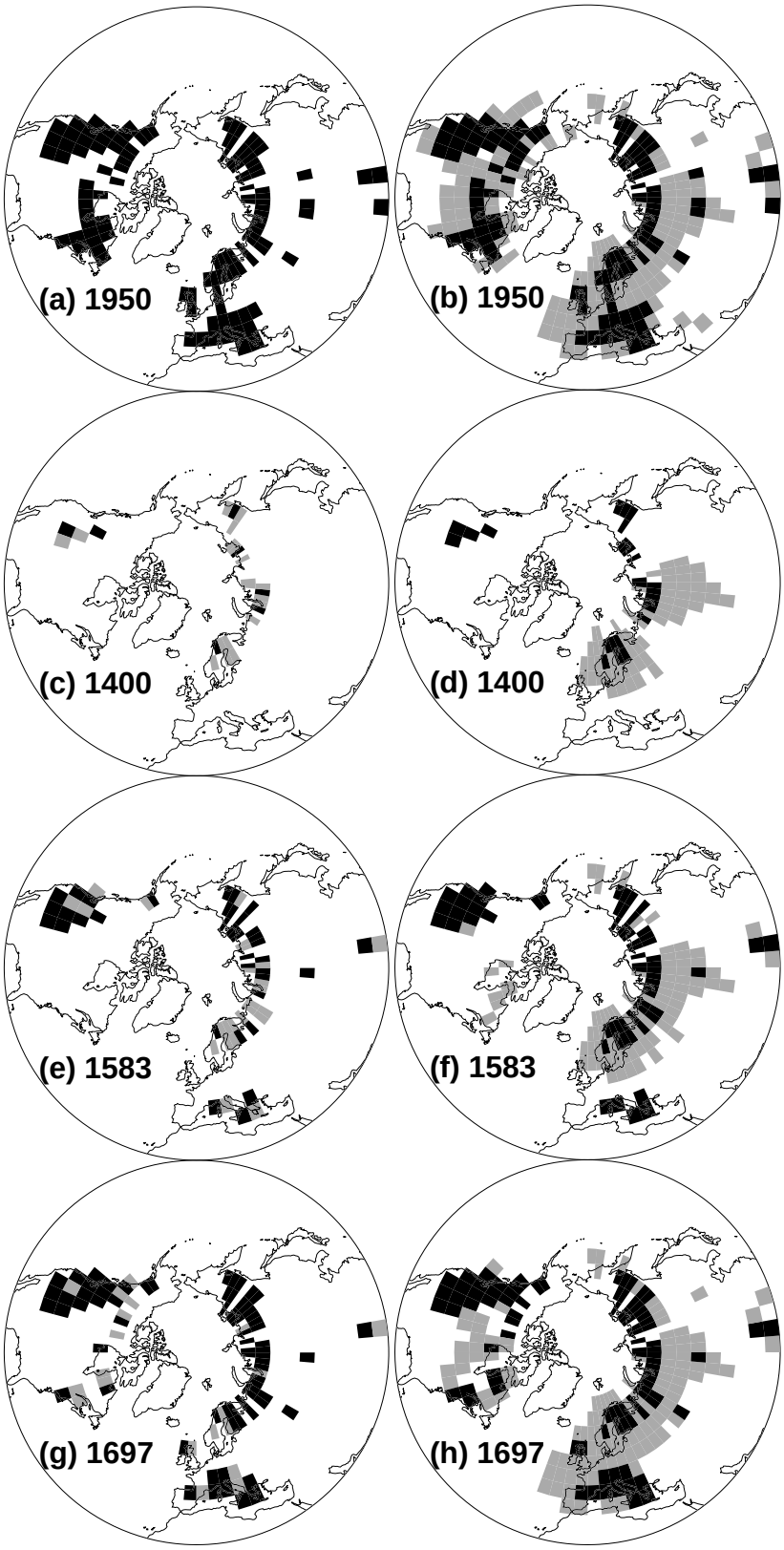


Figure 3: Spatial coverage of the gridded data set, with real density measurements (black boxes in a, c, e, and g), with density data extended by regression against neighbours (grey boxes in c, e and g), with temperature estimates obtained by local calibration of density data (black boxes in b, d, f and h) and temperature estimates obtained using principal component regression (grey boxes in b, d, f and h). Coverage is shown for four periods: (a, b) full coverage 1890–1976; (c, d) data back to 1400; (e, f) data back to 1583; and (g, h) data back to 1697.

