Supporting Information

Mann *et al.* 10.1073/pnas.0805721105

SI Text

Proxy Dataset Details. Dendroclimatic data included a tree-ring network of 105 maximum latewood density (MXD) grid box (5° latitude by 5° longitude) tree-ring composite series (1, 2), 926 tree-ring series from the International Tree Ring Data Bank (see below for further details), and 5 additional tree-ring-based series (local temperature reconstructions and regional composite chronologies). The proxy dataset also includes (see below) 3 marine sediment series (from 2 locations), 14 speleothem series (from 7 locations), 19 lacustrine series (from 12 locations), 32 ice core series (from 26 locations), 15 marine coral series (from 10 locations), and 19 historical documentary series (from 15 locations). The dataset also includes 71 European composite surface temperature reconstructions back to A.D. 1500 based on a composite of proxy, historical, and early instrumental data (3). All proxy data were required to have temporal resolution no coarser than decadal to facilitate meaningful calibration against the instrumental record. Note that multiple series were used from a given location when more than one proxy variable was available (e.g., ice accumulation and oxygen isotopes from a particular ice core). See Table S1 and supplementary spreadsheet ("1290proxynames.xls") for additional details and sources for proxy data used.

A. Selection of Tree-Ring Data. Tree-ring data included 926 tree-ring series extracted from the International Tree Ring Data Bank (ITRDB, version 5.03: www.ncdc.noaa.gov/paleo/treering.html). All ITRDB tree-ring proxy series were required to pass a series of minimum standards to be included in the network: (*i*) series must cover at least the interval 1750 to 1970, (*ii*) correlation between individual cores for a given site must be 0.50 for this period, (*iii*) there must be at least eight samples during the screened period 1800–1960 and for every year used.

Series that were redundant with other compilations (e.g., series used in the aforementioned MXD network) were not included. Four other series were not included because of site-specific problems (e.g., the "ar054" series because of earthquake effects after 1811). Of the remaining series, 139 had to be replaced because of format errors in the chronology file on the ITRDB (136), or because sample depth data were missing from the chronology file (57). As a result, 926 series derived from data at the ITRDB were used in total. When sample depth data were absent, the raw ring-width data from ITRDB were used to recalculate the chronology using program ARSTAN (Version 6.05P), with the following settings: a) a single detrending fitting a cubic spline of 50% variance reduction function at 66.6.% the length of each sample, no stabilization of variance or autoregressive modeling, indices computed as ratio, that is measurement divided by curve, and chronology calculated as biweight robust mean estimate. Series were required to be complete over most of the calibration interval: 25%, 50%, and 75% of the ITRDB series used end in 1979, 1984, and 1991, whereas 105 of the 926 series end in 1995 or later. Missing values were in-filled as described below.

Tree-ring data are derived from the means of multiple detrended sample measurement series. The more flexible the detrending function used on each sample series, the greater the risk of discarding low-frequency climate information. Conversely, stiffer detrending functions result in a more conservative detrending, retaining lower-frequency climate signal, but possibly also nonclimatic information on similar time scales. Examples of such a conservative treatment would include taking the ratio of each year's tree ring value and that year's value for a fitted modified negative exponential function, or of a cubic spline with a 50% variance reduction function at 2/3 the length of the measurement series. Although the tree-ring data series we have used are all based on well replicated chronologies, and although the great majority of series used result from relatively conservative treatment of the kinds described above, even the most conservative standardization techniques necessarily limit the reliability of tree-ring reconstructions on time scales longer than a few centuries. Cook *et al.* (4) wrote "the maximum length of recoverable climate information is ordinarily related to the lengths of the individual tree-ring series used to reconstruct the millennia-long chronology." In our case, 75% of 926 ITRDB tree-ring series have a median segment length (Fig. S2) of 299 years or less. For the 40 (16) Northern Hemisphere records used in reconstructions starting at A.D.1200 (A.D.1000), 75% have a median segment length of 603 (739) years or less. Thus the capacity of the tree-ring chronologies used for periods before A.D.1200 to capture multicentennial variation is greater than for the whole ITRDB data set, but is nonetheless limited. The gridded maximum density dataset was developed differently by Osborn *et al.* (http://www.cru.uea.ac.uk/~timo/datapages/mxdtrw.htm) [as used by Rutherford *et al.* (2)] with the specific goal of capturing century to multicentury variability. In addition to the problem of segment length, there are potentially severe, time-dependent biases in tree-ring series, especially at their beginnings and ends (6). Of the five tree-ring series included in the dataset that were not obtained directly from the ITRDB or from the authors of the gridded maximum density dataset, two were standardized by using the spline approach mentioned above [Cook *et al.*'s New Zealand Reconstruction and the D'Arrigo *et al.* (7) Mongolian data] and three with the regional curve standardization approach (Cook *et al.*'s Tasmania reconstruction, Briffa *et al.*'s Tornetrask data, and Naurzbaev and Vaganov's (8) reconstruction of early summer temperature on the Taimyr peninsula). In the latter case, the RCS approach was modified by whitening the records to remove persistence.

Selection of Non-Tree-Ring Data. We selected non-tree-ring proxies with the goal of obtaining data from as wide a geographic area as possible, but at the same time we tried to limit the selection to records that were reasonably well dated and where the original analysts had shown that there was a paleoclimatic signal associated with the proxy. Given the diversity of non-tree-ring proxies, this is a much more subjective process than that which we were able to apply to the tree-ring data. One advantage of the non-tree-ring proxy series used is that, in most cases, there is little reason to believe *a priori* that there are any problems with the series that are likely to eliminate the reliability of multicentury to millennial time scale information.

S2. Preprocessing of Proxy Data. When records were available at subannual resolution, they were averaged to obtain annual mean values. To avoid aliasing bias, records with only decadal resolution were first interpolated to annual resolution and then low-pass filtered to retain frequencies $f < 0.05$ cycle per year (the Nyquist frequency for decadal sampling). The year 1995, at or shortly after which many proxy data terminate, was used as the upper limit for calibration. Because of the evidence for loss of temperature sensitivity after ≈1960 (1), MXD data were eliminated for the post-1960 interval. The RegEM algorithm of Schneider (9) was used to estimate missing values for proxy series

terminating before the 1995 calibration interval endpoint, based on their mutual covariance with the other available proxy data over the full 1850–1995 calibration interval. No instrumental or historical (i.e., Luterbacher *et al.*) data were used in this procedure.

European temperature reconstructions of Luterbacher *et al.* (3) available at 0.5° latitude/longitude resolution and at monthly (A.D. 1659–1998) or seasonal (A.D. 1500–1658) temporal resolution were upscaled to a 5° grid and annually averaged (Jan–Dec) to yield 71 distinct annual grid box series back to A.D. 1500.

Screening Procedure. To pass screening, a series was required to exhibit a statistically significant ($P < 0.10$) correlation with either one of the two closest instrumental surface temperature grid points over the calibration interval (although see discussion below about the influence of temporal autocorrelation, which reduces the effective criterion to roughly $P < 0.13$ in the case of correlations at the annual time scales).

The screening process was performed on annual time scales for annually-resolved proxies, and on decadal time scales for decadally resolved proxies. Screening was performed separately for the full available overlap interval between proxy and instrumental data (1850–1995) and the shorter calibration intervals (1896–1995 and 1850–1949) used for validation experiments. We assumed $n = 144$ nominal degrees of freedom over the 1850–1995 (146-year) interval for correlations between annually resolved records (although see discussion below about influence of temporal autocorrelation at the annual time scales), and $n = 13°$ of freedom for decadal resolution records. The corresponding one-sided $P = 0.10$ significance thresholds are $|r| = 0.11$ and $|r| = 0.34$, respectively. For the shorter (100-year) calibration intervals, we assumed $n = 98$ nominal degrees of freedom for annually resolved records, and $n = 8°$ of freedom for decadal resolution records. The corresponding one-sided $P = 0.10$ significance thresholds are $|r| = 0.13$ and $|r| = 0.42$ respectively. Owing to reduced degrees of freedom arising from modest temporal autocorrelation, the effective P value for annual screening is slightly higher ($P \approx 0.128$) than the nominal ($P = 0.10$) value. For the decadally resolved proxies, the effect is negligible because the decadal time scale of the smoothing is long compared with the intrinsic autocorrelation time scales of the data.

Although 484 (≈40%) pass the temperature screening process over the full (1850–1995) calibration interval, one would expect that no more than ≈150 (<13%) of the proxy series would pass the screening procedure described above by chance alone. This observation indicates that selection bias, although potentially problematic when employing screened predictors (see e.g. Schneider (5); note, though, that in their reply, Hegerl *et al.* (10) contest that this is actually an issue in the context of their own study), does not appear a significant problem in our case.

Instrumental Surface Temperature Dataset. Gaps in the individual annual mean (Jan–Dec) grid box surface temperature data available from 1850–2006 were in-filled by using the RegEM procedure with ridge regression described by Schneider (2001) and described in Rutherford *et al.* (11). A small number of grid boxes were not in-filled because of the availability of too few monthly values in the raw data (see Fig. S3).

Statistical Validation Results. As some (European) instrumental information from Luterbacher *et al.* (3) was incorporated back to A.D. 1500, statistical validation exercises are not entirely independent of the instrumental record for networks that include these predictors. To gauge any potential artificial inflation of skill in reconstructions using these data as predictors, separate validation experiments and reconstructions back to A.D. 1500 were also performed without using these predictors.

The details of the early and late validation exercises for all of the various alternative predictor networks (e.g., "all proxy," "screened proxy," "frozen A.D. 1000," "no tree-ring," "no Luterbacher") and target instrumental series (NH, SH, and global, land, combined land plus ocean, and using both CRU and ICRU instrumental series) are provided in supplementary spreadsheets ("eiv-validation.xls" and "cps-validation.xls").

Validation results are plotted in Fig. S4 for the specific NH, SH, and global mean validation experiments described in the manuscript.

Supplementary Reconstruction Results. The full set of global, SH, and NH reconstructions passing validation in the CPS and EIV experiments are shown in Figs. S5 and Fig. S6.

Sensitivity Analyses (NH Temperatures). ***Sensitivity to use of tree-ring data in early centuries.*** We examined the influence of use of tree-ring data on the resulting long-term CPS and EIV reconstructions in Fig. S7 below.

Potential data quality problems. In addition to checking whether or not potential problems specific to tree-ring data have any significant impact on our reconstructions in earlier centuries (see Fig. S7), we also examined whether or not potential problems noted for several records (see Dataset S1 for details) might compromise the reconstructions. These records include the four Tijander *et al.* (12) series used (see Fig. S9) for which the original authors note that human effects over the past few centuries unrelated to climate might impact records (the original paper states "Natural variability in the sediment record was disrupted by increased human impact in the catchment area at A.D. 1720." and later, "In the case of Lake Korttajarvi it is a demanding task to calibrate the physical varve data we have collected against meteorological data, because human impacts have distorted the natural signal to varying extents"). These issues are particularly significant because there are few proxy records, particularly in the temperature-screened dataset (see Fig. S9), available back through the 9th century. The Tijander *et al.* series constitute 4 of the 15 available Northern Hemisphere records before that point.

In addition there are three other records in our database with potential data quality problems, as noted in the database notes: Benson *et al.* (13) (Mono Lake): "Data after 1940 no good—water exported to CA;" Isdale (14) (fluorescence): "anthropogenic influence after 1870;" and McCulloch (15) (Ba/Ca): "anthropogenic influence after 1870".

