Geophysical Research Letters

Supporting Information for

Revising mid-latitude summer-temperatures back to AD 600 based on a wood density network

Lea Schneider[1], Jason E. Smerdon[2], Ulf Büntgen[3], Rob J. S. Wilson[4], Vladimir S. Myglan[5], Alexander V. Kirdyanov[6] and Jan Esper[1]

[1]Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany

[2]Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA

[3]Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland

[4]School of Geography and Geosciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK

[5]Institute for the Humanities, Siberian Federal University, Krasnoyarsk, 660041, Russia

[6]V.N. Sukachev Institute of Forest SB RAS, Akademgorodok, Krasnoyarsk, 660036, Russia

Contents of this file

Text S1
Figures S1 to S11
Tables S1 to S2

Text S1. Supplementary Methods

Proxy Selection and Chronology Building For this study we compiled the most complete set of MXD-chronologies longer than 600 years. All datasets consist of relict and recent wood samples providing a relatively homogeneous age structure throughout the millennium. While this is a preferable feature for a RCS-detrending, not all sites revealed sufficient homogeneity to perform RCS with only one Regional Curve. Distinct inconsistencies regarding the tree species, the exact sampling site or the measurement technics made a split RCS approach mandatory for Athabasca, Jämtland and Mangazeja. This approach comes at the expense of low-frequency variability and was only applied when the smaller data portion significantly increased chronology length or sample replication. For Jämtland, neither was the case and hence, the update (Gunnarson et al. 2011) could not be integrated.
An advantage of MXD is the negligible level-to-spread relationship in contrast to ring width (Cook and Peters 1997). Thus a power transformation is not essential, but was employed in a few cases (10-20%) for individual measurements.

Benchmarking and Pseudoproxy Experiments The reconstruction performance was benchmarked using red noise time-series of adjusted first order autocorrelation as an alternative proxy network. Based on 100 different noise realizations, each nest of the true reconstruction exceeded the 99% quantile of the red-noise benchmarking experiments using the R^2 and RE metrics ($R^2_{min/max}$= .08/.10 and $RE_{min/max}$ = -.37/-.24).
Pseudoproxy experiments (see *Smerdon* 2012 for a review) were performed using the surface temperature field from the CCSM4 last-millennium simulation completed for the PMIP3 project [*Landrum et al.*, 2013; *Taylor et al.*, 2012]. The pseudoproxy network was derived from a 5x5 interpolated version of the surface temperature field and was sampled based on the location of the original site records. Signal-to-noise ratios of each pseudoproxy chronology reflected the correlation between the actual proxy record and their corresponding locally observed temperatures. Land-only NH extratropical JJA temperatures served as the reconstruction target in the model, with all grid cells from 30°-90°N being weighted for area.

Frequency Spectra Frequency spectra were calculated for 5 PMIP3 last millennium simulations (BCC, CCSM, GISS, IPSL, MPI) [*Taylor et al.*, 2012] and six annually resolved NH temperature reconstructions [*Masson-Delmotte et al.*, 2013] (both linearly detrended) over their common period 1402-1960. The extraction of model data followed the procedure detailed above for the CRUTEM4v temperature data. All records were scaled with respect to 1923-1960 representing the most recent common period in the length of a calibration interval. In the multi-taper spectra [*Thomson*, 1982] frequencies below 0.0025 and above 0.5 cycles per year were eliminated to allow for spectral characterizations from 2-300 years. The remaining part was regularly binned and spectra expressed as average bin values to guarantee equal weighting of all frequencies. β-values in Fig. 4b were derived from the power law [*Franke et al.*, 2013] $P(f) \propto f^{-\beta}$, where P is the spectral energy and f the frequency. -β equals the linear slope of spectral energy plotted against frequency on a logarithmic scale.

Site name	*Country*	*Location*	*Species*	*Period*	*n*	*r*	*Reference*
Alaska	US	68.8N/142.4W	Picea glauca	1073-2002	246	0.31**	Anchukaitis et al. 2013
Altai	Russia	50.0N/88.0E	Larix sibirica	462-2007	59	0.60**	Myglan et al. 2012
Athabasca	Canada	52.3N/117.3W	Picea engelmannii	1072-1994	102	0.56**	Luckman et al. 2005
Camphill	Canada	68.3N/133.3W	Picea glauca	1175-1992	58	0.47**	Schweingruber, ITRDB
Jämtland	Sweden	63.5N/15.5E	Pinus silvestris, Picea abies	1107-1978	158	0.63**	Schweingruber et al. 1988
Lauenen	Switzerland	46.4N/7.3E	Picea abies	982-1976	206	0.33**	Schweingruber et al. 1988
Lötschental	Switzerland	47.5N/7.5E	Larix decidua	735-2004	180	0.62**	Büntgen et al. 2006
Mangazeja	Russia	66.7N/82.3E	Larix sibirica, Picea obovata	1246-1990	143	0.70**	Schweingruber, ITRDB
Nscan	Finland	67.5N/22.5E	Pinus silvestris	-181-2006	587	0.82**	Esper et al. 2012
Polarural	Russia	66.9N/65.6E	Larix sibirica, Picea obovata	778-2006	141	0.76**	Briffa et al. 2013
Pyrenees	Spain	42.5N/2.5E	Pinus uncinata	1044-2005	203	0.40**	Büntgen et al. 2008
Quebec	Canada	57.5N/76.0W	Picea mariana	1352-1989	83	0.59**	Schweingruber, ITRDB
Torneträsk	Sweden	68.2N/19.7E	Pinus silvestris	441-2010	124	0.79**	Melvin et al. 2013
Tyrol	Austria	47.5N/12.5E	Picea abies	1047-2003	227	0.42**	Esper et al. 2007
Zhaschiviersk	Russia	67.5N/142.6E	Larix sibirica	1311-1991	55	0.30*	Schweingruber, ITRDB

Table S1. Site information of the 15 MXD-datasets. 'Period' covered by MXD samples is given in years AD, 'n' denotes the number of samples and 'r' the Pearson correlation coefficient against a nearby CRUTEM4v grid cell over the 1901-1976 period. ** indicates 95% significance, * 90% significance.