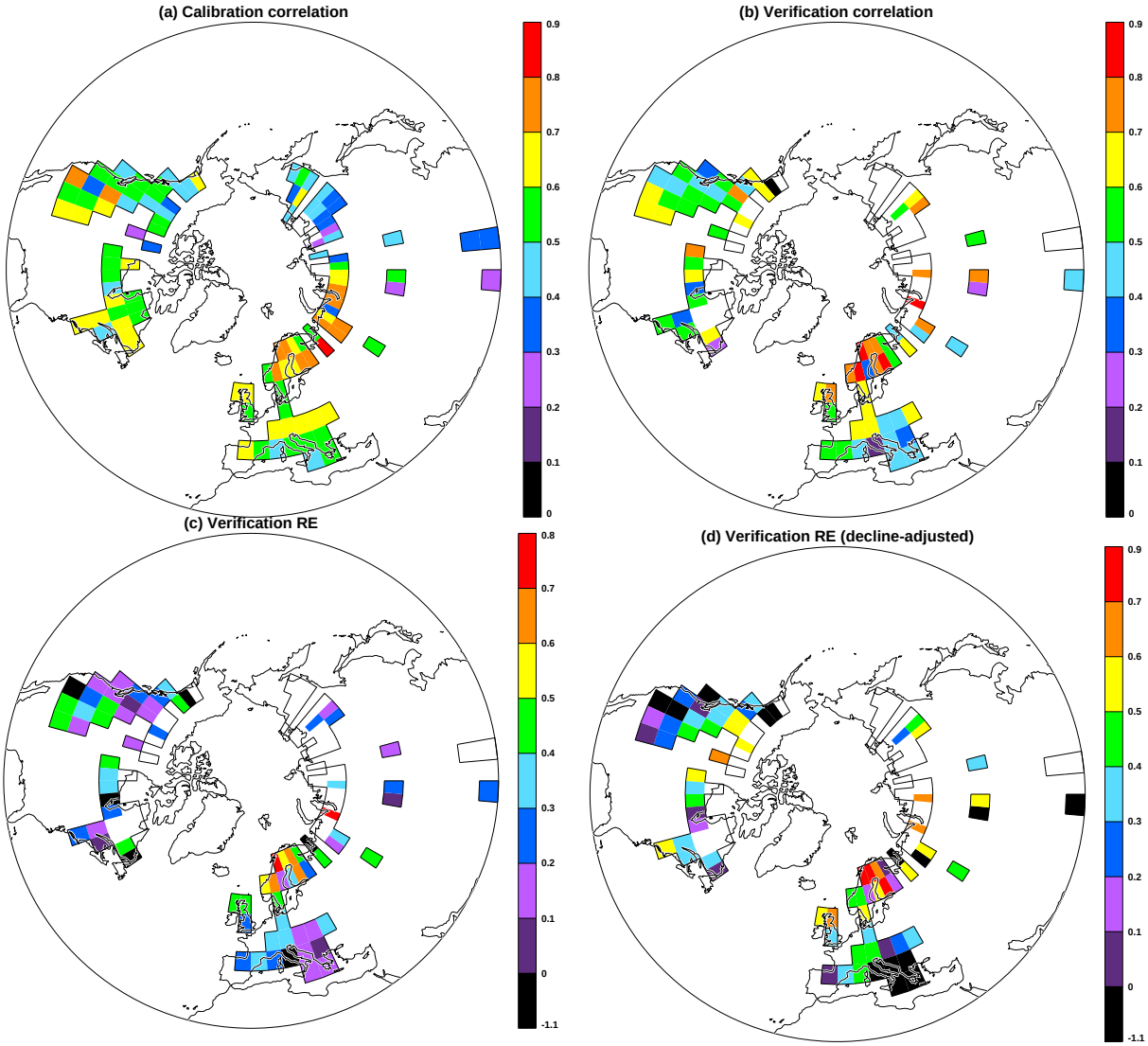


Figure 4: (a) Correlations between high-pass filtered density and temperature over the correlation period, 1911–1990. (b) Correlations between unfiltered density and temperature over the verification period, 1856–1910. (c) RE statistic computed over the 1856–1910 verification period for the reconstructions generated by linear regression between high-pass filtered density and temperature. (d) As (c), except for the reconstructions generated by matching the variance between the decline-adjusted density and temperature. The black outline marks the region with tree density data, and white boxes within this outline are those for which there was insufficient temperature data for either calibration or verification.

