Effect of scaling and regression on reconstructed temperature amplitude for the past millennium

Jan Esper and David C. Frank
Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

Robert J. S. Wilson
School of GeoSciences, Grant Institute, Edinburgh University, Edinburgh, UK

Keith R. Briffa
Climatic Research Unit, University of East Anglia, Norwich, UK

[1] Examination of large-scale millennial-long temperature reconstructions reveals a wide range of datasets and methods used for calibration. Proxy time series are commonly calibrated against overlapping instrumental records, representing different seasons, Northern Hemisphere latitudinal bands, and including or excluding sea surface temperature data. Methodological differences include, using scaling or regression, the calibration time period, and smoothing data before calibration. We find that these various approaches alone can result in differences in the reconstructed temperature amplitude of about 0.5°C. This magnitude is equivalent to the mean annual temperature change for the Northern Hemisphere reported in the last IPCC report for the 1000–1998 period. A more precise assessment of absolute reconstructed temperature amplitudes is necessary to help quantify the relative influences of forcing mechanisms in climate models. **Citation:** Esper, J., D. C. Frank, R. J. S. Wilson, and K. R. Briffa (2005), Effect of scaling and regression on reconstructed temperature amplitude for the past millennium, *Geophys. Res. Lett.*, *32*, L07711, doi:10.1029/2004GL021236.

1. Introduction

[2] Four recently published large-scale temperature reconstructions with annual resolution have been used to represent temperature variations over the last 1000 years: *Briffa* [2000, hereinafter referred to as Briffa00], *Esper et al.* [2002, hereinafter referred to as Esper02], *Jones et al.* [1998, hereinafter referred to as Jones98], and *Mann et al.* [1999, hereinafter referred to as Mann99]. These records were developed using tree ring data alone or using multi-proxy data, and are reported to represent different regions (e.g. Northern Hemisphere (NH) extra-tropics, or full NH; Table 1).

[3] The records vary in how they portray the transition from the Medieval Warm Period into the Little Ice Age, and the temperature amplitude (here defined as the difference between the warmest and coldest decades in °C; Table 1, Figure A1[1]) reconstructed for the past millennium. The differences in amplitude result from varying low frequency characteristics inherent in the proxy records [*Esper et al.*,

2004], and also from the specific calibration approaches and instrumental 'targets'. The proxy records have been scaled or regressed, using different instrumental data, and different calibration periods both in their initial presentation and also in subsequent compilations (Table 1, and references therein). These varying approaches affect the assessment of the magnitude of recent climate change in a longer-term context.

[4] The objective of this paper is to survey the calibration methods and data used in recent literature, and demonstrate their influence on reconstructed temperature amplitudes. Specifically, we compare:

[5] (*i*) using either annual or warm season (A–S) temperature data (i.e. influence of season),

[6] (*ii*) spanning the full NH or the 20°–90°N latitudinal bands (influence of latitude),

[7] (*iii*) including or excluding sea surface temperature data (influence of sea), and

[8] (*iv*) using the 1856–1980 or 1900–77 periods for calibration (influence of period).

We discuss differences between scaling and regression, and detail the effect of smoothing the proxy and instrumental data prior to calibration. Reasons for greater or lesser sensitivities of the proxy data to certain instrumental targets are addressed.

2. Instrumental Temperature Data

[9] Large-scale temperature averages provided by the Climatic Research Unit were used [*Jones et al.*, 1999] to assess the different approaches listed in Table 1. The spatial coverage of these data – i.e. the percent of gridboxes having data – is, for NH land and sea surface temperatures, ~90% in the 1950s, ~50% in the 1900s, and <20% in the 1860s. Therefore, not only are the data more uncertain back in time, but the average also becomes biased towards Europe, North America, and areas in Asia. The sparser coverage also results in slightly increased variance before 1880 [*Jones et al.*, 1997; *Jones and Moberg*, 2003]. Similar considerations also apply to the proxy reconstructions themselves.