	RCS 1	RCS 2	RCS SF
Detrending	RCS	RCS	RCS + Signal Free
RC-smoothing	100 yrs spline	Age-dependent spline	Age-dependent spline
Variance Stabilization	100 yrs spline through absolute departure	Replication-dependent (51 yrs running window)	Replication-dependent (51 yrs running window)
Software	ARSTAN (Cook and Krusic)	RCSigfree (Cook, Krusic and Melvin)	RCSigfree (Cook, Krusic and Melvin)

Table S2. Detrending approaches. RCS1 is the detrending setup finally chosen. RCS2 and RCS SF are alternative detrendings assessed in this study. Besides the basic detrending choice (1st row), differing Regional Curve smoothings (2nd row), and variance-stabilization techniques (3rd row) were tested. Additionally, two software packages were used.

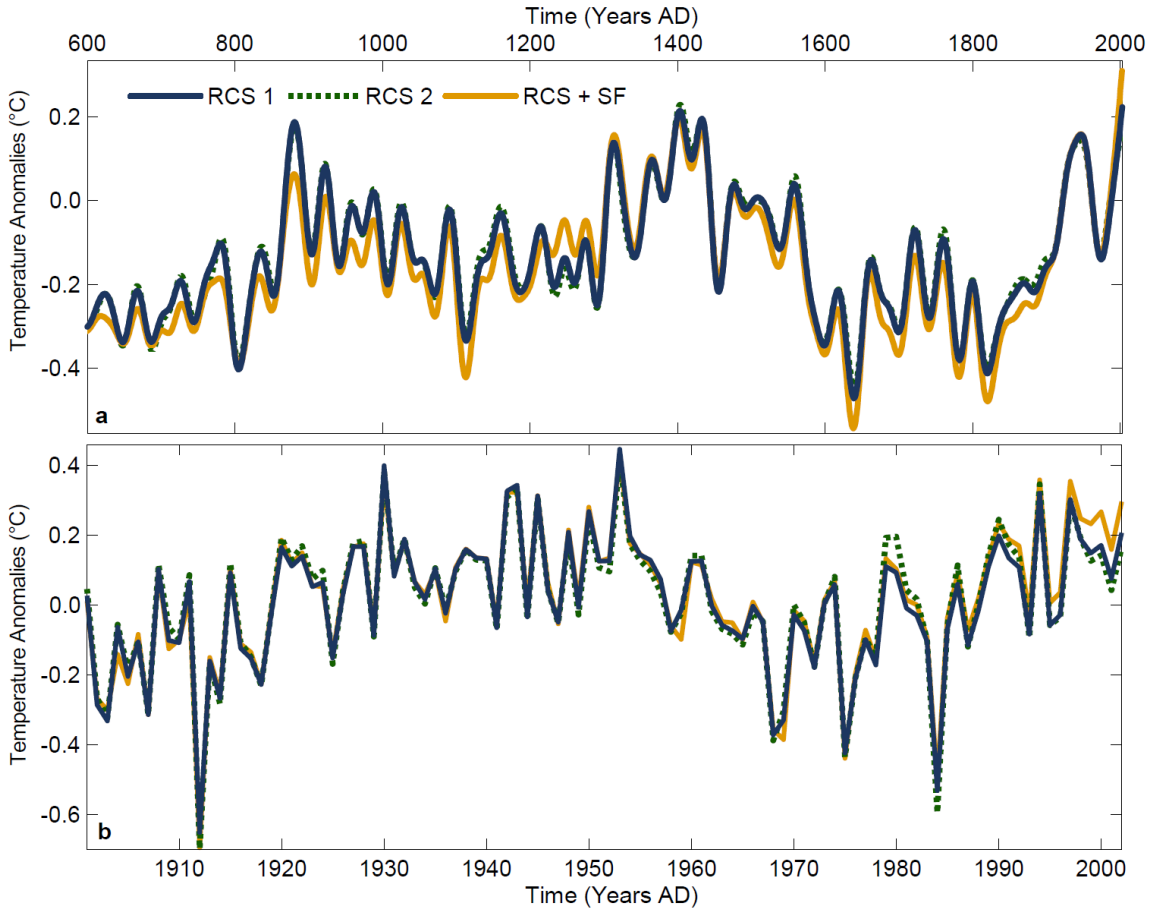

Figure S1. Reconstructions based on different detrending methodologies (see Tab. S2). **a**, Three low-pass filtered (30-years) reconstructions over the past 1400 years. **b**, The unsmoothed records over the 20th century. Temperature anomalies refer to the 1901-1976 period.