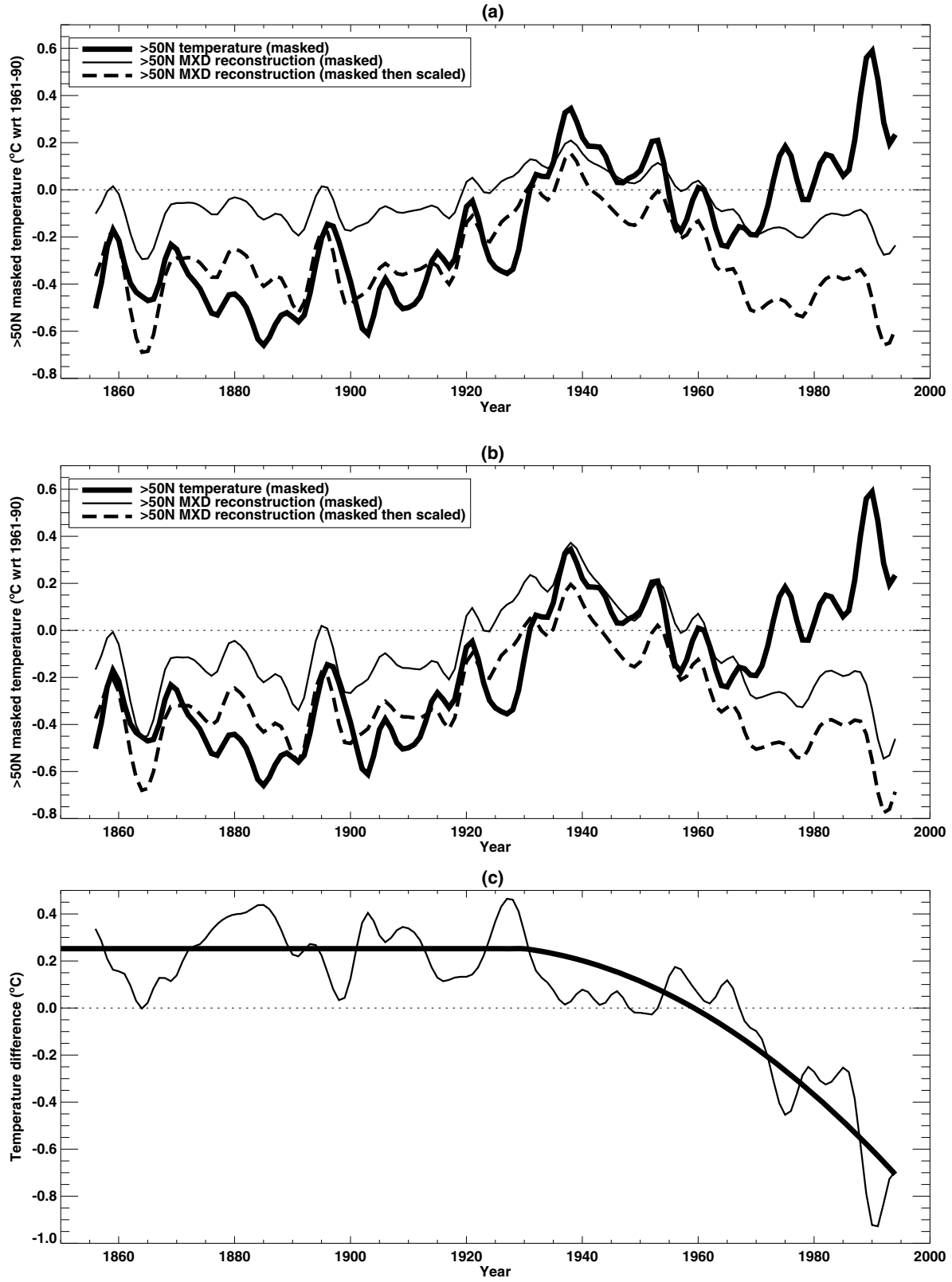


Figure 5: (a, b) Observed (thick line) and reconstructed (thin line) temperature averaged over all boxes north of 50°N where both observations and reconstructions have data, and the 1856–1960 least-squares fit (dashed line) of the thin curve on the thick curve. Reconstructions are generated by (a) linear regression, and (b) variance matching, between high-pass filtered density and temperature series. (c) The difference between the observed and reconstructed series in (b), together with a smooth function fitted to the difference series. All curves have been smoothed with a 10-year filter.