[10] Statistical characteristics of the instrumental targets, such as variance and trend behavior, are relevant for proxy record calibration, and thus to reconstructed temperature

---

[1]Auxiliary material is available at ftp://ftp.agu.org/apend/gl/2004GL021236.

Table 1. Large-Scale Temperature Reconstructions

Source	Data	Max. Response	Cold/Warm[a]	Calibration Instrumental Record	Period
		Original Papers			
Briffa00	tree rings	warm season T	1810s/1950s	–[b]	–
Mann99	multi-proxy	annual T	1460s/1940s	annual T, NH land + sea	1902–1980
Esper02	tree rings	warm season T	1290s/1950s	annual T, 30°–70°N land[c]	1856–1992
Jones98	multi-proxy	warm season T	1690s/1930s	warm T, NH land + sea	1961–1990
		Papers Providing Composite Figures of Several Reconstructions			
Briffa and Osborn [2002]				annual T, 20°–90°N land	1881–1960
Mann et al. [2003]				annual T, NH land + sea	1856–1980
Briffa et al. [2004]				warm T, 20°–90°N land	1881–1960
Jones and Mann [2004]				annual T, NH land + sea	Various

[a]Cold/warm indicates the coldest and warmest decade of each reconstruction.

[b]The Briffa00 record was not calibrated in the original paper, but in *Briffa and Osborn* [2002].

[c]In the original paper, the Esper02 record was scaled against the Mann99 record over the 1900–77 period. The instrumental data and period shown here are taken from *Cook et al.* [2004].

amplitudes. The lower frequency trends are quite similar between the instrumental temperature averages considered here (see Figure A2), with the largest difference occurring between summer and annual data. The early, 19th century, warm season data display relatively higher temperatures, with the 1860s decade being 0.34°C warmer than the 1900s decade. Standard deviations (STDV) of the large-scale instrumental averages are slightly lower for data averaged over larger areas, e.g. NH instead of 20°–90°N, and land and sea surface instead of land only. Warm season temperatures possess slightly lower STDV than annual temperatures. For the data compared here (Figure A2), STDV range between 0.19°C and 0.26°C over the 1856–1980 period.

3. Principles of Scaling and Regression

[11] Large-scale proxy time series are typically scaled to, or regressed against instrumental climate data. Here 'scaling' refers to the equalization of the mean and STDV of a proxy time series to the corresponding values of an instrumental temperature record over a defined period of overlap (i.e. the calibration period), e.g. 1856–1980 or 1900–1977.

[12] In cases where least squares linear regression is used for calibration [e.g., *Briffa and Osborn*, 2002], the resulting regressed temperature amplitude (T_R) equals the amplitude obtained from scaling (T_S) multiplied by the Pearson correlation coefficient between a proxy and instrumental record during the calibration period. Thus, the reconstructed temperature amplitude is reduced by the square root of unexplained variance in the regression. The rationale for performing regression instead of scaling is to allow a best (least-squares) estimate of temperatures during the calibration period. This reduces the reconstructed variability when a weaker agreement between proxy and instrumental record exists by not reconstructing error variance as determined by regression. Scaling, in contrast, results in homogeneous variability between the instrumental target and proxy, at the expense of inflated error variance.

[13] For example, when the Jones98 record is scaled to annual 20°–90°N land and sea surface temperatures (period 1856–1980), $T_S = 0.62$°C from the coldest to the warmest decade (for Jones98, the 1690s and 1930s respectively). The correlation between Jones98 and instrumental data over the same period is 0.55, resulting in $T_R = 0.34$°C.

[14] The $\sigma_{pre}/\sigma_{ins}$ quotient of a proxy time series, where σ_{pre} is the STDV of the pre-instrumental period (here, 1000–1855) and σ_{ins} the STDV of the instrumental period (here, 1856–1980), can be used to quantify amplitude changes over the past millennium, triggered by scaling (Figure A1). $\sigma_{pre}/\sigma_{ins}$ is, for example, low in Mann99 (0.66, with $\sigma_{pre} = 0.12$ and $\sigma_{ins} = 0.18$) and high in Esper02 (1.13, with $\sigma_{pre} = 0.12$ and $\sigma_{ins} = 0.10$). Consequently, reconstructed temperature variations using the Esper02 record are systematically larger than those from Mann99, when scaling is applied.

