www.sciencemag.org/cgi/content/full/326/5957/1256/DC1

Supporting Online Material for

Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly

Michael E. Mann,* Zhihua Zhang, Scott Rutherford, Raymond S. Bradley, Malcolm K. Hughes, Drew Shindell, Caspar Ammann, Greg Faluvegi, Fenbiao Ni

*To whom correspondence should be addressed. E-mail: mann@meteo.psu.edu

Published 27 November 2009, *Science* **326**, 1256 (2009)
DOI: 10.1126/science.1177303

This PDF file includes:

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/326/5957/1256/DC1)

SOM Data (computer codes and data, packaged as multiproxySpatial09.zip)

Supporting online material for Mann *et al* 'Global Signatures and Dynamical Origins of the "Little Ice Age" and "Medieval Climate Anomaly"'

Materials and Methods

Proxy Data. Details of the Mann et al (*S1*) proxy data set used are provided by Table S1 and Fig. S1. Dendroclimatic data included a tree ring network of 105 maximum latewood density ("MXD") gridbox (5° latitude by 5° longitude) tree-ring composite series (*S2-S4*), 926 tree-ring series from the International Tree Ring Data Bank (see ref S1 for further details), and 5 additional tree-ring based series (local temperature reconstructions and regional composite chronologies). The proxy dataset also includes (see ref. S1) 3 marine sediment series (from two 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 original Mann et al (S1) proxy dataset also included 71 European composite surface temperature reconstructions back to AD 1500 based on a composite of proxy, historical, and early instrumental data (S5). These data were not used in the present study, so that gridbox level assessments of skill would be entirely independent of information from the instrumental record.*

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).

Separate experiments were performed using a "screened" subset of the full proxy data set in which proxy records were screened for a local temperature signal based on their correlations with co-located instrumental data. These and other details, including sources, of the proxy data are provided in ref. *S1* [note: a recent correction was made to the details of the screening as described in ref. *S1*. Due to an "off-by-one" error in the degrees of freedom employed in the original screening that has been brought to our attention, the critical p values used for screening decadally-resolved proxy data are actually in the range p=0.11-0.12 rather than the nominal p=0.10 critical value cited. This brings the critical p value closer to the effective p value used for annually-resolved proxies (nominal value of p=0.10, but effective value actually closer to p=0.13 owing to the existence of significant serial correlation in many of the annual proxy data). It is worth noting that the precise thresholds used in the screening are subjective and therefore somewhat immaterial—our use of statistical validation exercises provides the best test of the reliability of any data screening exercises.

In this study, the use of the full "all proxy" data set is emphasized, as this yields considerably longer-term evidence of reconstruction skill. "Screened proxy" results are only provided for comparison.

All data used in this study are available in "SOM Data."

Instrumental Surface Temperature Data. Gaps in the individual annual mean (Jan-

Dec) gridbox surface temperature data available from 1850-2006 were infilled using the RegEM procedure with ridge regression described by Schneider (*S6*) [see also ref. *S7*]. A small number of grid boxes were not infilled due to the availability of too few monthly values in the raw data (see Fig. S2). The infilled gridbox data were averaged to yield the frozen grid "IHAD" and "ICRU" hemispheric global mean temperature series used in ref. *S1*. For this reason, comparisons of hemispheric and global means diagnosed from the spatial reconstructions of this study are made with the ICRU and IHAD reconstructions of ref. *S1*. Note that these yield slightly different reconstructions from the time-dependent sampling "CRU" and "HAD" estimates of Brohan et al (*S8*) also used in ref. *S1*.

RegEM CFR Procedure. The surface temperature field is reconstructed by calibrating the proxy network against the spatial information contained within the instrumental annual mean surface temperature field (*S8*) over a period of overlap between proxy and instrumental data (1850-1995) using the hybrid frequency-domain 'RegEM' CFR procedure described in ref. *S9* where high-frequency (f>0.05 cycles/year) and low-frequency (f<0.05 cycles/year) components of the reconstruction are calibrated separately.