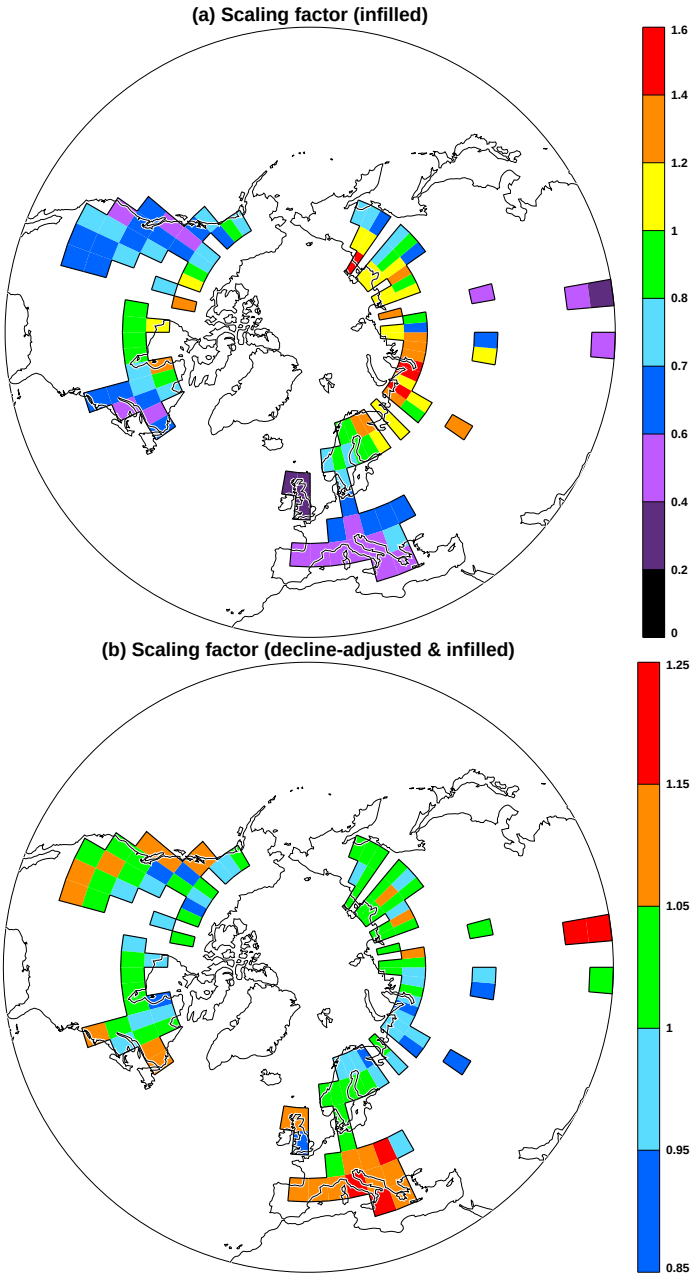


Figure 6: Scaling factors applied to the normalised tree density data to calibrate them into temperature reconstructions: (a) factors required to make the high-frequency variance match the observed temperature high-frequency variance; and (b) factors that are subsequently applied based on the variance of the unfiltered "decline-adjusted" calibrated density and the variance of the unfiltered observed temperature.