We therefore performed additional analyses as in Fig. S7, but instead compaired the reconstructions both with and without the above seven potentially problematic series, as shown in Fig. S8.

Leverage of Individual Records (CPS approach). We examine the influence of individual proxy records on the resulting CPS reconstructions in Fig. S10.

Robustness of Reconstructions with respect to Calibration Period. We examine the issue of sensitivity to calibration interval used in Fig. S11.

Analyses Using Synthetic "Pseudoproxy" Networks. Comparison of EIV and CPS with networks of varying sparseness; alternative CPS scaling #1 – trend matching; and alternative CPS scaling #2 – signal matching.

Comparisons of various NH reconstructions using alternative scaling conventions. We examined the impact of the various alternative CPS scalings discussed above on the CPS reconstructions of this study in Figs. S15 and S16.

SI Methods. We performed separate SH and NH reconstructions using both CPS and EIV methods. Global mean reconstructions were defined simply as the arithmetic average of the Northern and Southern Hemisphere reconstructions. Reconstructions were initiated at the first centennial starting date (e.g., A.D. 400) for which the reconstruction passes statistical validation (i.e., significance breaches the 95% level). As a safeguard against the inclusion of redundant statistical predictors and statistical overfitting, the set of additional series that become available at the start of each subsequent century (e.g., A.D. 500, A.D. 600, etc.) were admitted into the network only when their use results in an increase in scores for both skill metrics relative to the set used for the previous century.

CPS Approach. We used the CPS approach implementation described by Mann *et al.* (16). All proxy records in our network with an apparent local temperature signal (as determined through application of the screening procedure described above), were retained. The records were decadally smoothed by using a low-pass filter with cutoff $f = 0.1$ cycle per year, as in Mann and Jones (17) to ensure roughly common effective temporal resolution among annual and decadally resolved proxy series. The series were then converted to standardized anomalies by centering and scaling them to have zero mean and unit decadal standard deviation over the calibration interval. All proxies available within a given instrumental surface temperature grid box were then averaged and scaled to the same mean and decadal standard deviation as the nearest available 5° latitude by longitude instrumental surface temperature grid box series over the calibration period. Of the 226 total grid boxes available back to at least A.D. 1800 (see Fig. 1*B*), 136 extend back to A.D. 1500, 21 to A.D. 1000, and 11 to year 0. The gridded proxy data were then areally weighted, spatially averaged over the target hemisphere, and scaled to have the same mean and decadal standard deviation (we refer to the latter as "variance matching") as the target hemispheric (NH or SH) mean temperature series. In regions within North America and Eurasia with especially dense spatial coverage, grid boxes were first averaged over 15° resolution boxes before inclusion in the spatial average to diminish potential bias due to variations in spatial sampling density (see Fig. 1*B*). As a cross-check, we considered two alternative CPS scaling procedures. The first procedure attempts to preserve signal amplitudes by matching trends rather than variances over the calibration interval (we refer to this as "trend matching"). The second method divides the "variance matched" reconstruction by an estimated signal-to-noise amplitude ratio diagnosed over the calibration period (we refer to this as "signal matching"). This latter procedure is equivalent to the more sophisticated procedure used by Hegerl *et al.* (18) in the limit where uncertainties in the instrumental data are negligible relative to those in the proxy data. These two alternative procedures generally produced similar results, with isolated exceptions (see above). Analyses with synthetic proxy data suggest that the simpler variance matching procedure is likely to yield more skillful reconstructions than these two alternative procedures, particularly with sparse proxy data networks (compare Tables S2–S4).

Mann *et al.* (16) performed tests of the CPS method as described above using synthetic "pseudoproxies" derived from a long forced climate model simulation. They concluded that the approach is likely to produce skillful reconstructions with as few as a dozen proxy series (the rough number used in many past studies by e.g., Bradley and Jones (19); Crowley and Lowery (20); Esper *et al.* (21) for signal-to-noise ratios similar to those in our screened network (SNR $\approx$ 0.4). However, as the network size becomes increasingly sparse, it becomes increasingly likely that certain individual features of the reconstructions (e.g., a particular cooling episode) may be entirely spurious, arising from the sampling statistics of a small noisy sample (see above).

Regularized Errors-in-Variables Approach. Here, networks of available proxy data (global, hemisphere-confined, and locally screened) were used as statistical predictors in the estimation of a single target predictand defined by the NH or SH hemispheric mean instrumental series, using the regularized expectation-maximization or "RegEM" algorithm (9). RegEM has been applied to the problem of proxy-based climate field reconstruction (CFR) in several past studies [Mann and Rutherford (22); Zhang *et al.* (23); Rutherford *et al.* (2); and Mann *et al.* (16). Here, unlike CFR applications, we used RegEM to estimate only the single global/hemispheric series of interest, rather than a spatially-explicit field, so that the procedure serves as an iterative regularized variant on multivariate regression. In tests using model-generated pseudoproxy data (24), this method was found to yield very similar results to the full CFR version of the RegEM procedure.

We used the version of the RegEM algorithm described by Mann *et al.* (24) in which truncated total least squares (TTLS), rather than ridge regression, is used as a regularization scheme, and where a "hybrid" procedure is used to calibrate information separately in "interdecadal" (periods longer than 20 years; includes annually and decadally resolved proxies) and "interannual" (periods shorter than or equal to 20 years; includes only annually resolved proxies) bands. The resulting annual reconstructions were then decadally smoothed as described above. All predictors were first standardized by centering and normalizing them to have zero mean and unit standard deviation, respectively, over the calibration interval. For simplicity, all predictors were given unit weight in the analysis, as previous tests have shown a broad insensitivity to precise data weighting procedures (see ref. 24). Unlike earlier studies [e.g., Mann *et al.* (25)] no dimensional reduction procedures were applied to dense regional proxy networks, because the RegEM approach already employs an explicit regularization procedure to guard against colinearity of predictors (24). As a further safeguard against potentially nonrobust results, a minimum of seven predictors in a given hemisphere was required in implementing the EIV procedure.

The reliability of this procedure in reconstructing long-term temperature trends has been demonstrated previously (24) in tests using synthetic proxy data generated from model simulations of the past millennium with attributes (i.e., signal-to-noise ratios and noise "redness") similar to those estimated above for the proxy dataset used in this study (as noted above, we estimate SNR $\approx$ 0.4, whereas the average annual autocorrelation of the network is small, $\rho \approx 0.4$, over all intervals used).

Without further modification (e.g., in the absence of the quantitative estimates of uncertainties for each data series), the RegEM procedure assumes that relative errors are uniform over all series. In synthetic tests described by Mann *et al.* (24), results are found to be broadly insensitive to the relative assumed error amplitudes.

1. Briffa KR, *et al.* (2001) Low-frequency temperature variations from a northern tree-ring density network. *J Geophys Res* 106:2929–2941.
2. Rutherford S, *et al.* (2005) Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to methodology, predictor network, target season and target domain. *J Climate* 18:2308–2329.
3. 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.
4. Cook ER, Briffa KR, Meko DM, Graybill DA, Funkhouser G (1995) The 'segment length curse' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* 5:229–237.
5. Schneider T (2007) Uncertainty in climate sensitivity estimates. *Nature*, 446, E1.

6. Melvin T (2004) Historical growth rates and changing climatic sensitivity of boreal conifers. Ph.D. thesis, Climatic Research Unit, East Anglia, UK.
7. D'Arrigo RD, Wilson R, Jacoby G (2006) On the long-term context for 20th century warming. *J Geophys Res* 111:D03103.
8. Naurzbaev MM, Vaganov EA (2000) Variation of early summer and annual temperature in east Taymir and Putoran (Siberia) over the last two millennia inferred from tree rings. *J Geophys Res* 105:7317–7326.
9. Schneider T (2001) Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J Climate* 14:853–871.
10. Hegerl G, Crowley TJ, Hyde WT, Frame DJ (2007) and Hegerl *et al.* reply. *Nature*, 446:E2.
11. Rutherford S, Mann ME, Delworth TL, Stouffer R (2003) Climate field reconstruction under stationary and nonstationary forcing. *J Climate* 16:462–479.
12. Tiljander M, Saarnisto M, Ojala AEK, Saarinen T (2003) A 3000-year palaeoenvironmental record from annually laminated sediment of Lake Korttajärvi, Central Finland. *Boreas* 26:566–577.
13. Benson L, *et al.* (2003) Influence of the Pacific Decadal Oscillation on the climate of the Sierra Nevada, California and Nevada. *Quaternary Research* 59:151–159.
14. Isdale PJ, Stewart BJ, Tickle KS, Lough JM (1998) Palaeohydrological variation in a tropical river catchment: A reconstruction using fluorescent bands in corals of the Great Barrier Reef, Australia. *Holocene* 8:1–8.
15. McCulloch M, *et al.* (2003) Coral record of increased sediment flux to the inner Great Barrier Reef since European settlement. *Nature* 421:727–730.
16. Mann ME, Rutherford S, Wahl E, Ammann C (2005) Testing the fidelity of methods used in proxy-based reconstructions of past climate. *J Climate* 18:4097–4107.
17. Mann ME, Jones PD (2003) Global surface temperature over the past two millennia. *Geophy Res Lett* 30:1820.
18 Hegerl GC, Crowley TJ, Hyde WT, Frame DJ (2006) Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature* 440:1029–1032.
19. Bradley RS, Jones PD (1993) ''Little Ice Age'' summer temperature variations: Their nature and relevance to recent global warming trends. *Holocene* 3:367–376.
20. Crowley TJ, Lowery TS (2000) How warm was the Medieval Warm Period? A comment on 'Man-made versus natural climate change'. *Ambio* 39:51–54.
21. Esper J, Cook ER, Schweingruber FH (2002) Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability. *Science* 295:2250–2253.
22. Mann ME, Rutherford S (2002) Climate reconstruction using 'Pseudoproxies'. *Geophys Res Lett* 29:1501.
23. Zhang Z, Mann ME, Cook ER (2004) Alternative methods of proxy-based climate field reconstruction: application to summer drought over the conterminous United States back to A.D. 1700 from tree-ring data. *Holocene* 14:502–516.
24. Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. *J Geophys Res* 112:D12109.
25. Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* 392:779–787.
26. Huang S, Pollack HN, Shen P-Y (2000) Temperature trends over the past five centuries reconstructed from borehole temperature. *Nature* 403:756–758.
27. Mann ME, Rutherford S, Bradley RS, Hughes MK, Keimig FT (2003) Optimal surface temperature reconstructions using terrestrial borehole data. *J Geophys Res* 108:4203.
28. Rutherford S, Mann ME (2004) Correction to ''Optimal surface temperature reconstructions using terrestrial borehole data.'' *J Geophys Res* 109:D11107.