As in ref. *S1*, reconstructions were produced at the decadal timescale, both using all available proxy data, and a subset that passes screening for possessing a local temperature signal (see "Proxy Data" subsection above). The number of surface temperature modes i.e. Empirical Orthogonal Function (EOF)/Principal Component (PC) pairs),retained in the analysis (M) and the low-frequency truncation parameter (K) in the RegEM procedure were determined using the criteria of ref. *S9* with minor modifications described below (see "Revised RegEM Selection Rules" subsection).

Reconstructions were performed as far back as a skillful global mean land+ocean ('IHAD') reconstruction was found possible in ref. *S1*. This corresponds to AD 500 for the "full" proxy network, and AD 1300 for the "screened proxy" network.

Following the procedure of ref. *S1* designed to avoid the introduction of redundant predictors, a forward 'stepwise' reconstruction approach was used in which the proxy network was only updated at each century step when use of the additional proxy data that become available led to improved validation scores (we based the validation skill metric on the multivariate Northern Hemisphere *RE* scores, owing to the relative dearth of gridboxes available in the Southern Hemisphere, and the generally low levels of skill for the Southern Hemisphere region—see "Validation Results" subsection of "Supporting Text" section below).

The skill of the resulting gridbox temperature reconstructions was evaluated by use of a calibration/validation approach, invoking a "red noise" null hypothesis (see "Validation Exercises" subsection below). Uncertainties were estimated from validation residuals (see "Uncertainty Estimation" subsection below). All data used and MATLAB codes for performing the various analyses described in the main article and Supplementary Online Material are available in "SOM Data."

Revised RegEM Selection Rules. Additional RegEM reconstruction experiments with synthetic "pseudoproxy" networks since ref. *S9* have led us to a further optimization of the selection rules. We have found that choosing the truncation parameter K for the low-frequency component of the calibration which collectively resolve 33% of the low-

frequency multivariate data variance (rather than 50% as used in ref. *S9*), leads to a consistent, and sometimes substantial improvement in multivariate skill scores (though only a very modest improvement in skill scores for hemispheric mean quantities).

We performed a set of "pseudoproxy" experiments parallel to those ref. *S9* for both simulations (GKSS and CSM) and all 3 spatial networks (networks "A" of 104 pseudoproxies, network "B" of 18 pseudoproxies, and network "D" of 208 pseudoproxies) used in ref. *S9*, to investigate whether more conservative objective criteria might lead to slight improvements in skill scores. We defined skill in terms of the preferred (*S1,S9*; see "Validation Exercises" subjection below) validation metrics *RE* and *CE*, though r^2 was examined for completeness as in ref. S9. First, we allowed the # of surface temperature modes retained (*M*) to be reduced by 1 (i.e. *M-1*) compared to what is dictated by the criteria of ref. *S9* This led to a degradation in skill scores in nearly all experiments (Table S2) relative to the original criterion (*M*) of ref *S9*. This potential alternative choice was thus rejected. Next, we allowed the criterion for retained variance in the low-frequency TTLS procedure to be decreased relative to the 50% criterion used to choose the low-frequency truncation parameter *K* in the RegEM algorithm. Alternative choices considered were 15%, 25%, 33%, and 40% retained variance. We found that a 33% retained variance criterion on average yielded consistently optimal skill for both model simulations and using all three pseudoproxy networks. Skill for hemispheric mean reconstructions was only modestly improved (usually at most 1-2%) relative to the 50% criterion used in ref. *S9* [and in the 'EIV' hemispheric and global reconstructions described by ref. *S1*]. The resulting hemispheric reconstructions differ only slightly (Fig. S3). Multivariate skill scores, however, were considerably improved, with multivariate *RE* scores typically improving by 4% to 8% and multivariate *CE* scores typically improving from 6% to 11% depending on the simulation and proxy network.

We have accordingly adopted this revised selection rule in the reconstructions developed in the current study.