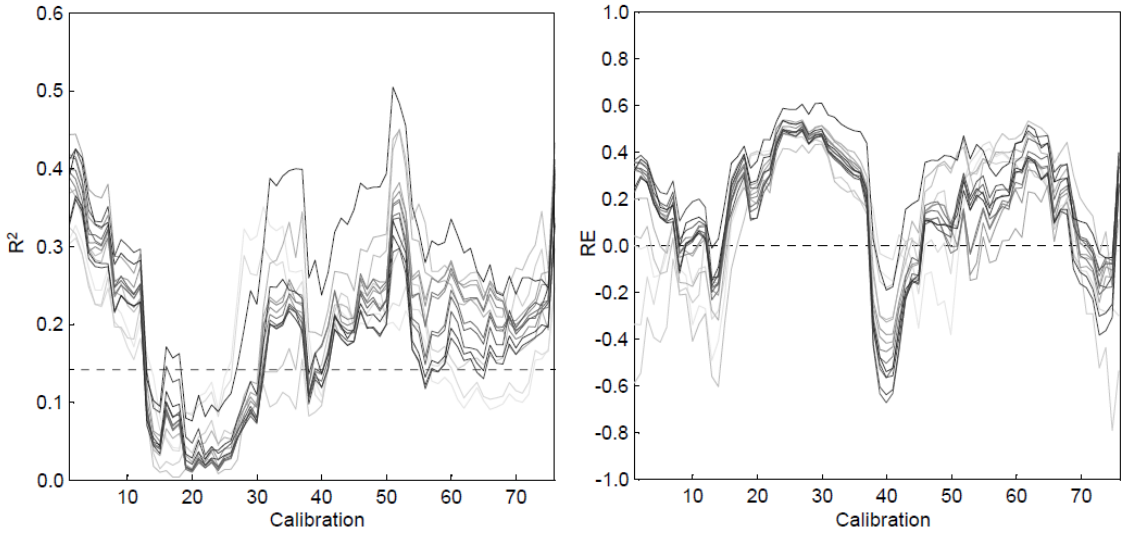

Figure S2. Validation metrics of 76 different calibration/validation periods and 15 nests (grey lines). The first calibration (validation) uses the 1901-1938 (1939-1976) period. This period is shifted by one year with each calibration so that the second setup uses 1902-1939 (1940-1976 and 1901) for calibration (validation), resulting in 76 values for each nest (plotted on the y-axis). Colors brighten with thinning replication. Dashed lines indicate $p < 0.01$ (one sided) for R^2 and zero for RE.

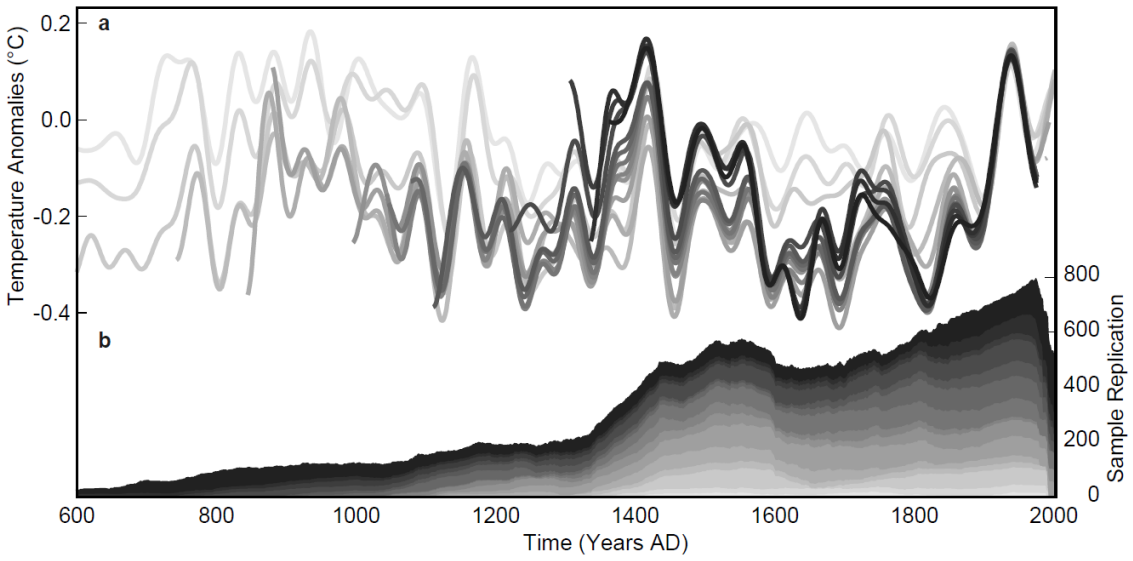

Figure S3. Reconstruction nests and replication of the single sites. a, Reconstructions calculated using changing site numbers, with the brightest (darkest) and longest (shortest) reconstruction based on 1 (15) record(s). All time-series are smoothed using a 30-year low-pass filter and expressed as anomalies w.r.t 1901-1976. The final reconstruction integrates the best-replicated reconstruction segments available at each period in time. **b**, Number of samples per site.