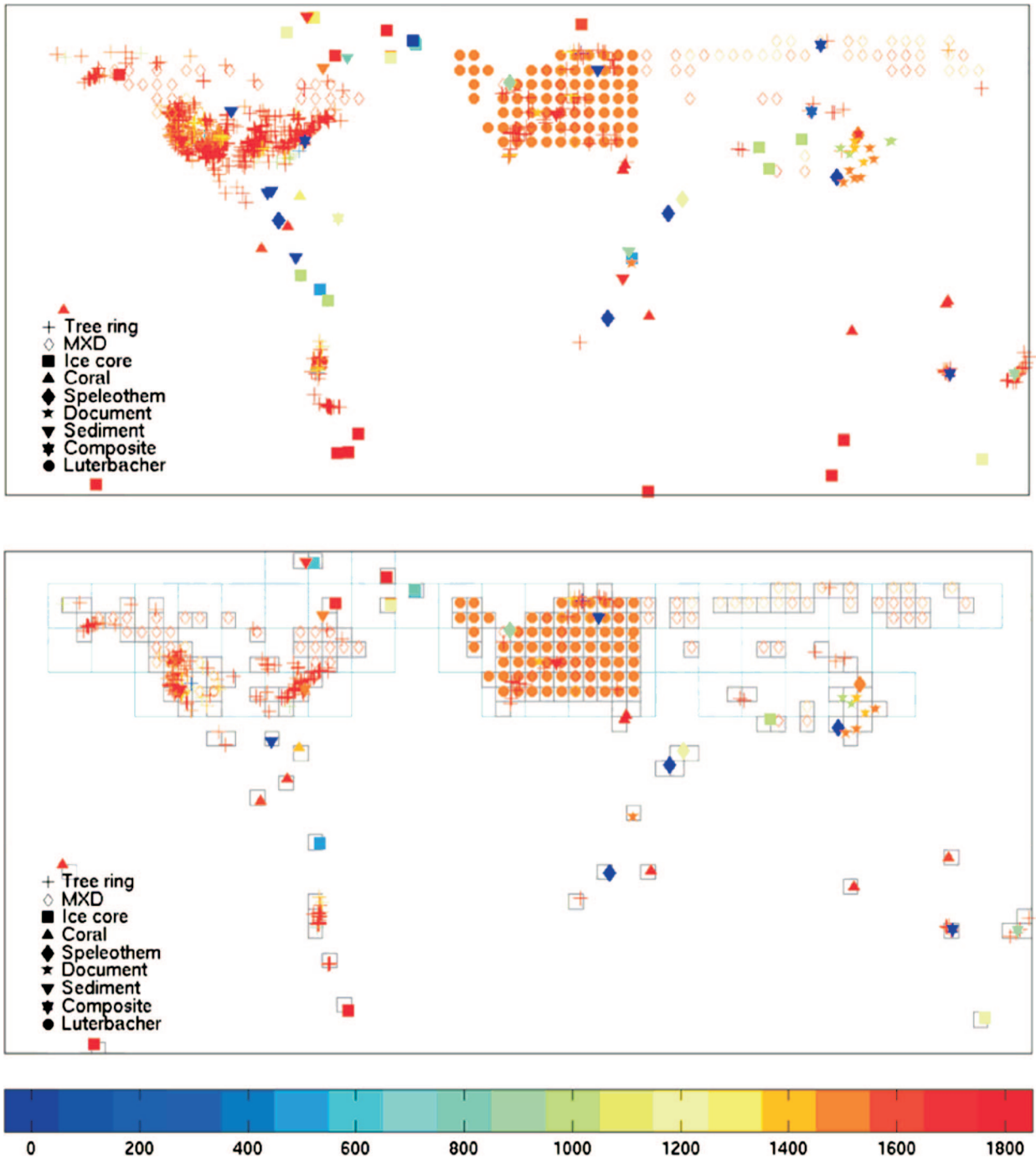


Fig. S1. Spatial distribution of full (*A*) and screened (*B*) (based on 1850–1995 interval) proxy database (see Table S1). Nine different proxy types are denoted with different symbols as shown in the map. Beginning dates of proxy records are represented by color scale shown. For the screened network, the location of 5° grid cells used in CPS reconstructions (and larger 15° regions discussed in *SI Methods* in *SI Text*) are superimposed. Color bar indicates how far back the various proxy records are available.

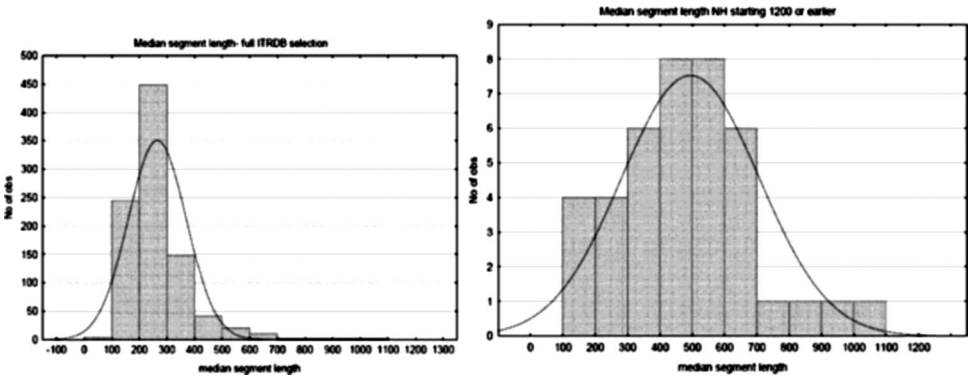


Fig. S2. Histogram of the median segment lengths of the 926 ITRDB chronologies (*Left*) and of those from the Northern Hemisphere starting in A.D. 1200 or earlier (*Right*).

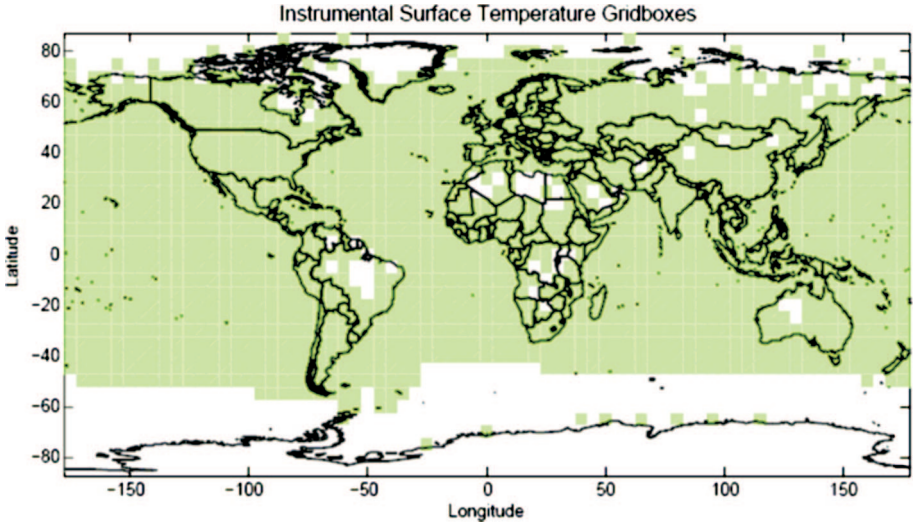


Fig. S3. Distribution of in-filled instrumental surface temperature grid boxes used in study.

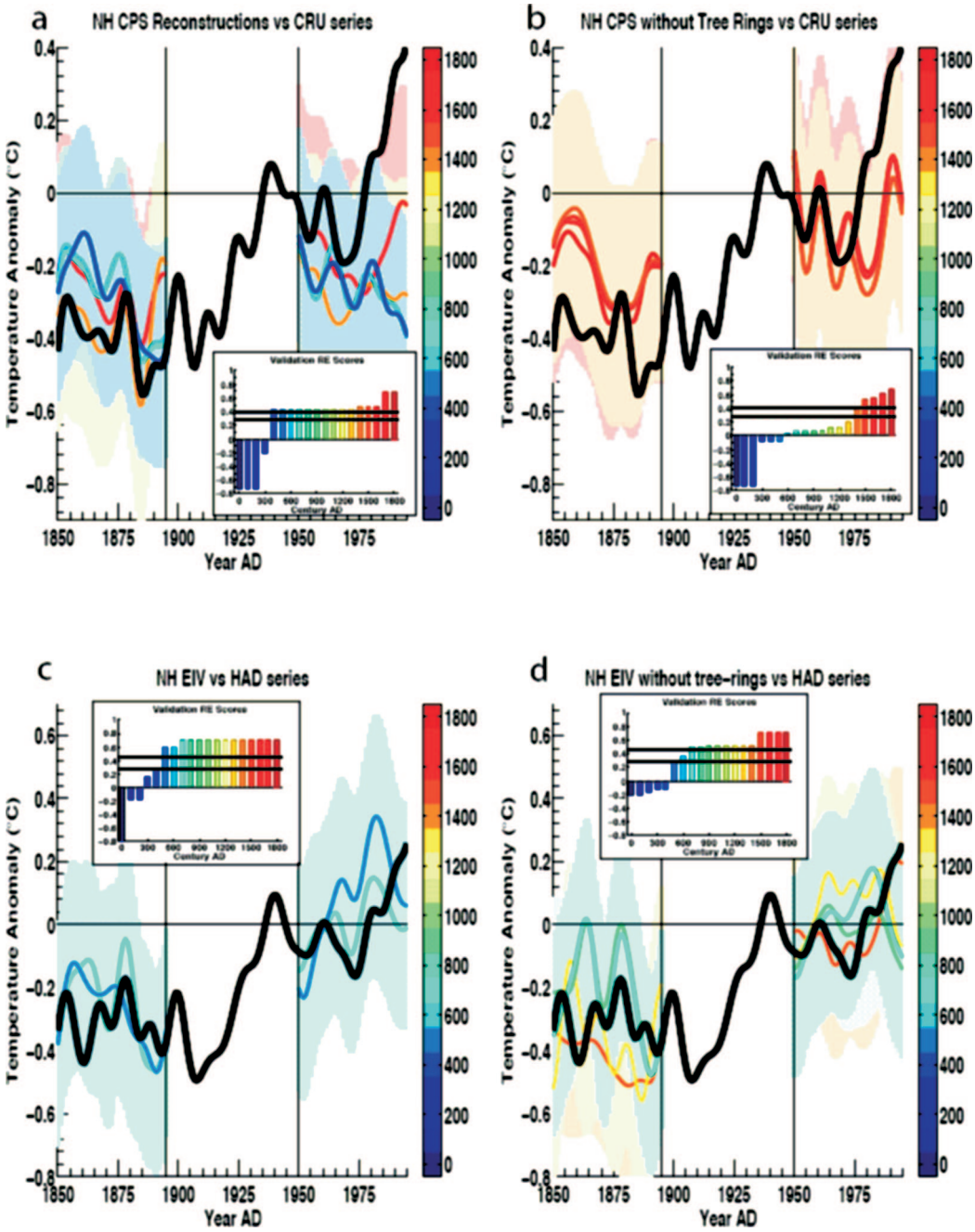


Fig. S4. As in Fig. 2 of the main paper, but showing results for NH land plus ocean (*a*) , SH land (*b*), and global land (*c*).