4. Scaling the Proxy Time series
4.1. Influence of Latitude, Sea, Season, and Period

[15] Temperature amplitudes (measured from the coldest to warmest decades; Table 1), resulting from scaling to the eight different instrumental targets over the 1856–1980 period, are 0.63–0.89°C for Briffa00, 0.51–0.73°C for Mann99, 0.93–1.31°C for Esper02, and 0.52–0.74°C for Jones98 (Figure 1a). Minimum amplitudes are obtained from scaling to warm season temperatures averaged over NH land and sea surface areas, and maximum amplitudes from scaling to annual temperatures averaged over 20°–90°N land only areas. Highest amplitudes are generally obtained from Esper02, and lowest amplitudes from Mann99 and Jones98. These tendencies are related to the internal structure of the reconstructions, as partly quantified by the $\sigma_{pre}/\sigma_{ins}$ quotient.

[16] The amplitude differences (in °C) shown in Figure 1a are compared for the instrumental categories (*i–iii*), with the maximum differences expressed in % for each proxy time series (Figure 1b). Accordingly, the influence of sea is largest at 24%, followed by the influence of season (15%) and latitude (14%). Interestingly, when scaling to warm season instead of annual mean temperatures over the period 1856–1980, reconstructed amplitudes are reduced, primarily due to the relatively greater temperatures of the early warm season data.

[17] If, however, the 1900–1977 period is used for scaling (instead of 1856–1980 as done in Figure 1a), amplitude increases, with a maximum of 49% (0.35°C) for Briffa00 are observed. This sensitivity to the time period depends upon the individual fit between a reconstruction and the target, and hence results in great differences from one reconstruction to another (Figure 1b).
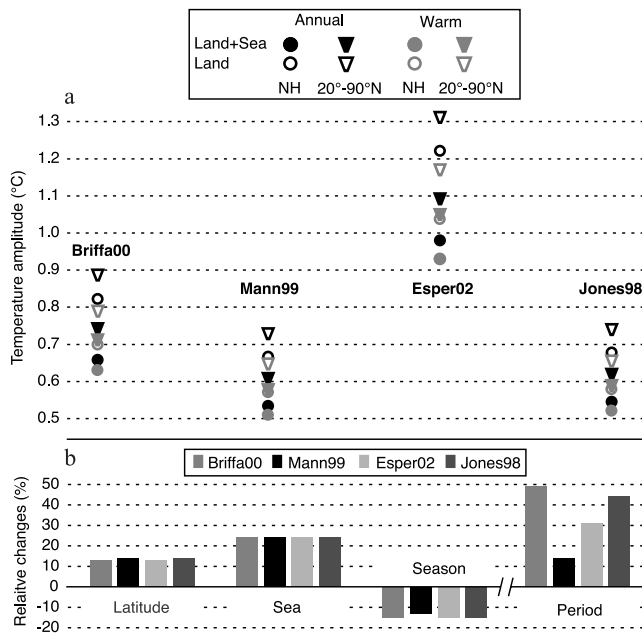
Figure 1. Temperature ranges of the Briffa00, Mann99, Esper02, and Jones98 reconstructions. (a) Amplitudes from coldest to warmest decades (see Table 1) after scaling the reconstructions to eight different temperature records (labeled on top) using the 1856–1980 period. (b) Maximum amplitude changes caused using observational data averaged over 20°–90°N (relative to full NH; influence of latitude), land only data (relative to land and sea surface; influence of sea), warm season temperatures (relative to annual; influence of season), and the 1900–1977 period (relative to 1856–1980; influence of period).

4.2. Scaling Examples

[18] To further illustrate scaling, comparisons between proxy and instrumental data are shown (Figure 2), detailing the under-prediction of 19th century instrumental data, when scaling over the 1900–1977 instead of the 1856–1980 period. This observation holds for Briffa00, Esper02, and Jones98 when using annual (Figure 2c), and for all reconstructions when using warm season temperature data (Figure 2d).