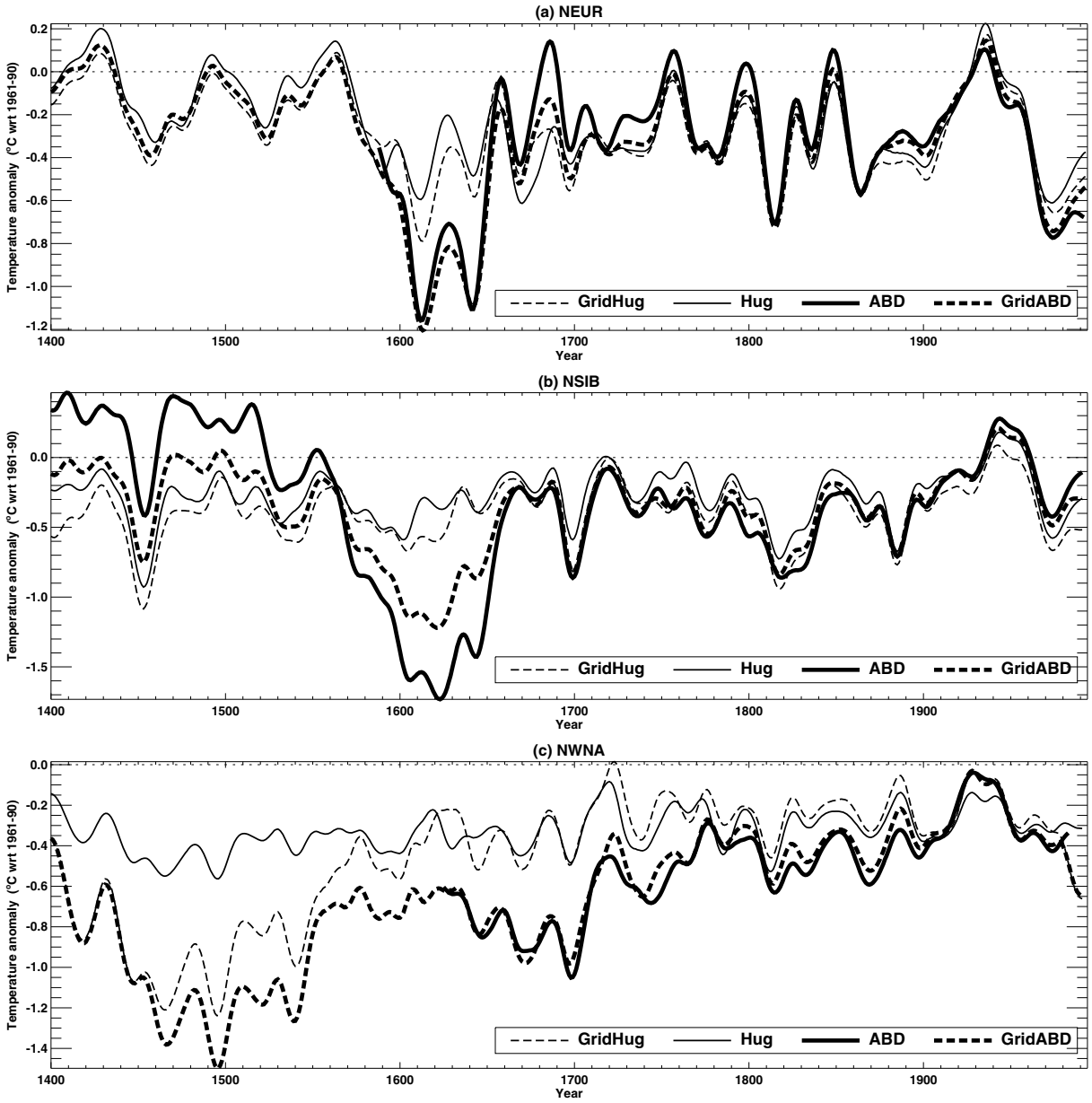


Figure 7: Regional-average temperature reconstructions for (a) northern Europe; (b) northern Siberia; and (c) northwestern North America. Continuous lines are the regional reconstructions published by Briffa *et al.* (2002a) using Hugershoff-standardised tree density (thin lines) and by Briffa *et al.* (2001) using ABD-processed tree density (thick lines). Dashed lines are the average of the gridded temperature reconstructions developed in this paper, using those grid boxes that lie in each region, based on Hugershoff-standardised tree density (thin dashed lines) and with additional low-frequency variations included (thick dashed lines). All series have been smoothed with a 30-year low-pass filter.