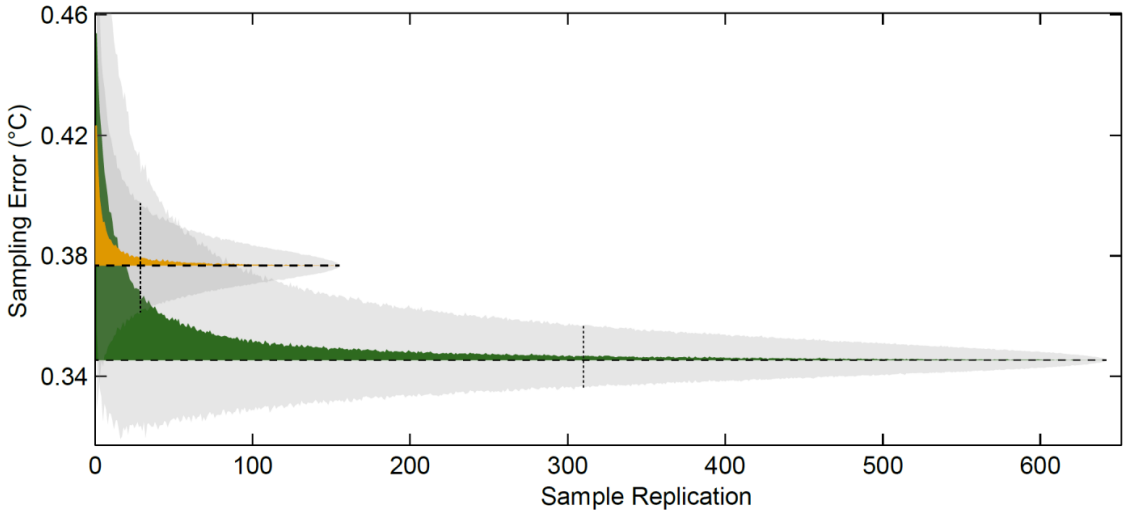

Figure S4. Replication error estimates. The green and yellow areas show the additional error introduced by decreasing sample replications for the nests including maximum and minimum numbers of sites, respectively. The dashed, horizontal lines represent the sampling error of the nests, here combined with the replication error to form the total reconstruction error. The latter error changes over time due to a declining number of old or relict tree material. From a total number of 642 (155) samples covering the 1921-1976 period, random draws produced stepwise changes in the total number of samples, from which the reconstruction error was recalculated. Monte Carlo simulations provide the mean standard error with respect to replication (upper bound of the green area) and the 95% confidence intervals (grey shades) derived from 1000 different draws. Dotted, vertical lines indicate the minimum replication actually found in the respective nests.