[19] The negative trend recorded in the Apr.–Sep. temperature data in the second half of the 19th century is generally not apparent in the proxy time series (Figures 2b and 2d). Mann99 shows a slight negative, and Briffa00, Esper02 and Jones98 a positive trend for this period. Note also that only annual temperatures show a significant increase from the 1860s to the 1970s, a feature seen in all reconstructions compared here.

[20] The poor fit with warm season temperatures also reduces the correlation results obtained when using original and low-pass filtered proxy and instrumental data (Figure A3). The high-pass components, however, exhibit slightly higher correlations with warm season temperatures for Briffa00, Esper02, and Jones98. Correlation results also indicate higher agreement between all proxy records and temperatures averaged over the full NH, instead of the 20°–90°N latitudinal band.

5. Effect of Regression

[21] When calibrating using linear regression, the variance of a proxy time series is always less than that of the instrumental data (Figure 3). Resulting amplitude reductions – solely dependent on the correlation between proxy and instrumental data – are generally larger for warm season than for annual, for land only than land and sea surface, and for 20°–90°N than NH data. Using the 1856–1980 period, the reductions are 42–67% (Briffa00), 17–51% (Mann99), 40–63% (Esper02), and 43–58% (Jones98) of the amplitudes obtained from scaling.

[22] As most reconstructions possess lower skill in capturing inter-annual variability, relative increases in amplitude are seen, if the instrumental and proxy data are low-pass filtered before regression [Cook et al., 2004]. For Briffa00, Esper02, and Jones98 correlations between 20-year high- and low-passed proxy and instrumental series are <0.3 and >0.8, respectively (Figure A3). While greater spatial coverage of the proxy data would be necessary to better capture higher frequency variations, smoothing the data results in a substantial loss of degrees of freedom.

6. Discussion

[23] The various scaling and regression approaches applied in recent literature to proxy-based temperature
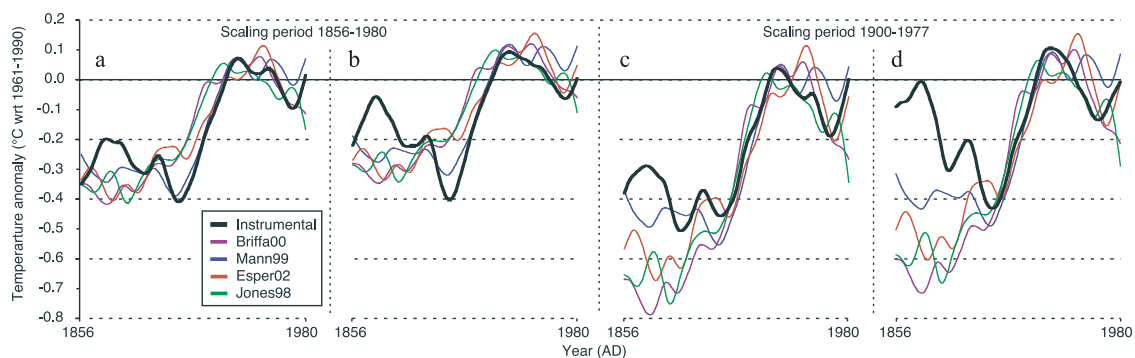