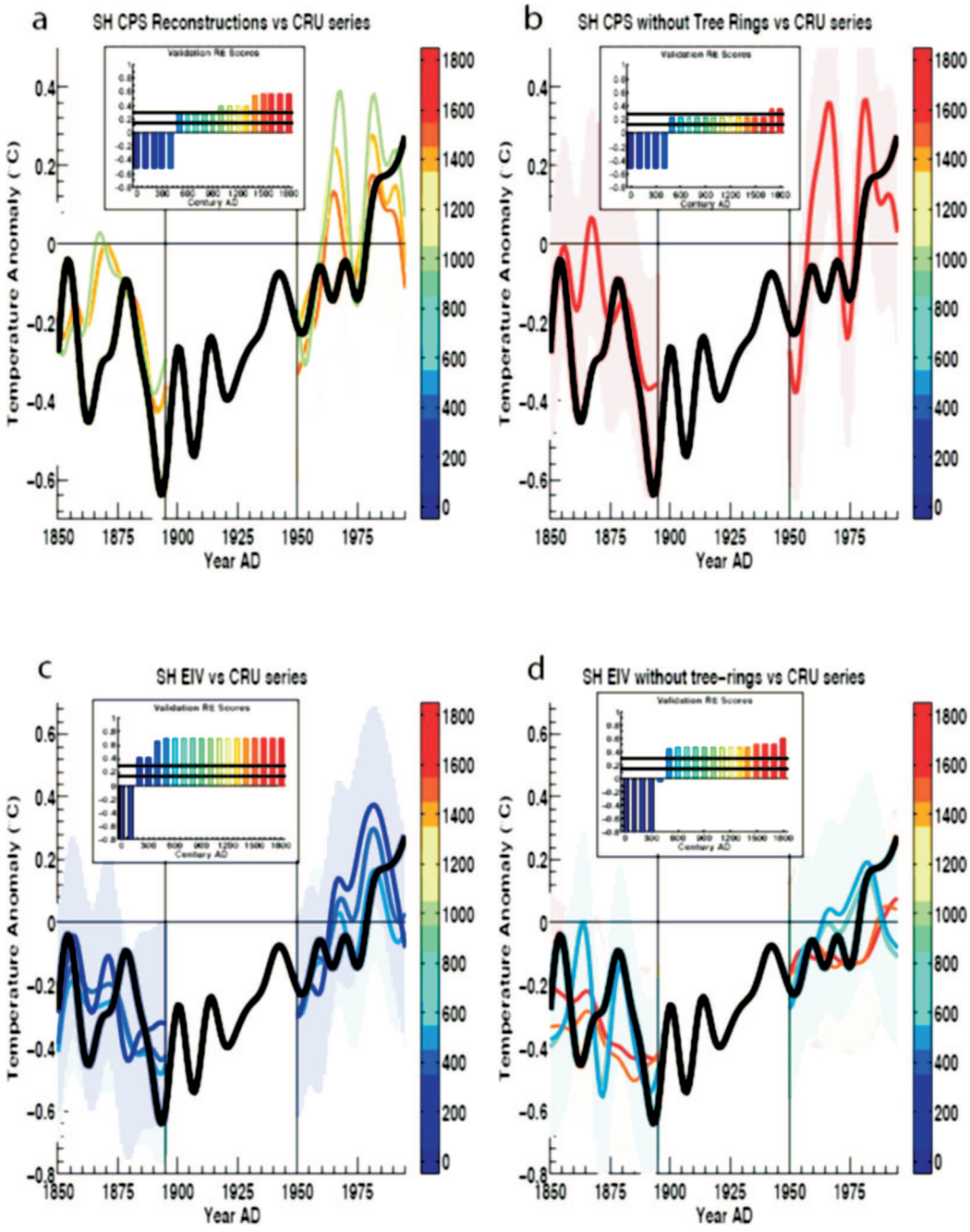


Fig. S4. (Continued).

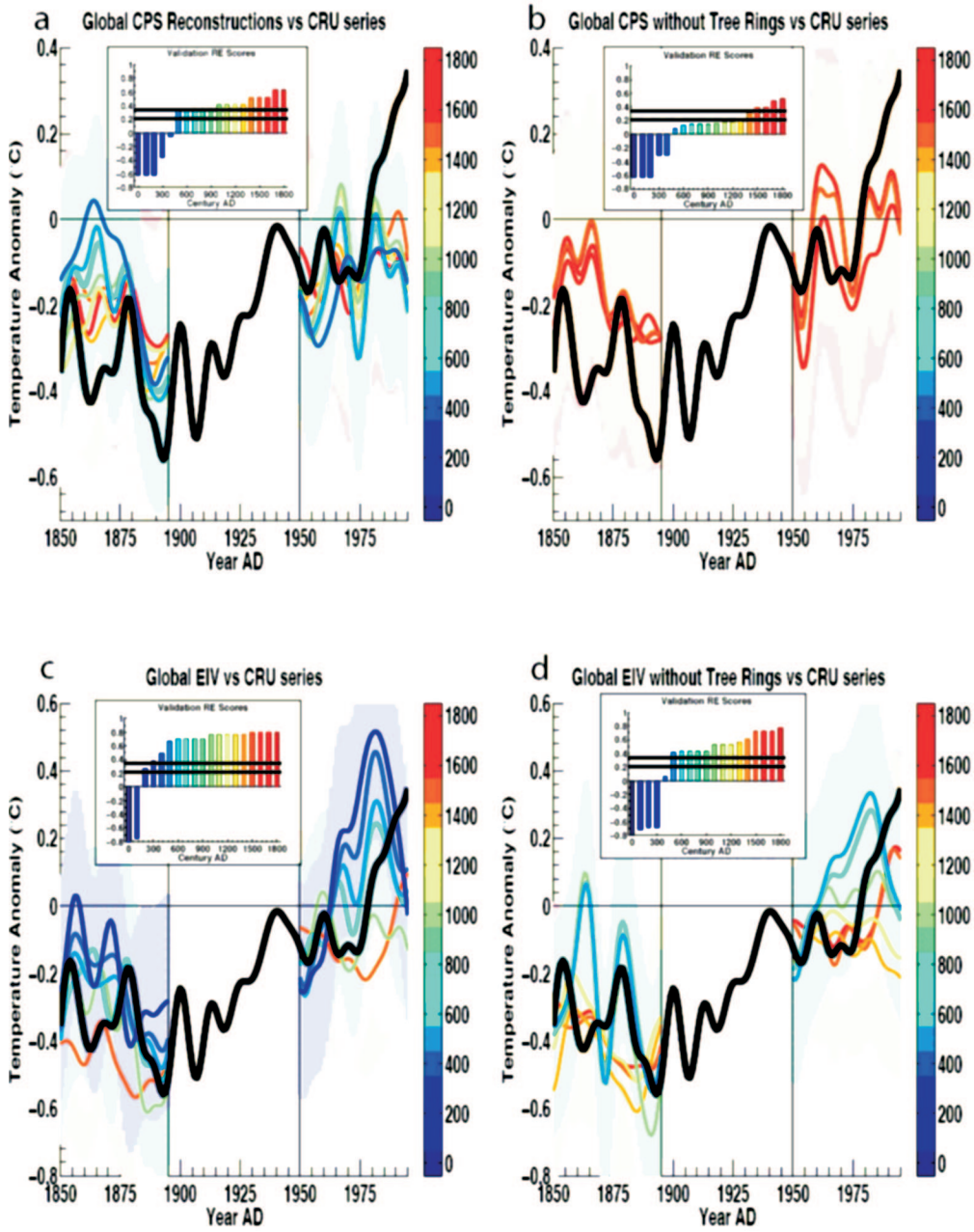


Fig. S4. (Continued).

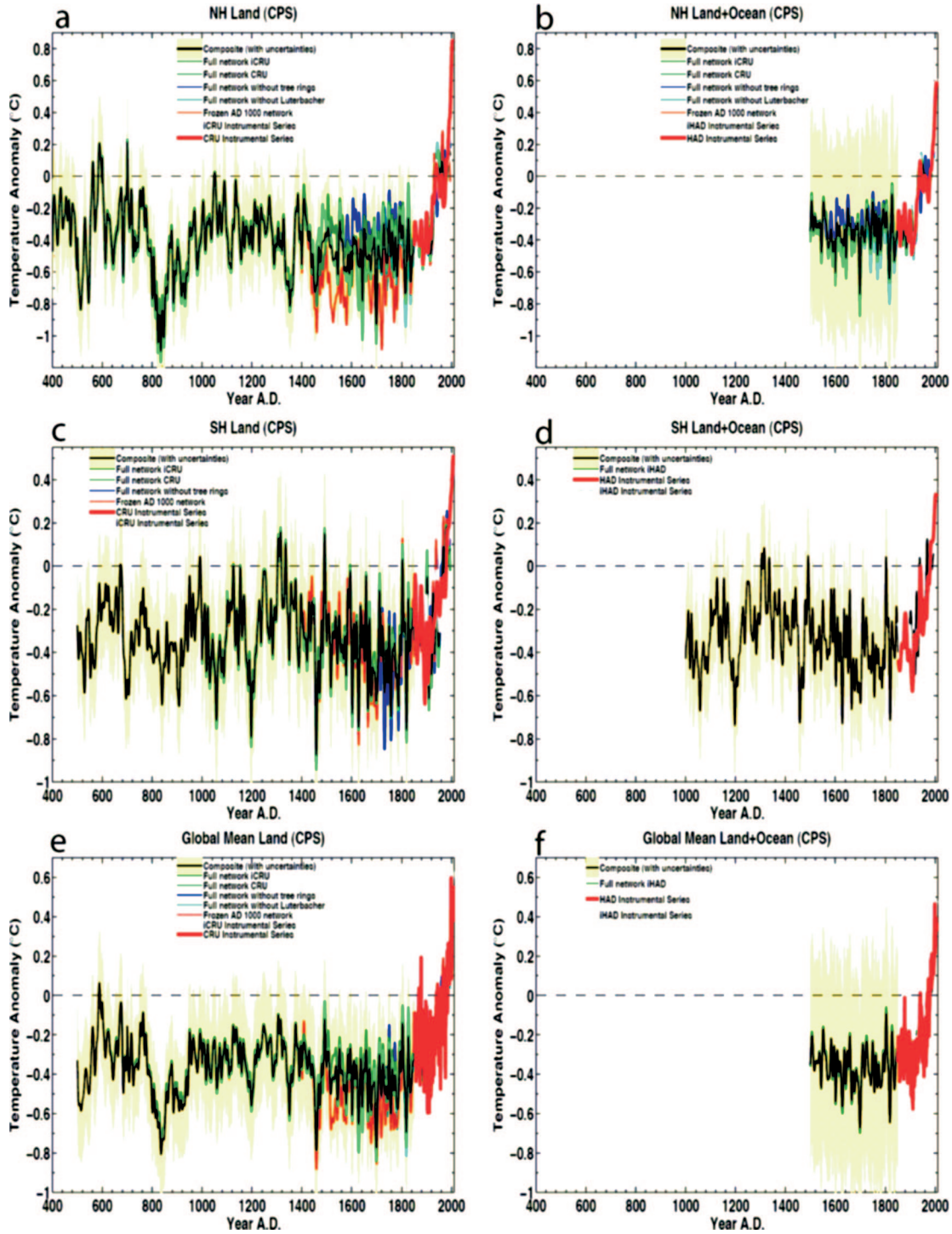


Fig. S5. Comparison of NH mean showing land (*a*); combined land plus ocean (*b*), SH land/land plus ocean mean (*c* and *d*), and global mean (*e* and *f*). Surface temperature reconstructions based on CPS method, and using the various proxy networks and target instrumental series described in text are shown. Only reconstructions that passed validation are shown. Shown also are composites of the multiple series shown, and their associated 95% confidence intervals. Decadally smoothed CRU land and Had land plus ocean instrumental series (1850–2006) are shown for comparison.