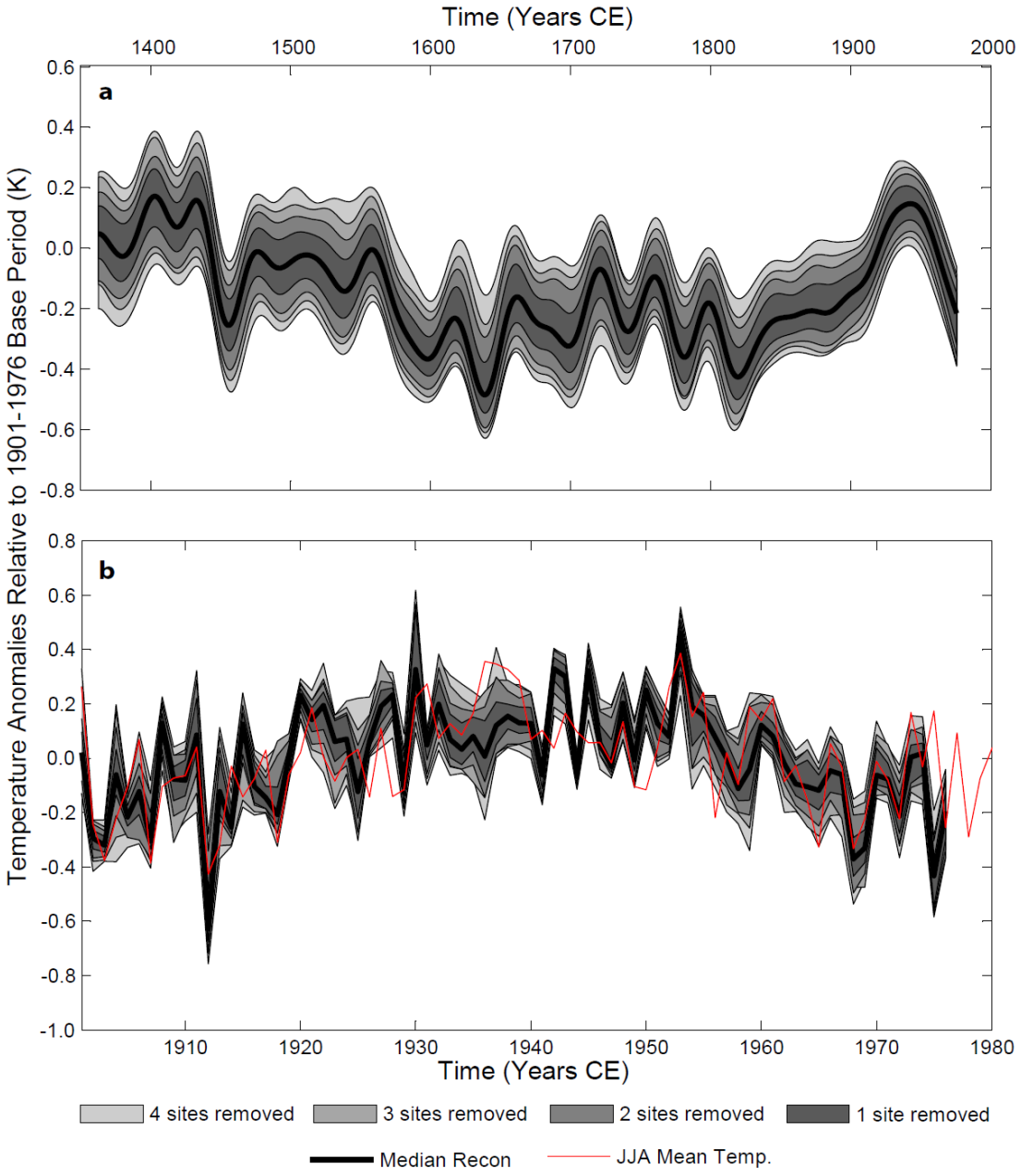

Figure S5. Jackknife tests. Uncertainty estimates for a truncated predictor network in the low (**a**) and high (**b**) frequency domain. In the upper panel records were smoothed with a 30-years spline. The upper and lower confidence intervals represent the maximum and minimum extent of alternative reconstructions based on a reduced number of site chronologies. For the removal of 2-4 chronologies the sites were randomly drawn and 100 realizations calculated.

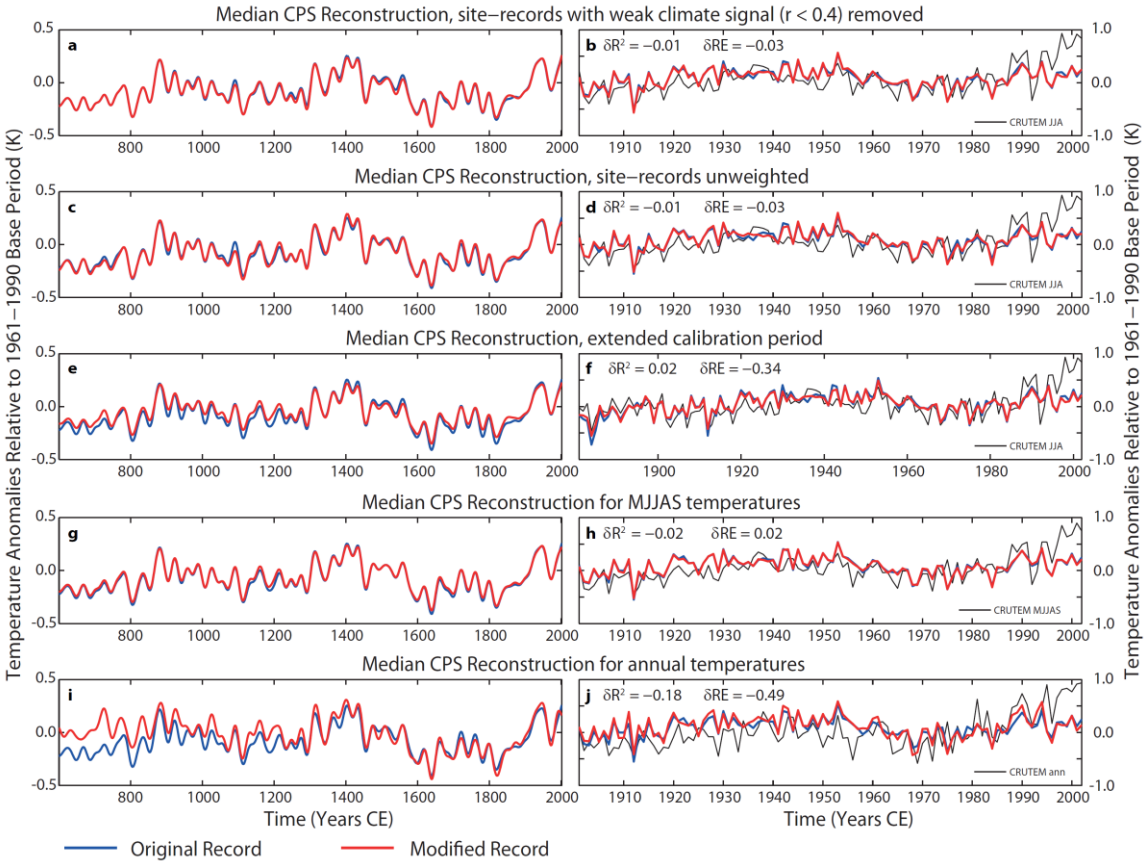

Figure S6. Alternative Calibrations. The original reconstruction was modified in multiple ways regarding the proxy network and the instrumental target. **a** and **b**, 3 site chronologies with local correlations of r < 0.4 were removed from the network. **c** and **d**, To warranty homogeneous spatial coverage proxy records were incorporated without weighting. **e** and **f**, The calibration period was extended back to 1881. **g** and **h**, The seasonal window was extended from May to September. **i** and **j**, Annual mean temperatures were used as a target. All data on the left hand side are smoothed with a 30-years spline, the right hand side shows the respective calibration period.