Figure 2. Scaling examples using the (a) and (b) 1856–1980 and (c) and (d) 1900–1977 periods. Scaling was performed using the original (un-filtered) series. Results smoothed with a 20-year filter. (a) Annual NH land and sea surface temperatures, (b) warm season (Apr.–Sep.) NH land and sea surface temperatures, (c) annual 20°–90°N land temperatures, and (d) warm season 20°–90°N land temperatures.
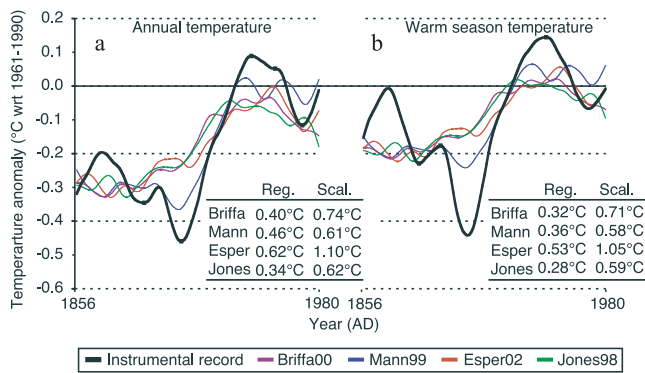
Figure 3. Regression examples against (a) annual and (b) Apr.–Sep. temperature records. The instrumental data were averaged over the 20°–90°N latitudinal band, and the calibration period is 1856–1980. The regression was performed using the original (un-filtered) series. Results smoothed with a 20-year spline filter. Inset tables show the amplitudes after regressing the reconstructions (Reg.) and, for comparison, after scaling them (Scal.).

records significantly affect the absolute temperature amplitude reconstructed for the past millennium. The range of these changes from scaling and regressing to different 'reasonable' instrumental targets easily approaches 0.5°C for decadal means. This range is on the order of the full temperature amplitude displayed in the last IPCC report for the past millennium. Consideration of temporally changing spatial coverage and uncertainty in both the instrumental and proxy data, as expressed by confidence limits accompanying such records, would further increase the range of amplitude estimates over the past millennium.

[24] When linear regression is used for calibration, the variance of a proxy record remains below that of the target data, leaving the visual impression that the recent dynamics are substantially larger than the historic ones when splicing such records together. Comparisons presented herein show that temperature amplitudes obtained from scaling vs. regression can differ by about 50% (wrt. scaling amplitude), as a result of the (predominantly) higher frequency error eliminated in regression. Consideration of the common signal between proxy and target as a function of frequency (e.g. via smoothing) can result in more similar amplitudes between proxy and target data, but is only practical when the data are long enough to prevent a dangerous reduction in the effective degrees of freedom associated with the calibration.

[25] Aside from the above considerations, (i) the choice of the calibration period, and (ii) the use of warm season instead of annual temperatures (because of significant miss-fit in the 19th century) are critical to reconstructed amplitudes. Less critical is the choice between land only or land and sea surface data, and, in particular, between data averaged over the full NH or 20°–90°N, as these different spatial averages are quite similar.

[26] For the selection of land only or land and sea surface data, the location and spatial response pattern of the regional proxy records should be considered. In cases where land-based proxy data also respond to regional sea surface temperatures (through climate teleconnections), the appli-

cation of both land and sea surface data seems reasonable. However, differences in the spectral character of the proxy and the temperature target might lead to underestimates of long-term temperature variability [*Osborn and Briffa*, 2004].

[27] The sensitivity of amplitude to changes in calibration period is strongly influenced by time dependent changes in the fit, and the pattern of trends between the instrumental and proxy series. The original temperature reconstructions reviewed used a wide range (e.g. 1961–1990 to 1856–1980) of calibration periods, which have influenced their long-term amplitudes. Assuming small and nonsystematic biases in the variance and spatial representativity of early instrumental data [*Jones and Moberg*, 2003], it would seem reasonable to calibrate over as long a period as possible to minimize the sensitivity to shorter-term changes in proxy and instrumental series relationships.

[28] The application of a long calibration period is complicated, however, by the substantially warmer Apr.–Sep. temperatures (relative to annual) during the middle of the 19th century. This warmth is broadly absent in the proxy records, even though three of them (Briffa00, Esper02, Jones8) are described as having an optimal response to summer temperatures. Possible hypotheses to help resolve this dilemma include (i) early warm season records indicate higher temperatures than actually occurred [*Chenoweth*, 1993; *Jones and Moberg*, 2003; *Moberg et al.*, 2003], (ii) the proxy records also contain substantial signals outside those of the warm season [*Jacoby et al.*, 1996]. Our tests have shown that the differences between annual and warm season temperatures and their varying fits with the proxy reconstructions are critical for the estimation of temperatures over the past millennium, as reflected by the high sensitivity to the calibration time period. Resolution of this dilemma should lead to less uncertainty in estimated temperature amplitudes.