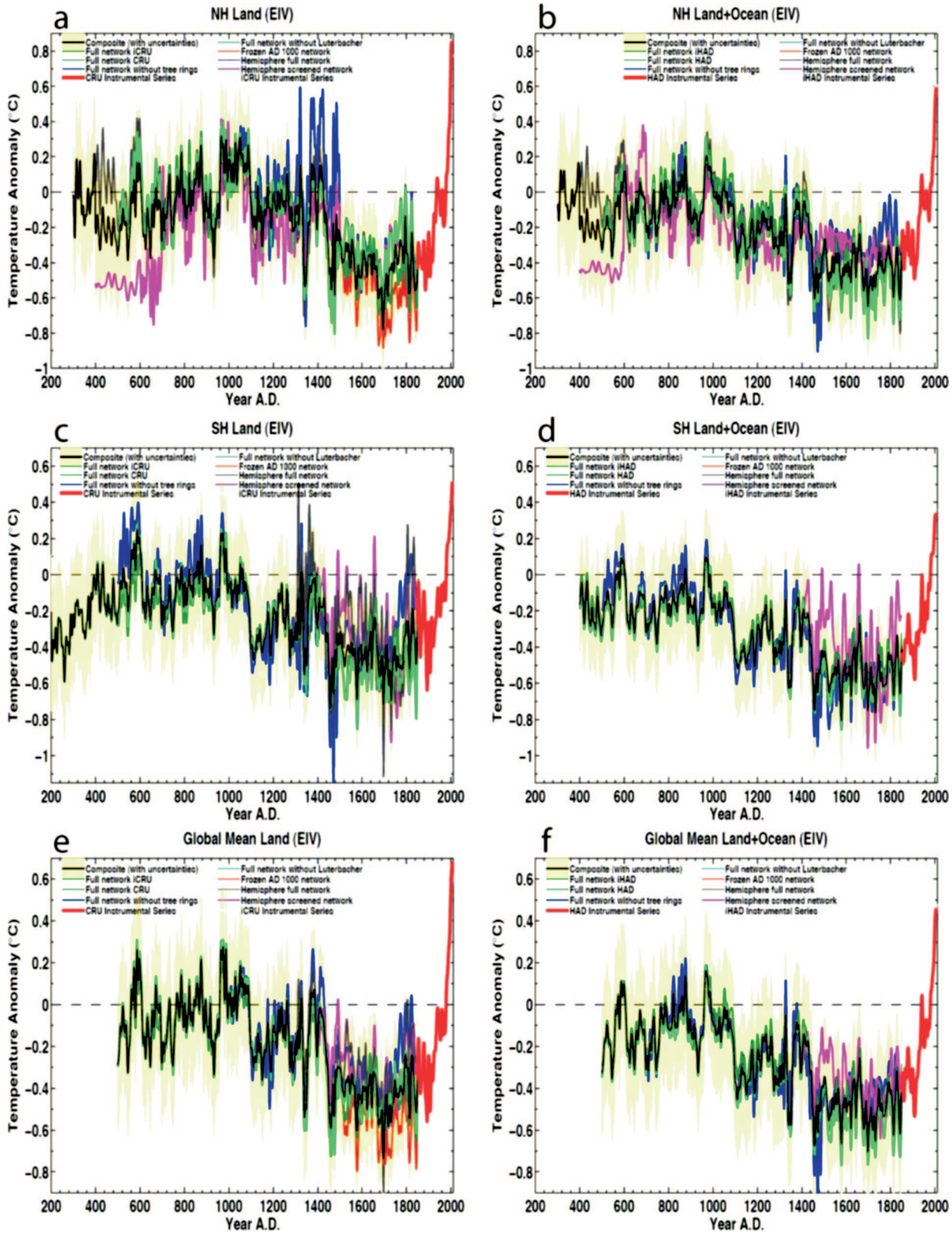


Fig. S6. Comparison of NH mean for land (*a*); combined land+ocean (*b*), SH land/combined land plus ocean mean (*c* and *d*), and global mean (*e* and *f*). Surface temperature reconstructions based on EIV method and using the various proxy networks and target instrumental series as described. Details are as in Fig. S4.

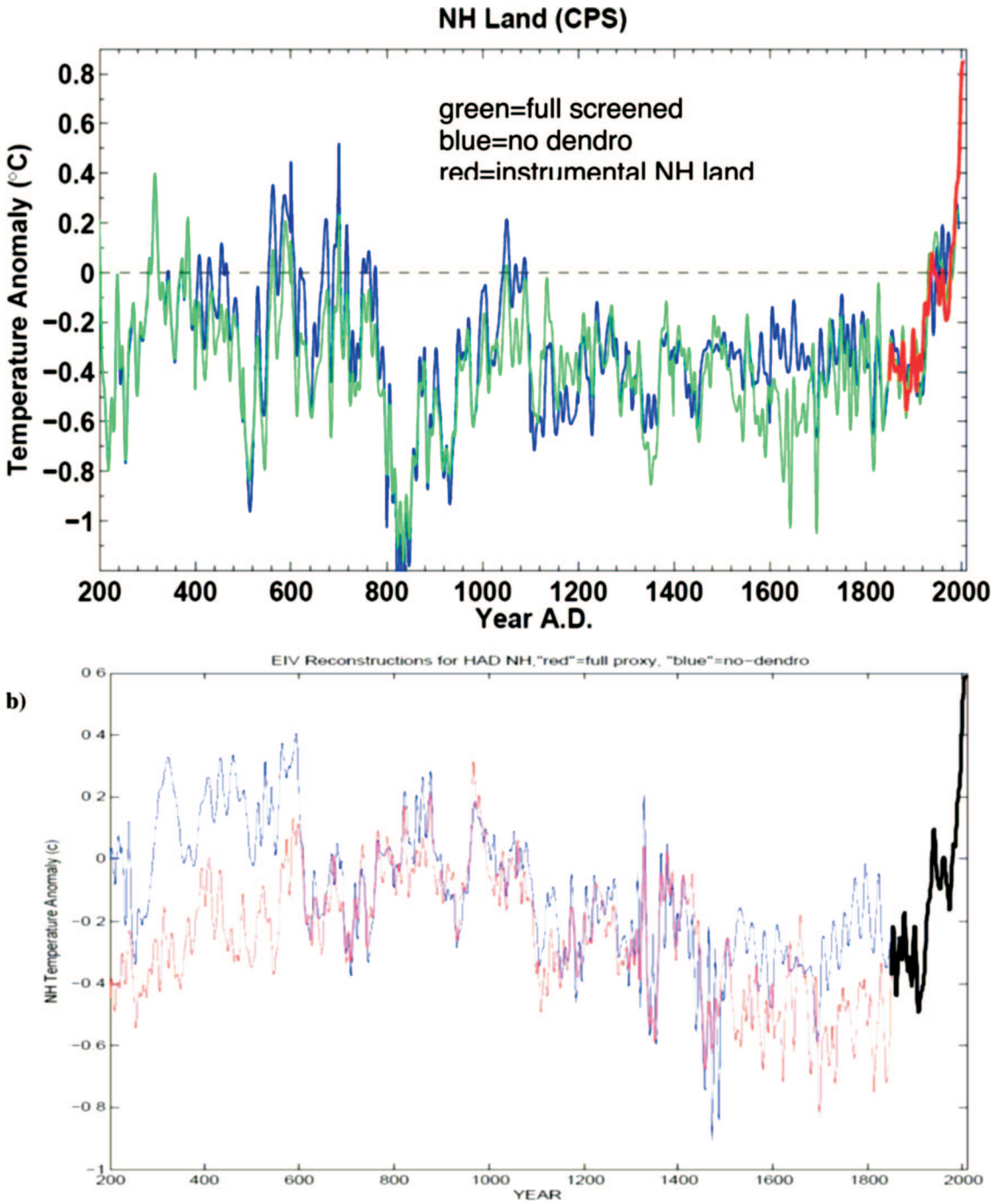


Fig. S7. Comparison of long-term CPS NH land (*a*) and EIV NH land plus ocean (full global proxy network) (*b*) with and without using tree-ring data [note, e.g., from Fig. S5*B* that the ''full proxy network'' (''no dendro'' network) reconstructions are not skillful before A.D. 500 (before A.D. 700). The point of this comparison is simply that the anomalousness of recent warming in the long-term context of the reconstructions does not depend on whether or not tree-ring data have been used.]

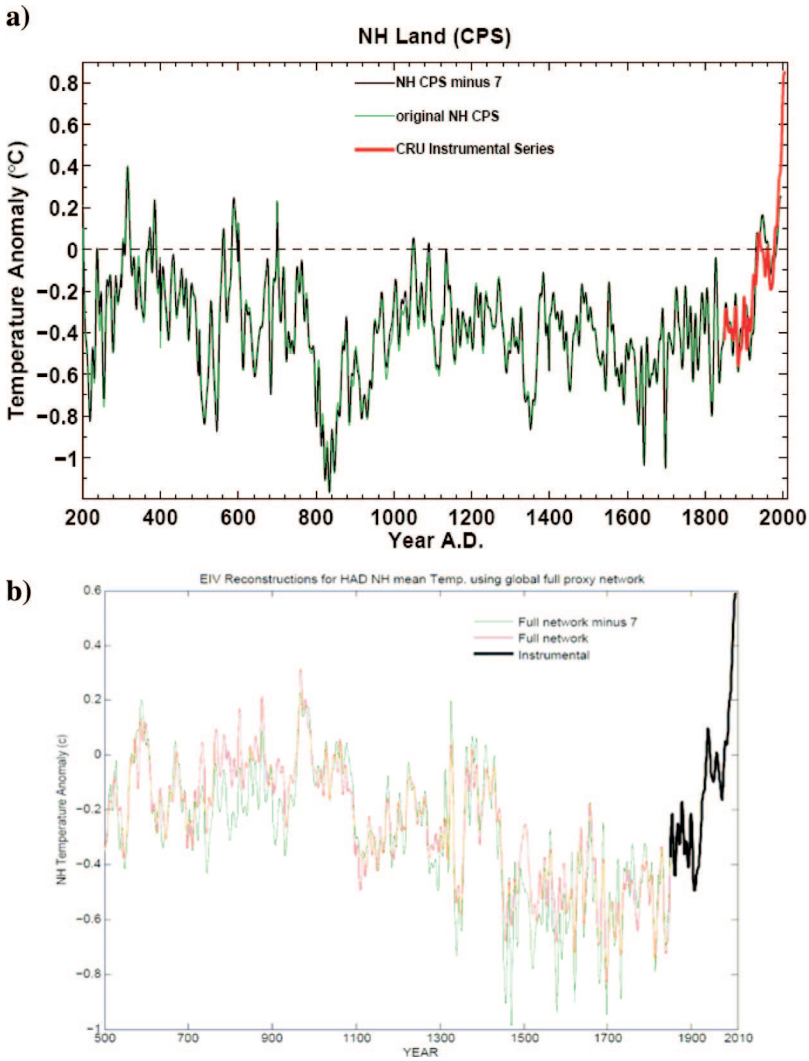


Fig. S8. Comparison of long-term CPS NH land (*a*) and EIV NH land plus ocean (*b*) reconstructions (full global proxy network) both with and without the seven potentially problematic series discussed.