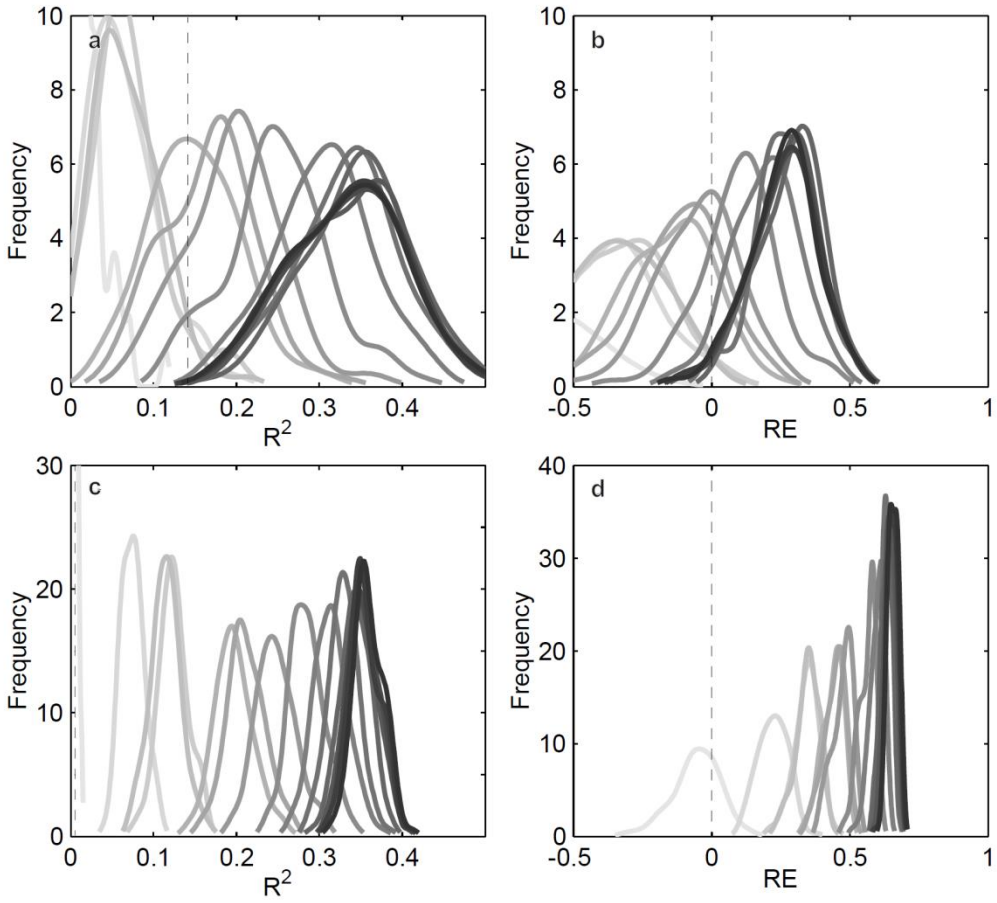

Figure S7. Validation metrics for pseudoproxy experiments. a and **b**, Nonparametric kernel-smoothing distributions of R^2 and RE of 100 different noise realizations using the sliding calibration and validation approach over the 1901-1976 period. Grey scales represent the 15 nests as in Fig. S2. **c** and **d**, Same as in **a**, but calculated over fixed and extended calibration and validation periods (1901-1976 and 850-1900). R^2 of nest #1 is ill-fitted in **a** and **c** due to its asymptotic approximation of zero. Dashed lines as in Fig. S2.