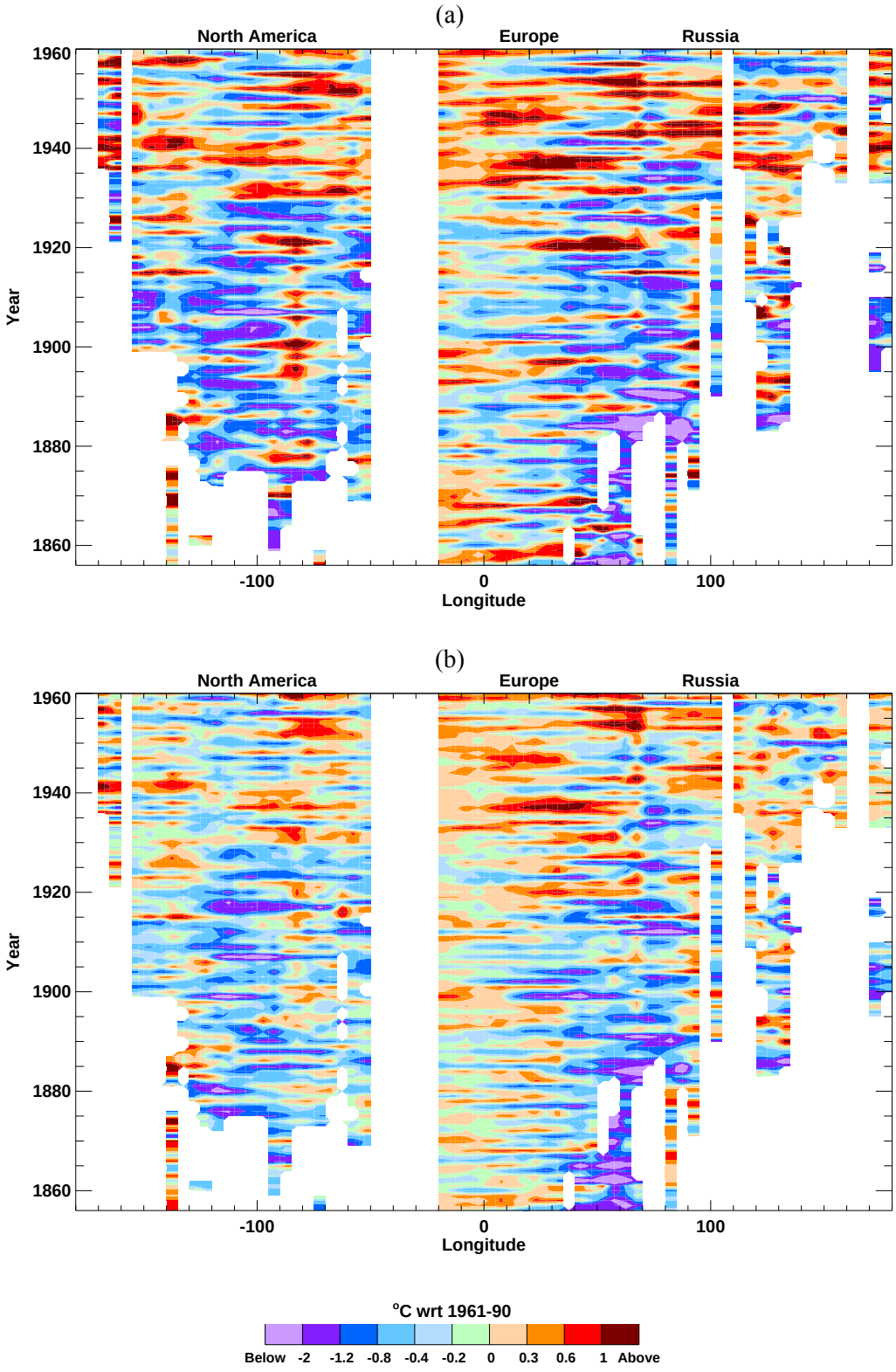


Figure 8: Longitude versus time variations in April-to-September temperature (K), averaged over the region north of 45°N and expressed as anomalies with respect to the 1961–1990 mean: (a) observed; and (b) reconstructed from the Hugershoff-standardised tree-ring density network, with coverage extended using principal component regression. Both panels were masked to limit their coverage to only those boxes and years where both data sets have values. The data were not smoothed.