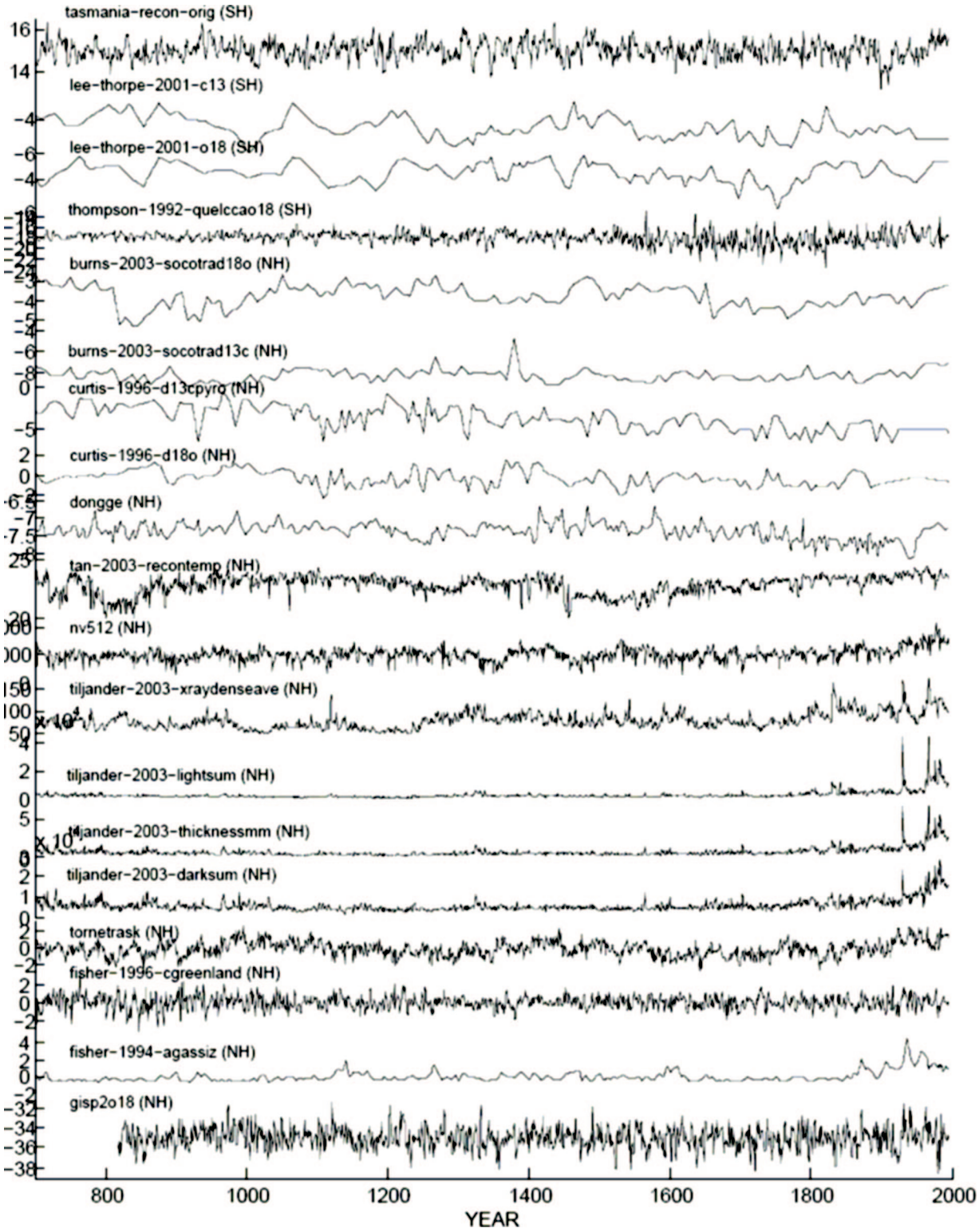


Fig. S9. Plots of the 19 (decadally smoothed) proxy records that pass the screening procedure back to at least A.D. 818, including four records from the Southern Hemisphere (labeled as SH) and 15 records from the Northern Hemisphere (labeled as NH).

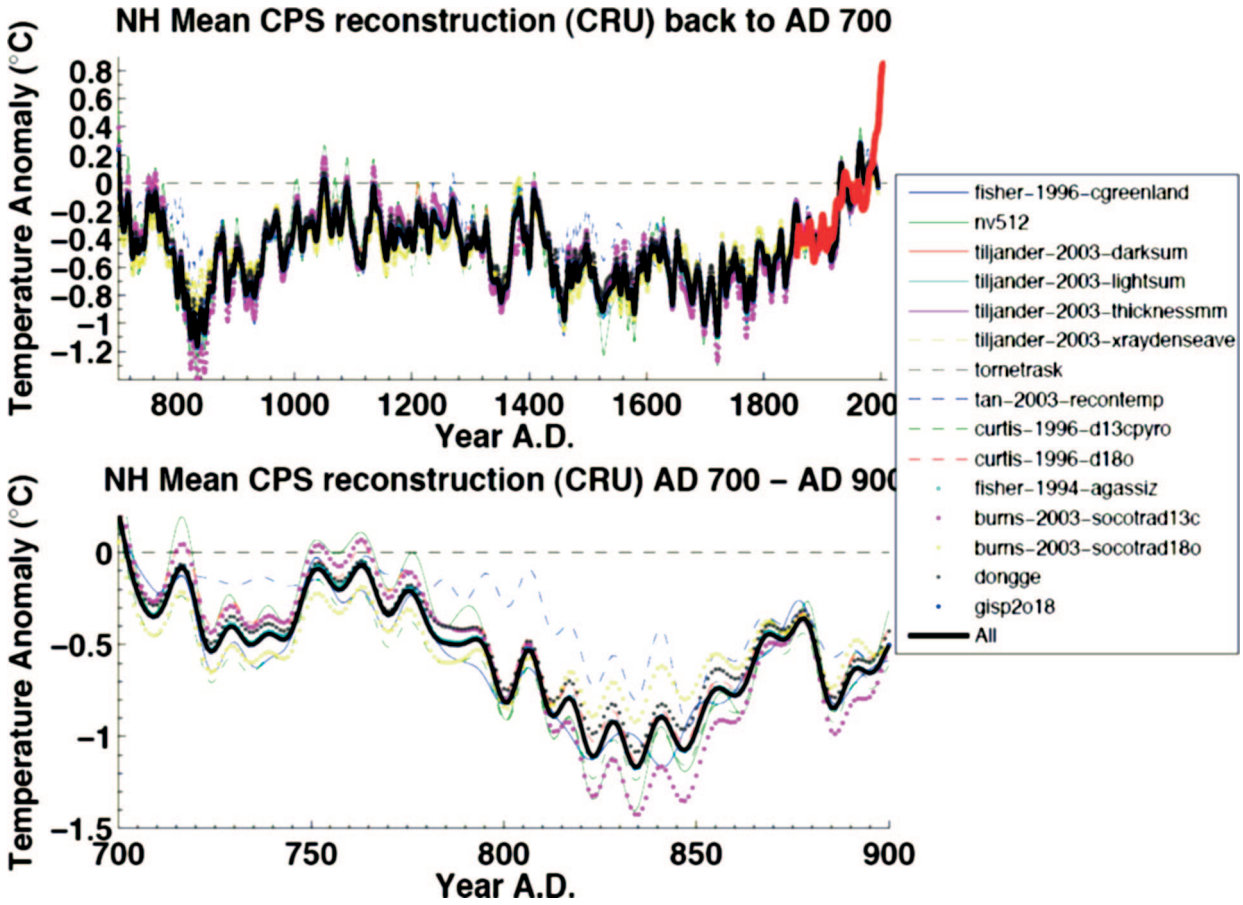


Fig. S10. Plots of the NH CPS land reconstruction withholding each of the 15 available Northern Hemisphere screened proxy records available back to the early 9th century as indicated.

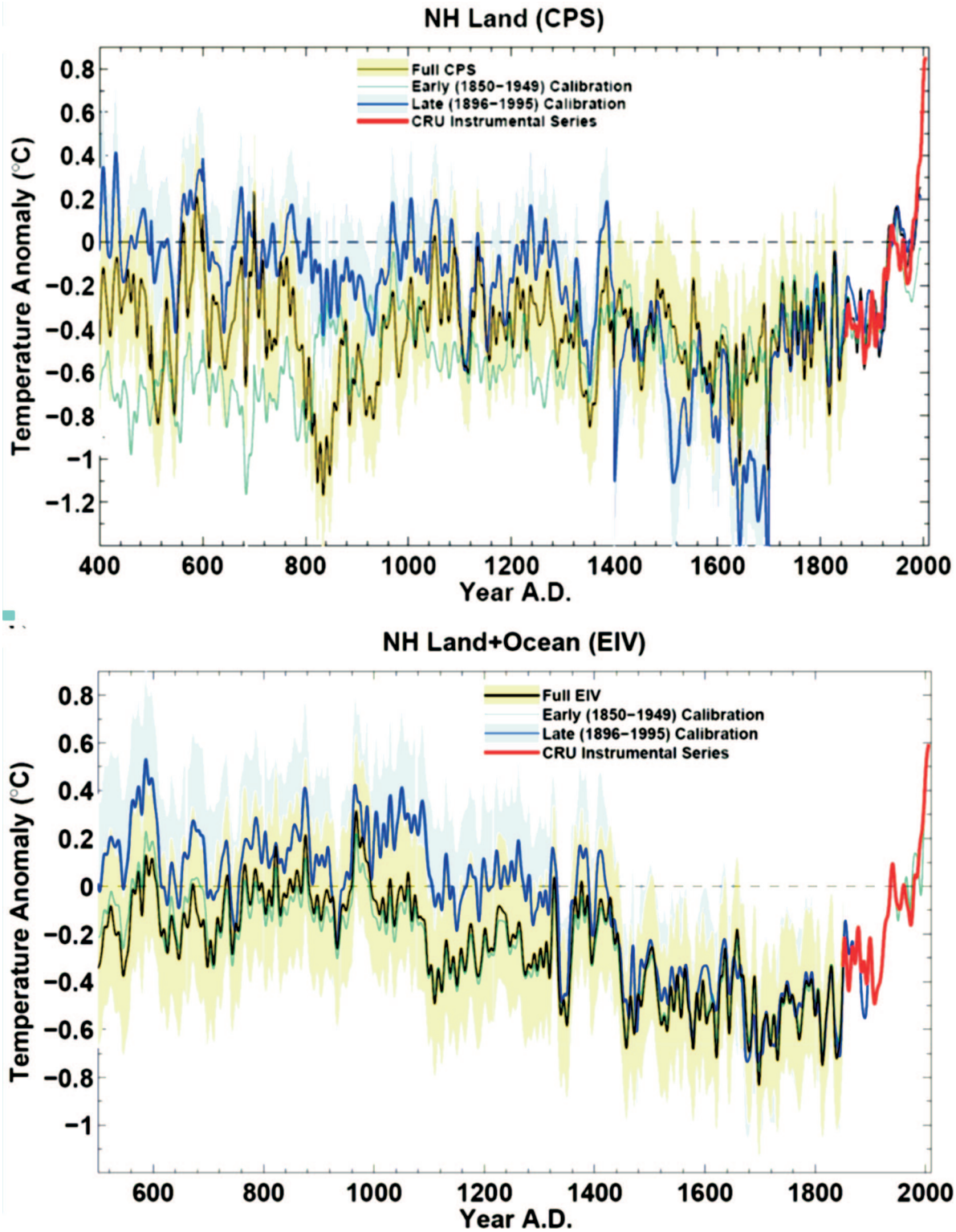


Fig. S11. Plots of the NH CPS Land reconstruction (*A*) and EIV Land plus ocean (full global proxy network) (*B*) that result from full period calibration, and early and late calibration intervals used in validation experiments.