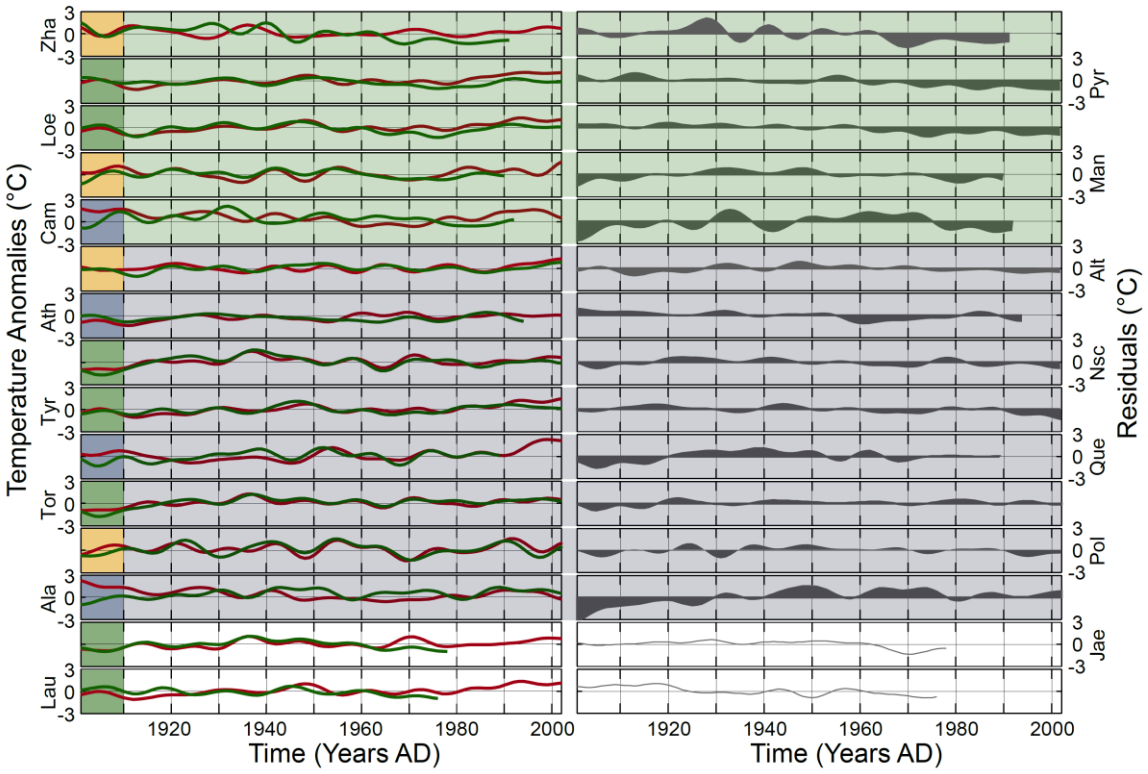

Figure S8. Local observational and modeled temperatures. Left panel shows the proxy-derived (green) and instrumental (red; CRUTEM4v) JJA temperatures at the 15 NH tree sites, all smoothed using a 10-year low-pass filter. Right panel shows the corresponding residuals emphasizing potential, late 20th century divergence issues. Sums of the residuals during the 1971-1990 period are decreasing from top to bottom, starting with largest late 20th century residuals in Zha. Referring to these sums a grouping in "diverging" (greenish background) and "non-diverging" (bluish background) proxy-records was established. The site-records from "Jae" and "Lau" are too short to be assigned to any of the groups. Green boxes on the left indicate sites from Europe, yellow from Asia, and blue from North America.

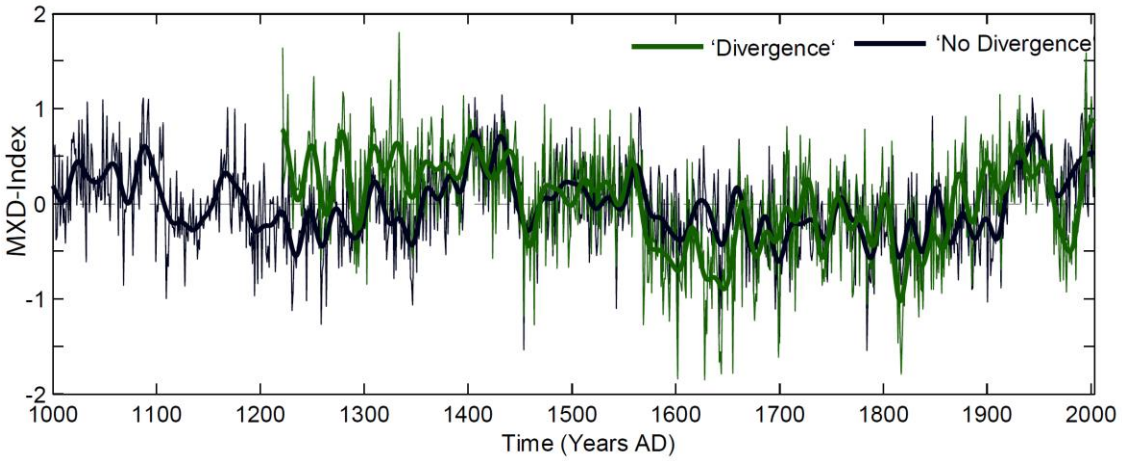

Figure S9. Mean records for potentially "diverging" and "non-diverging" MXD-chronologies (as classified in Fig. S6) smoothed with a 30-year low-pass filter. Records are truncated when based on less than 3 site-chronologies.