References

Briffa, K. R. (2000), Annual climate variability in the Holocene: Interpreting the message of ancient trees, *Quat. Sci. Rev.*, *19*, 87–105.

Briffa, K. R., and T. J. Osborn (2002), Blowing hot and cold, *Science*, *295*, 2227–2228.

Briffa, K. R., T. J. Osborn, and F. H. Schweingruber (2004), Large-scale temperature inferences from tree rings: A review, *Global Planet. Change*, *40*, 11–26.

Chenoweth, M. (1993), Nonstandard thermometer exposures at U.S. cooperative weather stations during the late nineteenth century, *J. Clim.*, *6*, 1787–1797.

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

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

Esper, J., D. C. Frank, and R. J. S. Wilson (2004), Low frequency ambition, high frequency ratification, *Eos Trans. AGU*, *85*, 113, 120.

Jacoby, G., R. D'Arrigo, and B. Luckman (1996), Millennial and near-millennial scale dendroclimatic studies in northern North America, in *Climatic Variations and Forcing Mechanisms of the Last 2000 Year*, *NATO ASI Ser., Ser. I*, vol. 41, edited by P. D. Jones et al., pp. 67–84, Springer, New York.

Jones, P. D., and M. E. Mann (2004), Climate over past millennia, *Rev. Geophys.*, *42*, RG2002, doi:10.1029/2003RG000143.

Jones, P. D., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations, *J. Clim.*, *16*, 206–223.

Jones, P. D., T. J. Osborn, and K. R. Briffa (1997), Estimating sampling errors in large-scale temperature averages, *J. Clim.*, *10*, 2548–2568.

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

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its variations over the last 150 years, *Rev. Geophys.*, *37*, 173–199.

Mann, M. E., R. S. Bradley, and M. K. Hughes (1999), Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.*, *26*, 759–762.

Mann, M. E., et al. (2003), On past temperatures and anomalous late 20th century warmth, *Eos Trans. AGU*, *84*, 256–258.

Moberg, A., H. Alexandersson, H. Bergström, and P. D. Jones (2003), Were southern Swedish summer temperatures before 1860 as warm as measured?, *Int. J. Climatol.*, *23*, 1495–1521.

Osborn, T. J., and K. R. Briffa (2004), The real color of climate change?, *Science*, *306*, 621–622.