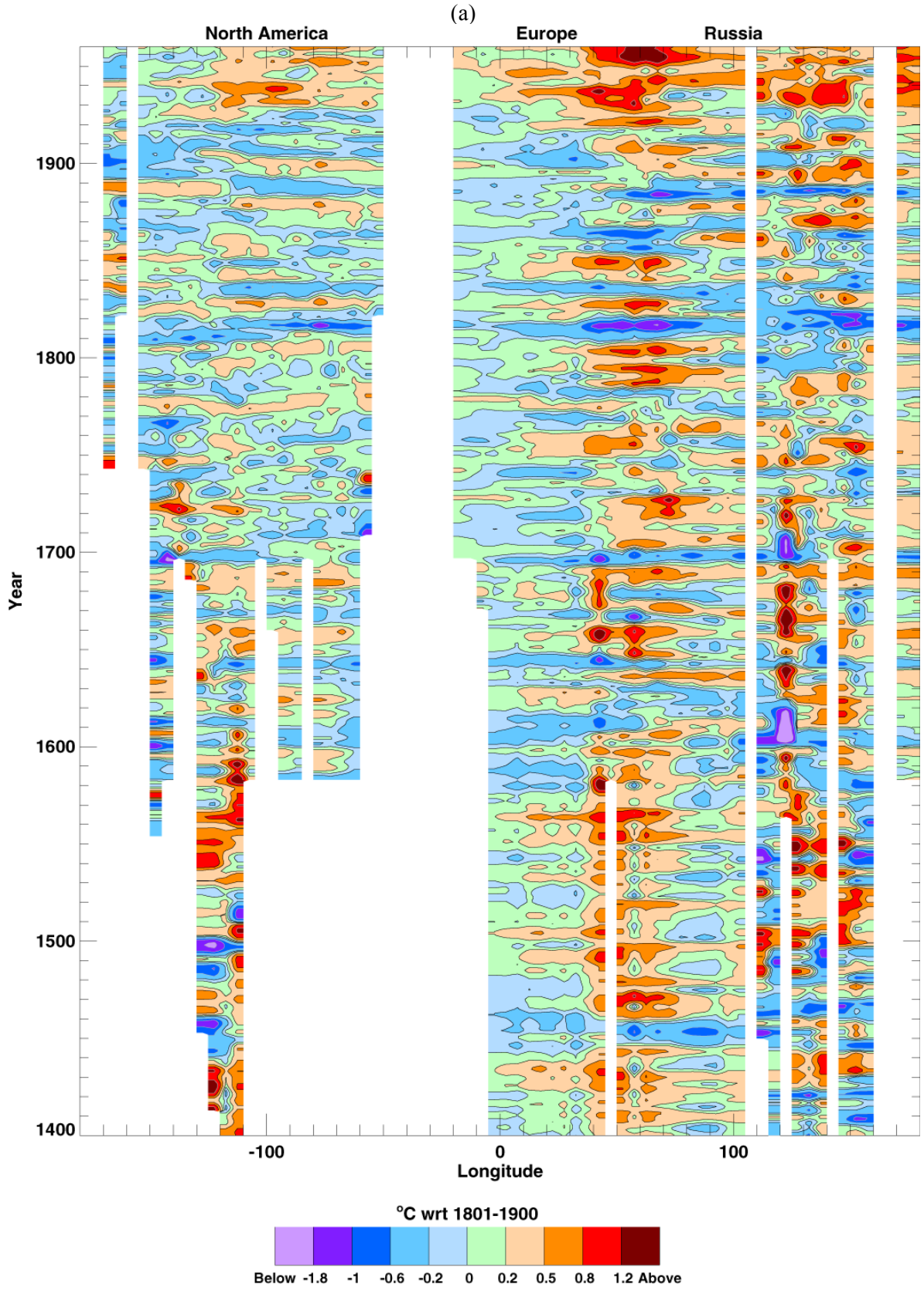


Figure 9: As Figure 8, but smoothed in time with a 10-year filter and expressed as anomalies with respect to the 1801–1900 mean, and (a) reconstructed from the Hugershoff-standardised tree-ring density network, with coverage extended using principal component regression; and (b) as (a), except additional low-frequency variability has been added (see section 6).