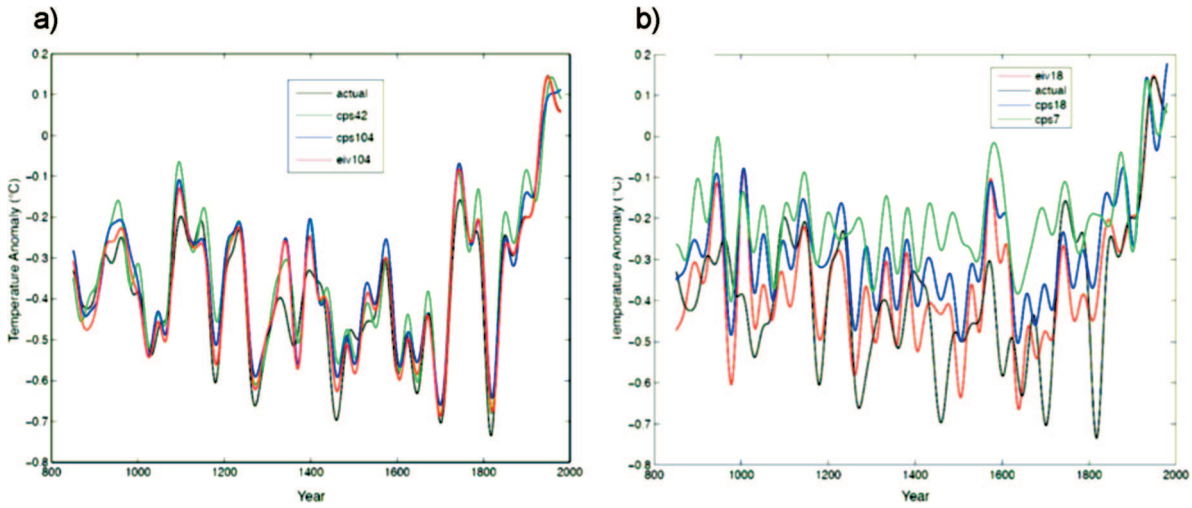


Fig. S12. Comparisons of CPS and EIV NH mean reconstructions using synthetic ''pseudoproxy'' networks calibrated over the 1856–1980 interval of a model simulation from A.D. 850 to A.D. 1999 using the NCAR CSM coupled model [see Mann *et al*. (16, 24)]. Results are shown for a large (104 pseudoproxy), moderate signal-to-noise ratio (SNR = 0.4) proxy network (*a*) and a smaller (18 pseudoproxy), low signal-to-noise ratio (SNR = 0.25) proxy network (*b*). The pseudoproxy networks are designed to have attributes similar to proxy networks used in the reconstructions described in the manuscript: Estimated SNR and network sizes for *a* and *b* are similar to those available for the actual proxy networks used in the present study back to A.D. 1300 and A.D. 0, respectively. Parallel CPS analyses are performed for networks by using a random subset of 40% of the pseudoproxy series (i.e., 42 of the 104 pseudoproxies for *a* and 7 of the 18 pseudoproxies for *b*) to reflect the smaller number (≈40% of total) of proxies typically available for the screened network required by CPS. All series smoothed to emphasize 40 year and longer time scales as in ref. 24.

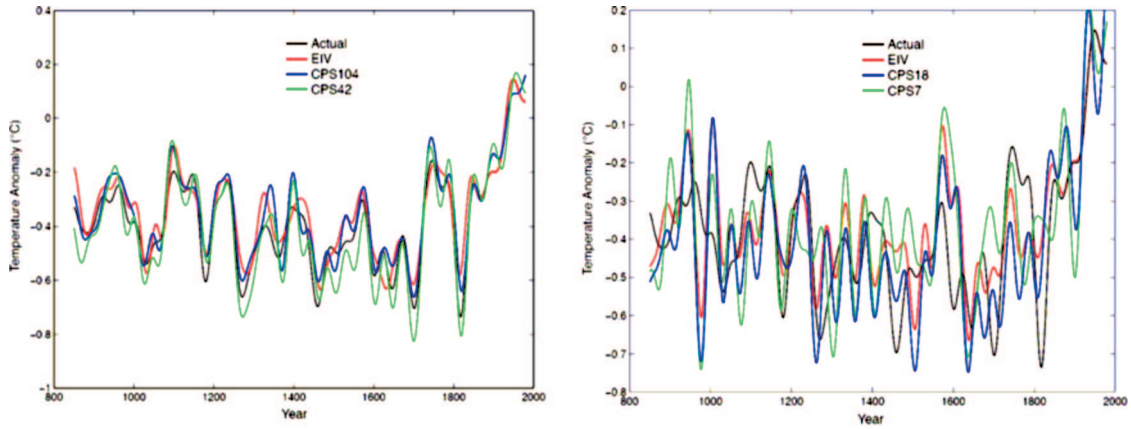


Fig. S13. As in Fig. S12, but where CPS scaling involves matching trend of composite to that for the target series over the (1856–1980) calibration interval.

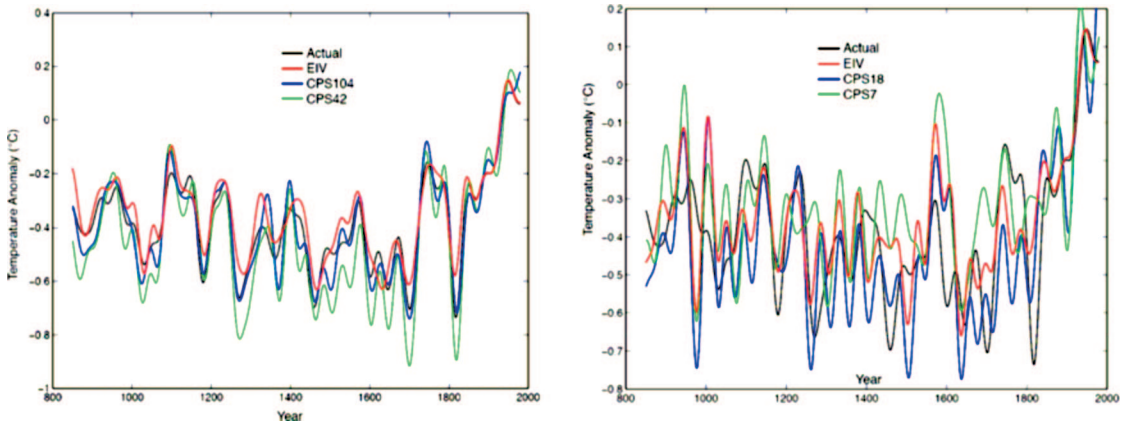
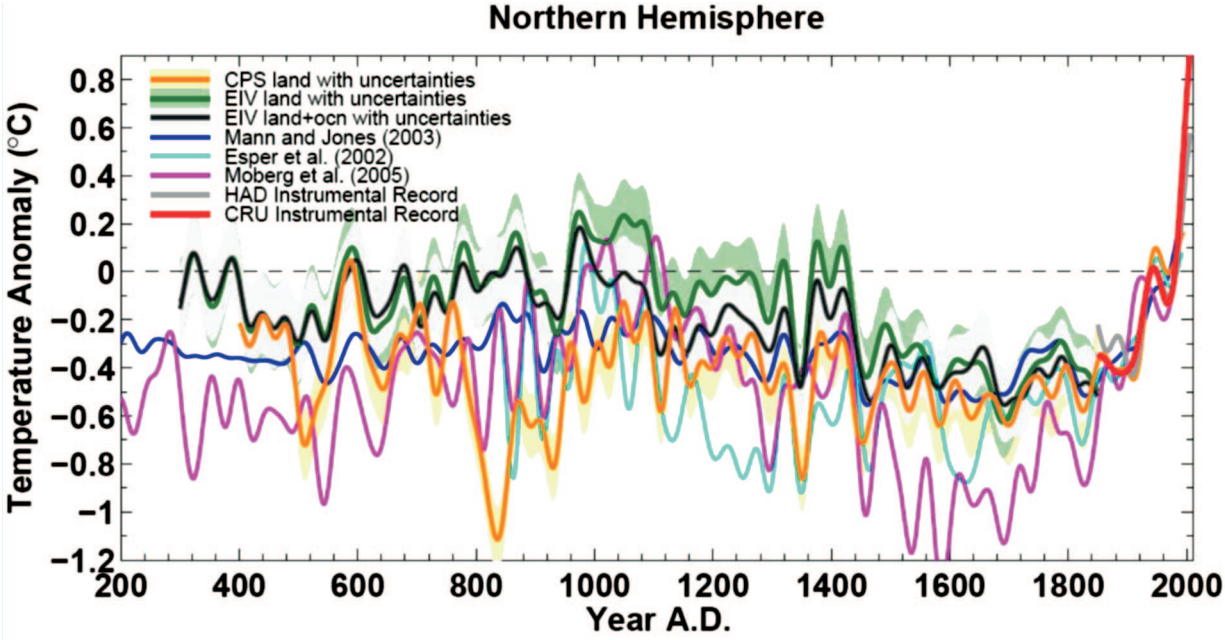


Fig. S14. As in Fig. S12, but where CPS scaling is based an alternative estimate of the ''signal amplitude'' of the reconstruction [if the calibration period correlation between the true NH mean temperature series and the proxy reconstruction is *r*, then the conventional (variance-matched) CPS reconstruction is rescaled by 1/*r* so that the estimated signal component of the record has the same decadal variance as the target series over the (1856–1980) calibration interval.



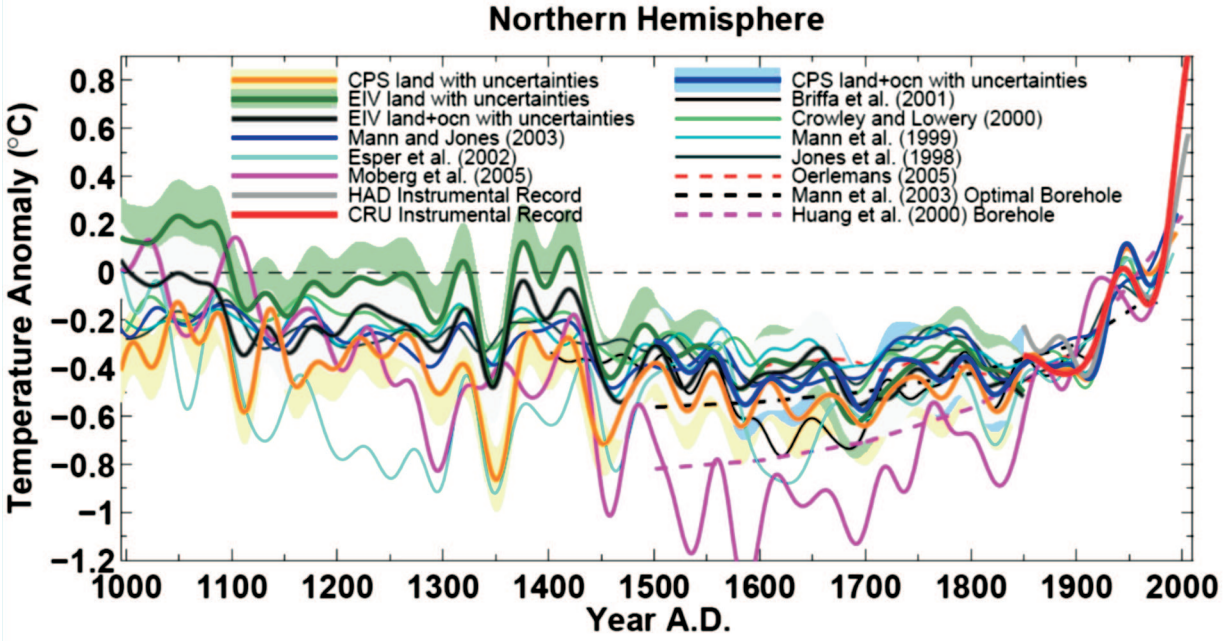


Fig. S15. As in Fig. 3 of the main paper, but where both the CPS reconstructions of this study and the other previously published reconstructions compared with (with the exception of the two borehole-based reconstructions Huang *et al.* (26) and Mann *et al.* (27)/Rutherford and Mann (28) have been scaled to have the same trend amplitude as the CRU NH land surface temperature record during the overlap interval (i.e., using trend matching discussed previously).