————————————

K. R. Briffa, Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK. (k.briffa@uea.ac.uk)

J. Esper and D. C. Frank, Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland. (esper@wsl.ch; frank@wsl.ch)

R. J. S. Wilson, School of GeoSciences, Grant Institute, Edinburgh University, Edinburgh EH9 3JW, UK. (rjwilson_dendro@blueyonder.co.uk)
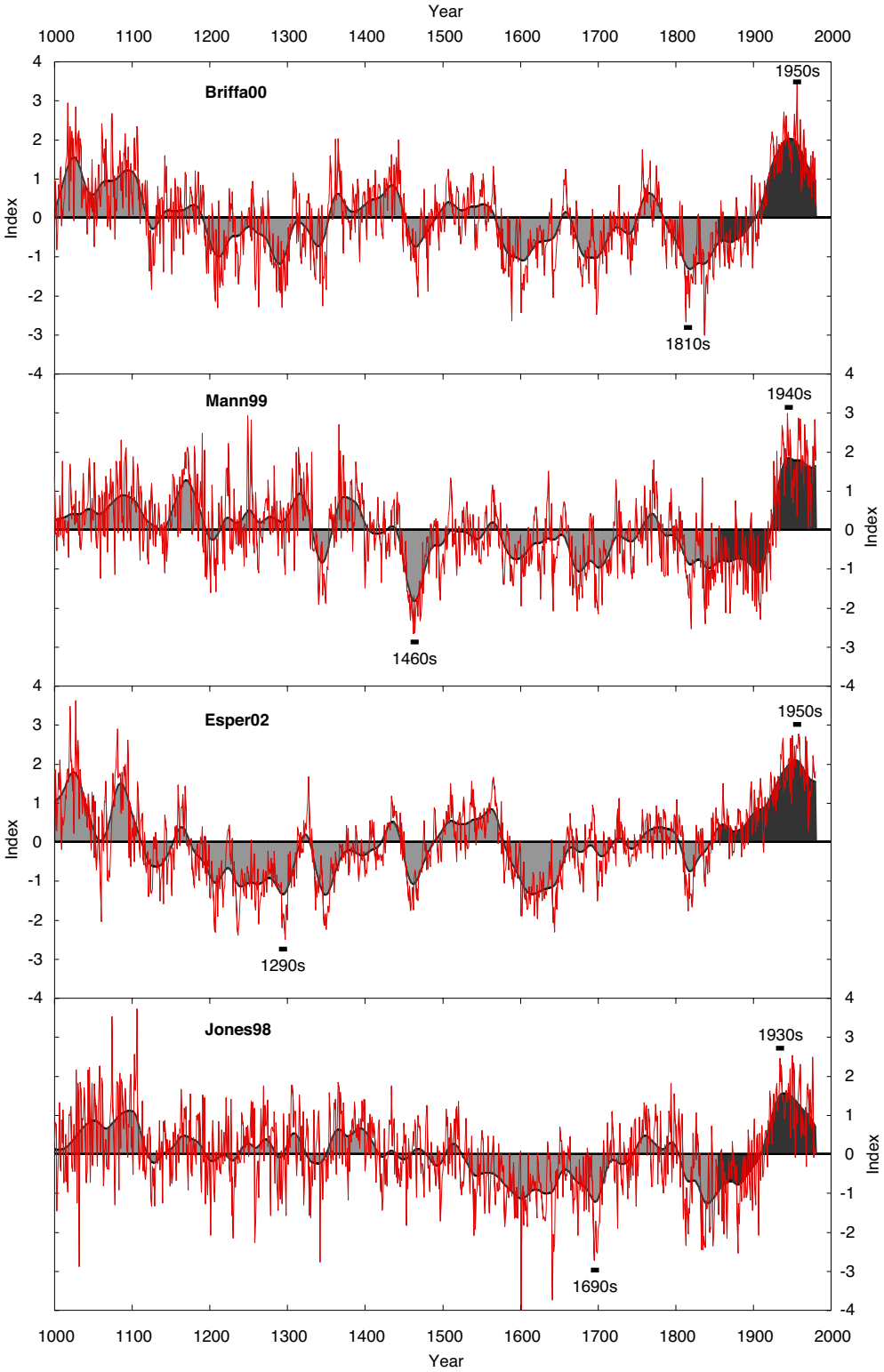
Figure S1 Millennial-long, large-scale temperature reconstructions with annual resolution. Records are normalized over the 1000-1980 period. The long-term trend is highlighted using a 40-year spline filter. The "pre-instrumental" period 1000-1855 is highlighted in grey, the "instrumental period" 1856-1980" is highlighted in black. Coldest and warmest decades of each reconstruction are indicated with bars.
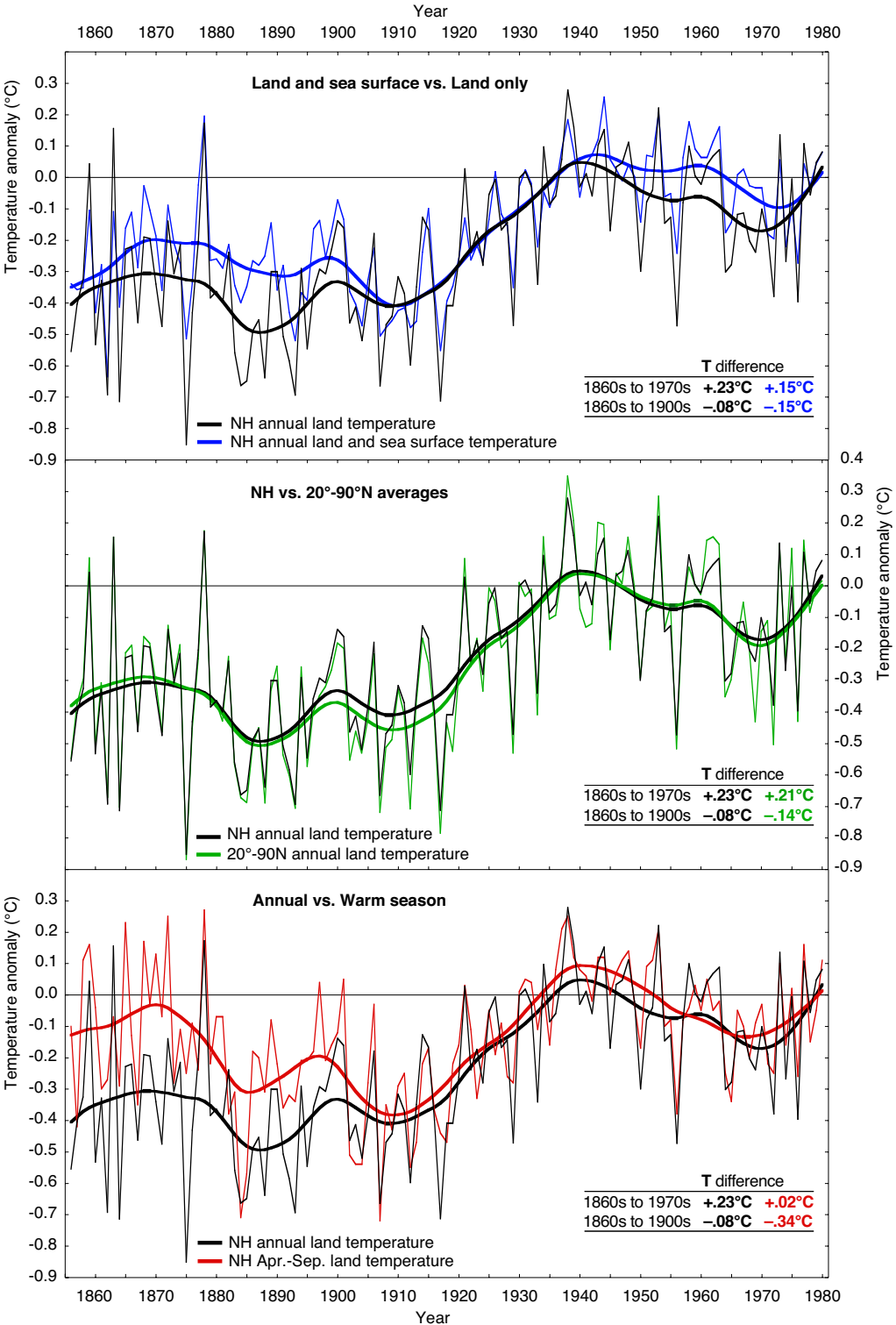
Figure S2 Large-scale mean temperature records. **(A)** Land and sea surface data versus land only data (both NH and annual). **(B)** NH data versus 20°-90°N data (both annual and land only). **(C)** Annual data versus Apr.-Sep. data (both NH and land only). Smoothed series are 20-year splines. Tables indicate the temperature differences (°C) between decadal means from the 1860s to 1970s, and 1860s to 1900s. Data provided by the Climatic Research Unit, UK.

Figure S3 Correlations between reconstructions and temperature records over the 1856-1980 period, using original, low-pass and high-pass filtered data. Low- and high-pass filters applied to reconstructions and observational records are 20-year splines. Results from correlations with NH average data are shown in blue, and with 20°-90°N average data in red. An = annual temperature, Wa = warm season (Apr.-Sep.) temperature, LS = land and sea surface temperature, L = land (only) temperature.