Validation Exercises. Split calibration/validation experiments were used in ref. *S1*, wherein the early and late half of the full available instrumental calibration interval (1850-1995) are alternatively used for statistical calibration and validation. Such experiments are possible when reconstructing only hemispheric and global mean series, since spatial gaps in the instrumental data (which grow increasingly large for the earlier data) have a minor influence on the target series during either early or late half intervals. Early calibration experiments are more problematic, however, in a spatial context since the EOFs of the instrumental surface temperature field are subject to significant bias if defined on the earlier sparse data. For this reason, we favored early validation experiments for skill evaluation. Validation experiments were performed for decadally-smoothed data as in ref.s *S1* and *S9*. Validation scores were evaluated at the gridbox level, and integrated 'multivariate' spatial skill scores were defined by integrating gridbox scores over the full domain or sub-domains (e.g. Northern Hemisphere and Southern Hemisphere). Gridbox skill scores were also calculated for global and hemispheric means, and were appropriately averaged over the appropriate gridboxes to estimate skill scores for regional averages and indices (see "Regional averages and Indices" subsection of "Supporting Text" section below).

A given gridbox was defined as being skillfully reconstructed if and only if it passed

validation at the p=0.05 level for both *RE* and *CE* skill metrics. We adopted the additional requirement that *RE*>0. Note that spatially-infilled values in the instrumental data (see 'Instrumental Surface Temperature Data' sub-section above) were not used in the calculation of validation scores. In addition, validation r^2 was used in the uncertainty estimation procedure (*S1,S9*), though this latter metric is not favored as a skill diagnostic for reasons discussed previously (see ref.s *S1* and *S9* for definitions and further discussion of the relative merits of these different validation metrics).

Only gridboxes passing validation over a given time interval were used in the calculation of regional spatial means. For defining 'skillful regions' in the 300 year LIA and MCA interval composites described in the main article with respect to a particular (*RE* or *CE*) skill measure, we required that a gridbox pass validation with respect to that skill measure at the p=0.05 level for at least a century-long sub-period of the 300 year interval.

Significance was measured relative to a red noise null hypothesis [based on an AR(1) fit to the gridbox surface temperature series over the calibration period, i.e., using the same Monte Carlo procedure employed in ref. *S1*, but at the gridbox level].

Uncertainty Estimation. Uncertainties for gridbox temperature reconstructions were calculated from the gridbox validation residuals, as in ref. *S1* for hemispheric and global mean reconstructions. Specifically, we define the 1 sigma decadal standard error σ_S in the reconstruction is defined by $\sigma_S^2=(1-r^2)\sigma^2$ where r is the validation period correlation coefficient, and $)\sigma^2$ is the decadal variance in the target series during the calibration interval. This definition in general yields more liberal (i.e. larger) estimate of uncertainty than the use of calibration *RE* scores as in previous work (*S10,SI1*) and has been shown in pseudoproxy experiments (*S9*) to reflect an accurate estimate of the intrinsic uncertainty in the reconstructions. As these uncertainties are determined solely from the cross-validation experiments, they do not include the potential additional component of uncertainty that might arise from the degradation of proxy data prior to the calibration/validation intervals, wherein the suitability and skill of proxy records are assessed. Thus, as noted previously in ref. *S1*, the assessed uncertainties should be considered minimum uncertainties, and in reality could be larger.

Supporting Text

Parameter Values Used in RegEM Reconstructions. Parameter values were estimated from the proxy and instrumental data using the selection rules discussed in the previous section, including the revised rule for selecting the low-frequency TTLS truncation parameter *K* described and tested in the previous section. The resulting RegEM parameter values are given in Table S3 for the various intervals of the reconstruction.

These parameter values indicate multiple spatial degrees of freedom in the reconstructions back through AD 500, contrasting with earlier proxy-based large-scale surface temperature reconstructions (*S10-11*) where only a single spatial degree of freedom was resolved prior to AD 1450. The choice of low-frequency TTLS truncation parameter *K* determines the actual number of statistical degrees of freedom in the reconstructions at interdecadal (f<0.05 cycles/year) timescales, while *M* determines the number of spatial temperature modes (i.e. EOF/PC pairs) used in both the low-frequency and high-frequency calibration of the hybrid reconstruction procedure (see ref. *S9* for further details). The number of low-frequency degrees of freedom decreases back in time