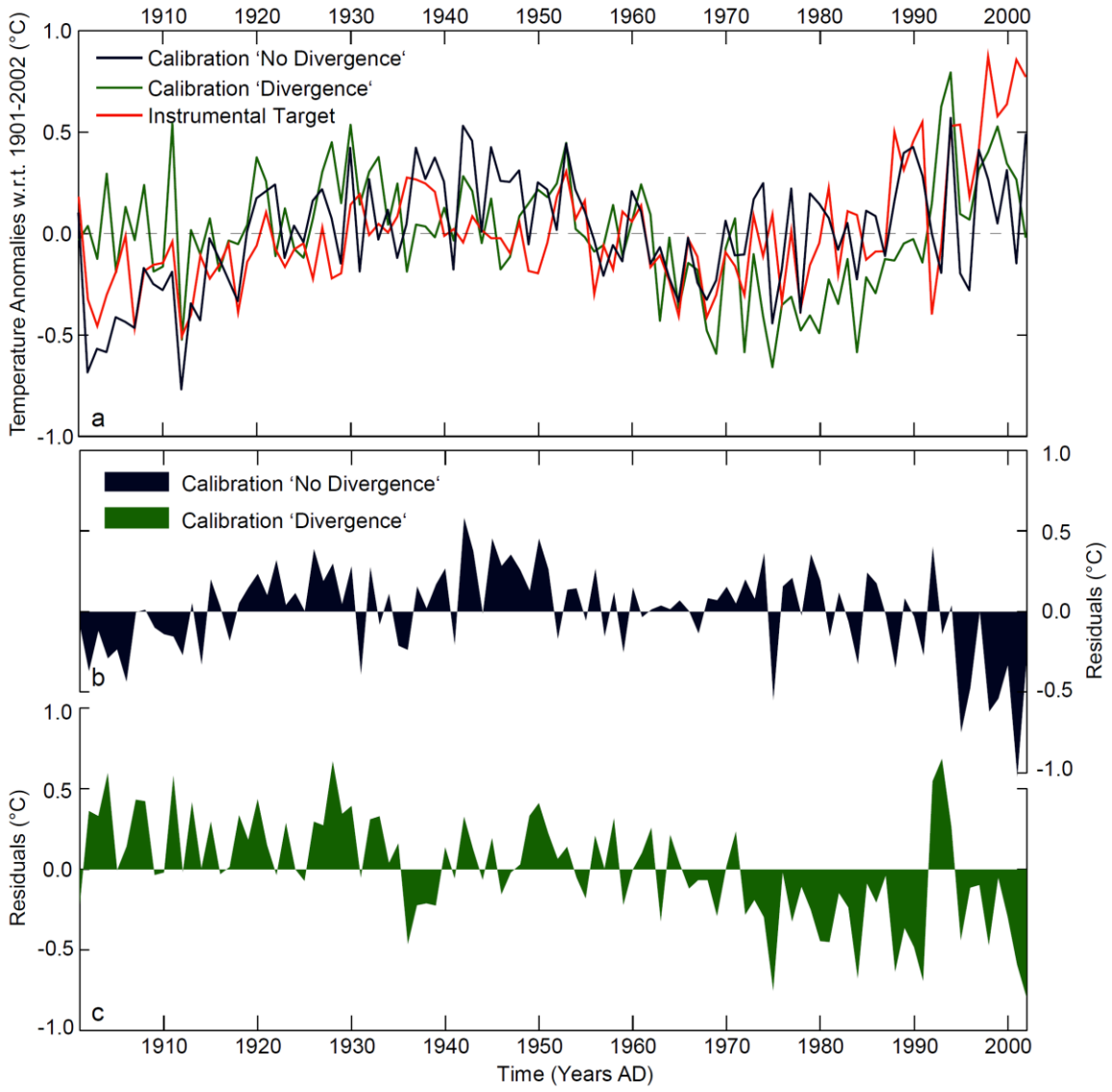

Figure S10. Reconstructions using mean curves of 5 "diverging" and 8 "non-diverging" records. a, Mean curves shown together with instrumental JJA temperatures from 1901-2002. **b** and **c**, Residuals from the observational temperature record are stronger in the "diverging" group, not only in the post 1970 period, but also during earlier periods. Warming temperatures at the turn of the millennium are not captured by the "diverging" or by the "non-diverging" data.

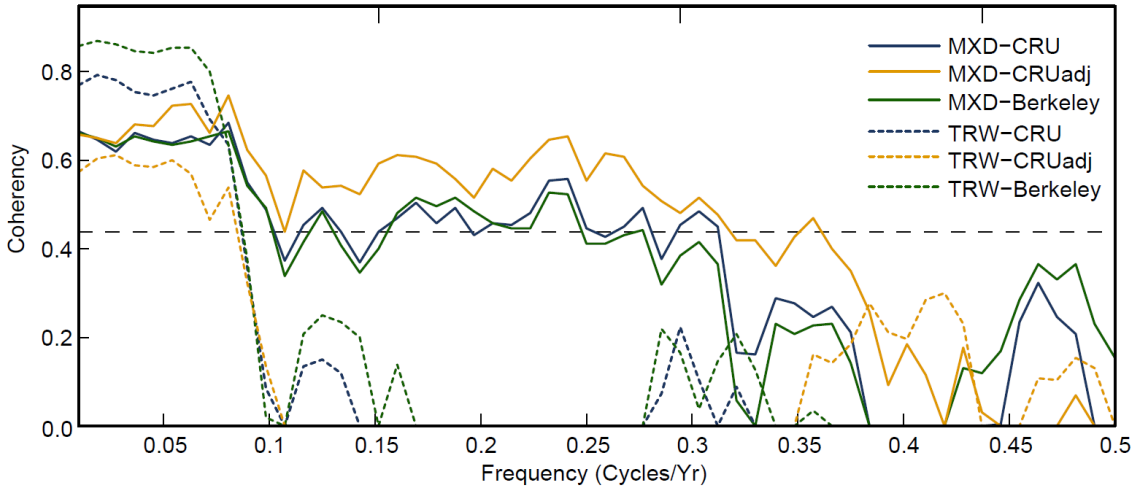

Figure S11. Coherency spectra for the 1881-1995 period. The multi-taper spectra are calculated using this reconstruction ("MXD") and the most recent tree-ring width-based reconstruction ("TRW", D'Arrigo et al. 2006) vs. the CRUTEM- and Berkeley-datasets as observational records [*Jones et al.*, 2012; *Rohde et al.*, 2013]. "CRU" and "Berkeley" refers to a simple JJA average of all grid-points between 30 and 90°N, whereas CRUadj includes an areal weighting and gapfilling as applied in this study.

References

Cook, E. R., and K. Peters (1997), Calculating unbiased tree-ring indices for the study of climatic and environmental change, *Holocene*, 7(3), 361-370.

Franke, J., D. Frank, C. C. Raible, J. Esper, and S. Brönnimann (2013), Spectral biases in tree-ring climate proxies, *Nat. Clim. Change*, *3*(4), 360-364.

Gunnarson, B. E., H. W. Linderholm, and A. Moberg (2011), Improving a tree-ring reconstruction from west-central Scandinavia: 900 years of warm-season temperatures, *Clim. Dyn.*, 36, 97-108.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res.-Atmos., 117.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng (2013), Last millennium climate and its variability in CCSM4, *Journal of Climate*, *26*(4), 1085-1111.

Masson-Delmotte, V., et al. (2013), Information from paleoclimate archives, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley, pp. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Smerdon, J.E. (2012), Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Climate Change, 3:63-77, doi:10.1002/wcc.149.

Rohde et al. (2013), A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview 1:1*, http://dx.doi.org/10.4172/2327-4581.1000101.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of Cmip5 and the experiment design, *B. Am. Meteorol. Soc.*, *93*(4), 485-498.

Thomson, D. J. (1982), Spectrum estimation and harmonic-analysis, *P. Ieee.*, *70*(9), 1055-1096.