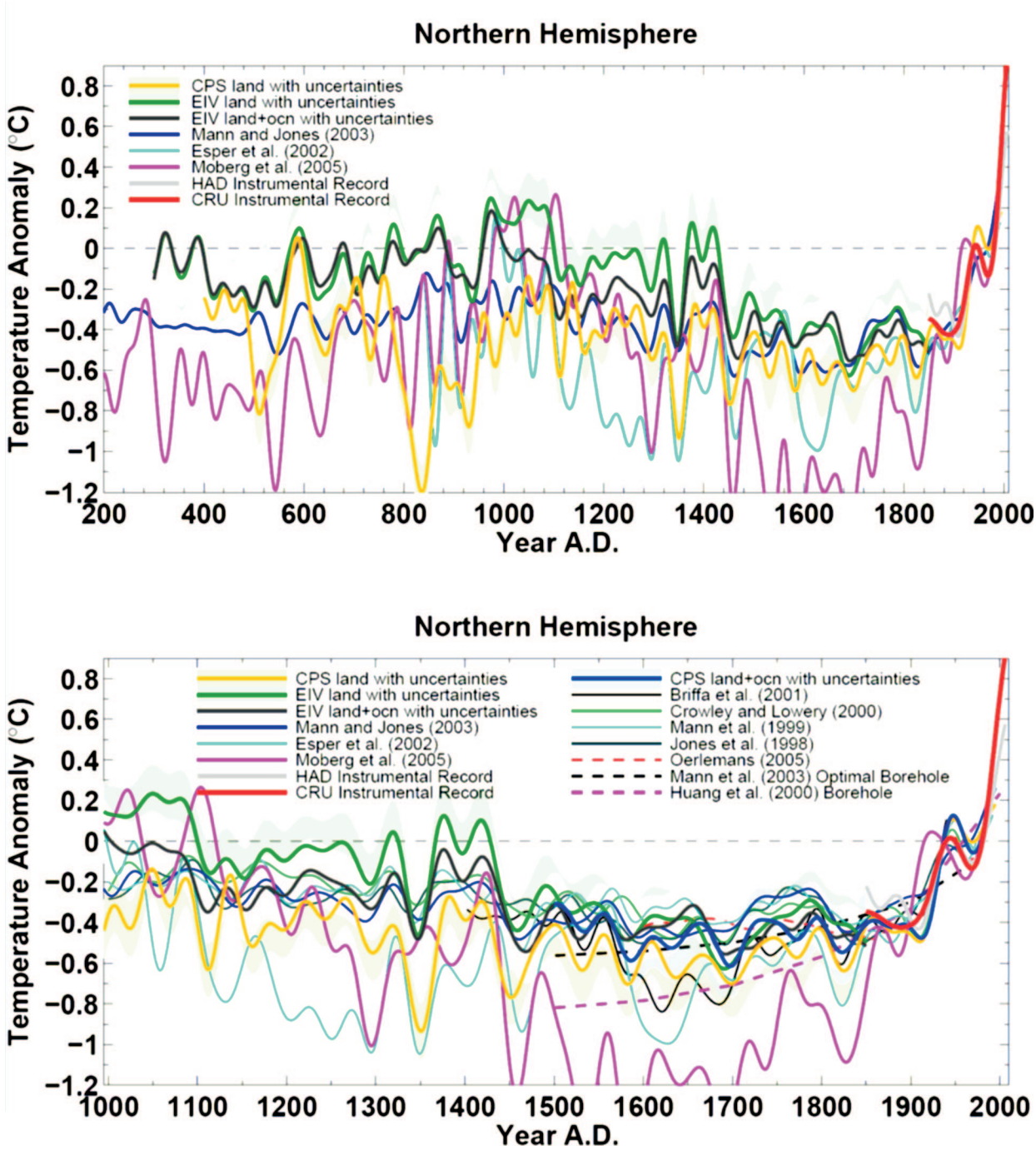


Fig. S16. As in Fig. 3 of the main paper, but where both the CPS reconstructions of this study and the other previously published reconstructions compared with (with the exception of the two borehole-based reconstructions Huang *et al.* (26) and Mann *et al.* (27)/Rutherford and Mann (28) have been scaled based on the estimated signal amplitude indicated from the estimated signal-to-noise ratio [if the calibration period correlation between the instrumental NH mean temperature series and the proxy reconstruction is *r*, than the conventional (variance-matched, i.e., as in Fig. 3 of the main paper) CPS reconstruction is rescaled by 1/*r* so that the estimated signal component of the record has the same decadal variance as the instrumental NH series] (i.e., using signal matching as discussed).

Table S1. Number of annually, decadally, and total proxy data available from different sources for different starting years for full proxy data set and screened proxy dataset

	Annually resolved proxies			Annually plus decadally resolved proxies		
Interval	Dendro	Others	Total	Dendro	Others	Total
			Full NH proxies			
1800–1855	889	104	993	889	147	1,036
1700–1799	563	95	658	563	137	700
1600–1699	276	91	367	276	132	408
1500–1599	150	89	239	150	127	277
1400–1499	105	18	123	105	46	151
1300–1399	62	16	78	62	40	102
1200–1299	41	12	53	41	36	77
1100–1199	29	12	41	29	34	63
1000–1099	18	12	30	18	28	46
900–999	15	10	25	15	25	40
800–899	13	9	22	13	24	37
700–799	11	8	19	11	23	34
600–699	11	8	19	11	22	33
500–599	7	7	14	7	21	28
400–499	7	7	14	7	21	28
300–399	6	7	13	6	21	27
200–299	4	6	10	4	20	24
100–199	3	6	9	3	20	23
0–99	2	6	8	2	19	21
			Full SH proxies			
1800–1855	143	22	165	143	30	173
1700–1799	78	12	90	78	17	95
1600–1699	39	9	48	39	13	52
1500–1599	22	8	30	22	12	34
1400–1499	14	8	22	14	12	26
1300–1399	8	8	16	8	12	20
1200–1299	6	7	13	6	11	17
1100–1199	5	7	12	5	11	16
1000–1099	4	7	11	4	9	13
900–999	2	7	9	2	9	11
800–899	2	6	8	2	7	9
700–799	2	6	8	2	7	9
600–699	1	6	7	1	7	8
500–599	1	6	7	1	7	8
400–499	1	4	5	1	4	5
300–399	1	4	5	1	4	5
200–299	0	4	4	0	4	4
100–199	0	4	4	0	4	4
0–99	0	4	4	0	4	4
			Full global proxies			
1800–1855	1,032	126	1,158	1,032	177	1,209
1700–1799	641	107	748	641	154	795
1600–1699	315	100	415	315	145	460
1500–1599	172	97	269	172	139	311
1400–1499	119	26	145	119	58	177
1300–1399	70	24	94	70	52	122
1200–1299	47	19	66	47	47	94
1100–1199	34	19	53	34	45	79
1000–1099	22	19	41	22	37	59
900–999	17	17	34	17	34	51
800–899	15	15	30	15	31	46
700–799	13	14	27	13	30	43
600–699	12	14	26	12	29	41
500–599	8	13	21	8	28	36
400–499	8	11	19	8	25	33
300–399	7	11	18	7	25	32
200–299	4	10	14	4	24	28
100–199	3	10	13	3	24	27
0–99	2	10	12	2	23	25
			Screened NH proxies			
1800–1855	305	98	403	305	115	420

	Annually resolved proxies			Annually plus decadally resolved proxies		
Interval	Dendro	Others	Total	Dendro	Others	Total
1700–1799	225	90	315	225	107	332
1600–1699	117	86	203	117	102	219
1500–1599	62	85	147	62	100	162
1400–1499	41	14	55	41	23	64
1300–1399	18	12	30	18	20	38
1200–1299	12	9	21	12	17	29
1100–1199	7	9	16	7	17	24
1000–1099	3	9	12	3	16	19
900–999	2	7	9	2	13	15
800–899	2	6	8	2	12	14
700–799	2	6	8	2	12	14
600–699	2	6	8	2	11	13
500–599	2	5	7	2	10	12
400–499	2	5	7	2	10	12
300–399	1	5	6	1	10	11
200–299	1	5	6	1	10	11
100–199	1	5	6	1	10	11
0–99	0	5	5	0	10	10
			Screened SH proxies			
1800–1855	50	14	64	50	14	64
1700–1799	31	9	40	31	9	40
1600–1699	13	7	20	13	7	20
1500–1599	8	6	14	8	6	14
1400–1499	5	6	11	5	6	11
1300–1399	2	6	8	2	6	8
1200–1299	1	5	6	1	5	6
1100–1199	1	5	6	1	5	6
1000–1099	1	5	6	1	5	6
900–999	0	5	5	0	5	5
800–899	0	4	4	0	4	4
700–799	0	4	4	0	4	4
600–699	0	4	4	0	4	4
500–599	0	4	4	0	4	4
400–499	0	3	3	0	3	3
300–399	0	3	3	0	3	3
200–299	0	3	3	0	3	3
100–199	0	3	3	0	3	3
0–99	0	3	3	0	3	3
			Screened global proxies			
1800–1855	355	112	467	355	129	484
1700–1799	256	99	355	256	116	372
1600–1699	130	93	223	130	109	239
1500–1599	70	91	161	70	106	176
1400–1499	46	20	66	46	29	75
1300–1399	20	18	38	20	26	46
1200–1299	13	14	27	13	22	35
1100–1199	8	14	22	8	22	30
1000–1099	4	14	18	4	21	25
900–999	2	12	14	2	18	20
800–899	2	10	12	2	16	18
700–799	2	10	12	2	16	18
600–699	2	10	12	2	15	17
500–599	2	9	11	2	14	16
400–499	2	8	10	2	13	15
300–399	1	8	9	1	13	14
200–299	1	8	9	1	13	14
100–199	1	8	9	1	13	14
0–99	0	8	8	0	13	13

Table S2. Decadal validation scores (based on full precalibration interval A.D. 850-1855) corresponding to NH reconstructions shown in Fig. S11 *A* and *B*

SNR = 0.4	*RE*	*CE*	r^2
CPS (42)	0.92	0.61	0.71
CPS (104)	0.94	0.70	0.76
EIV (104)	0.95	0.74	0.75
SNR = 0.25			
CPS (7)	0.55	−1.18	0.09
CPS (18)	0.68	−0.56	0.16
EIV (18)	0.76	−0.15	0.16

Provided are the standard *RE* and *CE* scores favored for skill evaluation. r^2 scores are provided for comparison, despite their demonstrated limited utility in skill evaluation [see Mann *et al.* (Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. *J Geophys Res* 112:D12109.) for skill metric definitions and other details].

Table S3. As Table S2 but where CPS scaling involves trend matching (see Fig. S13)

SNR = 0.4	*RE*	*CE*	r^2
CPS (42)	0.89	0.46	0.71
CPS (104)	0.94	0.69	0.76
SNR = 0.25			
CPS (7)	0.36	−2.11	0.09
CPS (18)	0.49	−1.47	0.16

Table S4. As Table S2, but where CPS scaling is based on estimated signal-to-noise ratio (see Fig. S14)

SNR = 0.4	*RE*	*CE*	r^2
CPS (42)	0.88	0.44	0.84
CPS (104)	0.92	0.63	0.87
SNR = 0.25			
CPS (7)	0.48	−1.55	0.29
CPS (18)	0.45	−1.67	0.40

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)
Dataset S4 (XLS)