as seen in Table S3, but it is always greater than one, indicating that there are multiple low-frequency degrees of freedom resolved in the reconstructions over the full interval. Back to AD 1600, our selection rules yield K=3 and M=6 which indicates that three degrees of freedom are available at low-frequencies to reconstruct the time histories of six spatial surface temperature patterns. The low-frequency time histories of the M=6 reconstructed patterns (i.e. the reconstructed low-frequency PC series) are thus not orthogonal [note that by contrast, the raw corresponding instrumental PC series themselves are by construction orthogonal at the annual—though not necessarily *interdecadal*---timescale over the calibration interval]. Rather, each low-frequency PC series represents a linear combination of the K=3 independent patterns of temporal evolution. Prior to AD 1600, our selection rules instead yield K=2 and M between 2 and 5 (depending on how far back). Thus, prior to AD 1600, the low-frequency time histories of the reconstructed patterns are represented as linear combinations of only K=2 statistically independent patterns of temporal evolution.

Some caution is therefore required in interpreting the details of the reconstructions, particularly prior to AD 1600. As noted in the main manuscript, the low value of the selected truncation parameter K may lead to greater apparent levels of similarity between regions (i.e fewer spatial degrees of freedom) than exists in the true underlying spatial temperature pattern, and likely leads to the various climate indices such as the PDO and AMO series exhibiting an artificially high level of similarity than their true underlying counterparts, particularly prior to AD 1600.

Validation Results. Spatial patterns of gridbox *RE* and *CE* decadal skill scores for various time intervals are provided in Fig. S4. *RE* and *CE* skill scores and associated statistical significance levels from Monte Carlo simulations are summarized in Tables S4 and S5. Multivariate scores for the full surface temperature field and scores for hemispheric/global mean series are provided in Table S4, while scores for regional averages and derived climate indices (see "Regional averages and indices" subsection below) are provided in Table S5. For the 'multivariate' (full field) skill scores, an integrated score was assessed over all gridboxes for which a reconstruction is attempted. For spatial means and indices, the skill scores reflect averages only of those gridboxes used in the calculation of the particular spatial mean or index (i.e. only gridboxes passing validation). For the multivariate full field scores, we provide for comparison so-called "PC-filtered" skill scores (Table S4). These scores indicate the level of skill achieved in the reconstruction of the field in the case where the actual, instrumental PC series, rather than proxy reconstructions thereof, are used to reconstruct the spatial field over the validation interval. These scores can be thought of as 'perfect reconstruction' scores, in the sense that they describe the maximum amount of variance that one could hope to resolve in the full field if the restricted set of PCs used to perform the reconstruction were reconstructed 'perfectly' by the proxy data. In other words, they reflect the best possible reconstruction that could be achieved given just the filtering of the raw data that is imposed by retaining a restricted set of empirical eigenvectors of the full field. These scores thus provide useful upper-limit skill estimates against which the skill scores obtained for the actual proxy reconstructions should be compared.

Examining the spatial pattern of skill scores (Fig. S4) we see that there is in general broad skill over much of the Northern Hemisphere and tropics for most intervals (more so for the "full proxy" reconstructions than for the "screened proxy" reconstructions. We

observe less evidence for skill in the southern hemisphere, particularly in the extratropics. As expected, the extent of "no skill" regions increases in general back in time, as the proxy networks become more sparse. The observations are consistent with the summary statistics reported in Tables S4 and S5. The lower levels of skill in the southern hemisphere is reproduced in the synthetic pseudoproxy experiments of ref. S9, and are interpreted as a clear consequence of the dearth of proxy data in the southern hemisphere (in particular, the absence of proxy records from the southern ocean). Our focus and interpretation of features in the climate reconstructions is closely guided by the validation exercises results. For example, we have in general avoided placing any emphasis on features in the southern hemisphere where spatial skill scores are typically low.

A few additional points are worthy of note. Firstly, we note that the hemispheric and global mean skill scores are modestly lower than those reported in ref. *S1* for hemispheric and global mean reconstructions using a simpler "EIV" implementation of RegEM where a single hemispheric or global mean series, rather than the underlying spatial field, was reconstructed from the same proxy data set. This result is unsurprising, as the RegEM CFR procedure used here is designed to optimize a spatial pattern rather than a single series or index. The differences between the hemispheric and global mean series are minor, and the main differences (see Fig. 1 of main article and also Fig S5) arise after AD 1500 where the predictor data sets differ in the two studies, as the historical records of ref. S1 were not used in this study (see "Proxy Data" subsection of "Materials and Methods" section above). When spatially averaged over smaller than hemispheric domains (e.g. Northern Hemisphere land only) the differences between the EIV and CFR-based reconstructions are somewhat more substantial. For example, the CFR reconstruction of Northern Hemisphere land region averaged temperature runs systematically a bit cooler in past centuries than its EIV-based counterpart, though the two estimates are well within estimated uncertainties (see "Regional averages and indices" subsection below, and in particular Fig S5).