(b)

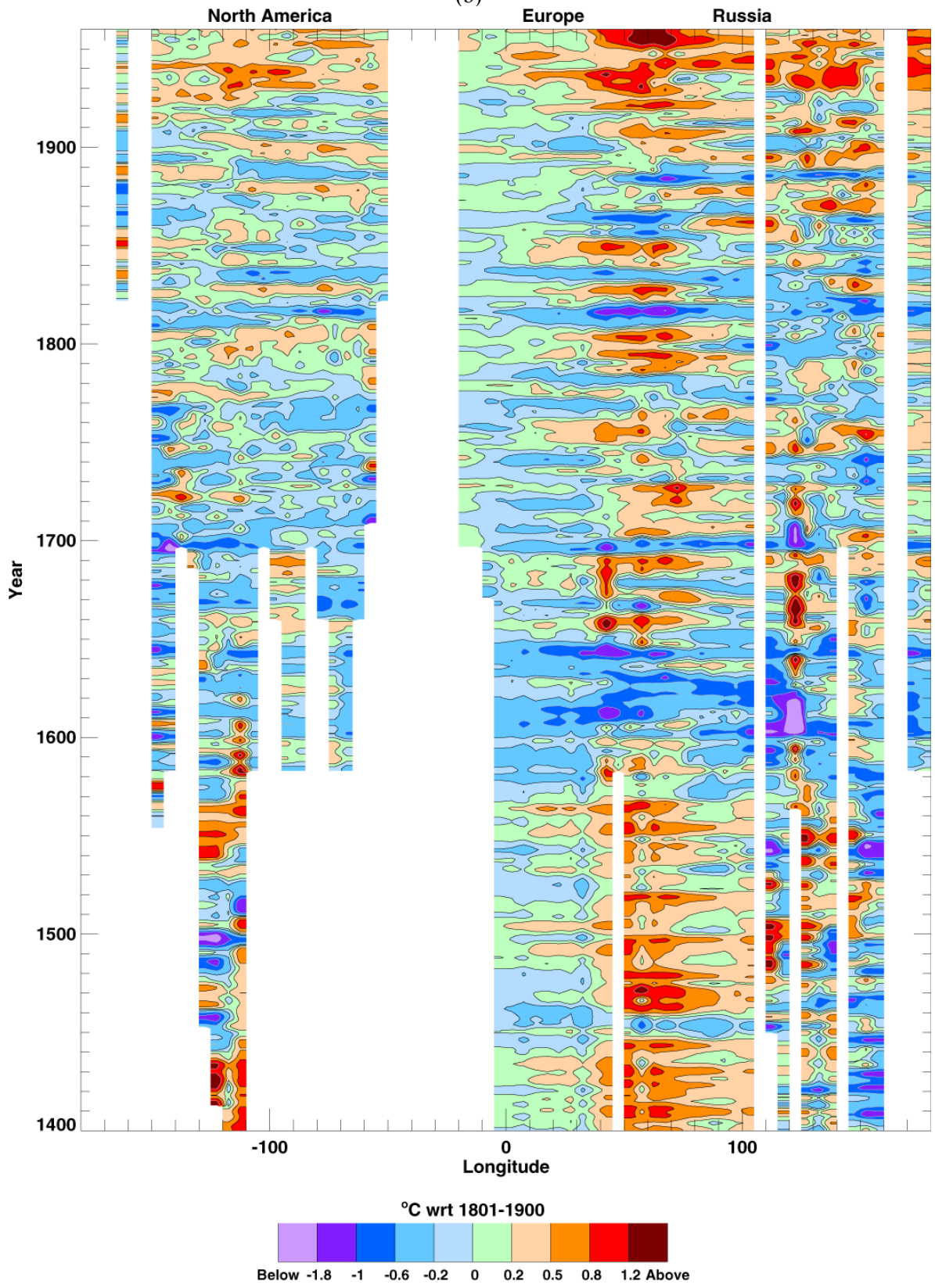


Figure 9: (continued)

University of East Anglia

Authors

Tim Osborn is a Reader in Climate Science in the Climatic Research Unit, School of Environmental Sciences, University of East Anglia.

Keith Briffa

Fritz Schweingruber

Phil Jones

Climatic Research Unit
School of Environmental Sciences
University of East Anglia