It is also important to note that the decadal multivariate/spatial skill scores are invariably substantially lower than the hemispheric and global mean scores, with a tendency, for example, for negative albeit statistically significant (relative to red noise null hypothesis) *CE* scores in many cases. This observation is in fact a very basic feature of proxy-based CFR that is thoroughly captured in experiments with synthetic 'pseudoproxy' data possessing attributes similar to those estimated for actual proxy data networks.(i.e. Table S2, Fig. S3, and ref. *S9*) Indeed, negative spatial *CE* scores are obtained *even* in the "perfect proxy" analyses discussed above, as the number of retained EOF/PCs is reduced to a handful (2 or 3). These findings are explained by the large nature of sampling fluctuations that are present over a short validation interval. In ref. *S9*, it is shown that over much longer validation intervals (which are not available in real-world reconstructions, but are of course available in the synthetic world of a climate reconstruction), positive scores are indeed typically obtained (indicating by definition true reconstructive skill) even when negative *CE* scores are found for the same reconstruction using much shorter validation intervals. It is precisely for these reasons that negative decadal multivariate *CE* scores obtained over short validation intervals are in many cases found to be skillful based on their estimated significance relative to a red noise null hypothesis, and should not be dismissed on the simple basis of a negative short

validation *CE* score.

Regional averages and indices. A variety of regional indices were calculated by averaging the reconstructed spatial patterns over various key regions. As described in the "Validation Exercises" subsection above, only gridboxes passing validation over a particular time interval were used in calculating the regional averages. The regional averages and indices were defined as follows (bold type indicates series that was already shown in Fig. 1 of main article)

A) Global SST (all global SST gridboxes).
B) Tropical: 27.5S-27.5N, 177.5W-177.5E
C) Extratropical: 27.5N-67.5N, 177.5W-177.5E
D) Northern hemisphere land: land gridboxes north of equator
E) Nino3: SST gridboxes over 2.5S-2.5N, 92.5W-147.5W
F) PDO: SST gridboxes over 22.5N-57.5N, 152.5E-132.5W
G) AMO: Atlantic SST gridboxes in Northern Hemisphere domain
H) MDR (Atlantic main development region for tropical cyclones): SST gridboxes over 7.5N-17.5N, 17.5W-62.5W
The various regional averages and indices as well as hemispheric and global series not already shown in the main article, are shown in Fig. S5, along with estimated 95% uncertainties and modern instrumental series.

Proxy Weights. The weights placed on different proxy records by RegEM can be determined by the regression coefficients for each proxy record used obtained at convergence (i.e. at the final iteration of the RegEM algorithm). In the hybrid version of RegEM used here, low-frequency and high-frequency components are estimated separately, as are consequently the weights on particular proxy records. Of primary interest to our analysis are the weights obtained for the low-frequency component of the reconstruction. Weights can be estimated separately for each of the *M* reconstructed PCs, and for the reconstruction overall, through the appropriate (i.e. eigenvalues-weighted) sum over the weights for each of the *M* low-frequency PC reconstructions. The low-frequency weights for both the individual PCs and overall reconstruction for each interval of the reconstruction are shown in Fig S6. The low-frequency weights for the overall reconstruction are averaged over the appropriate intervals to yield the MCA and LIA weighting patterns shown in Figure 2 of the main article.

EOFs and PC reconstructions. It is instructive to look at the behavior of the leading surface temperature modes and associated PC series during the modern instrumental interval and as reconstructed from the proxy data (Fig. S7), the first 3 of which The leading mode is characterized by broadly positive loadings over the global domain, and carries much of the global mean temperature signal. The modern increase exceeds the envelope of past variation spanning the past 1500 years. By contrast, the 2nd and 3rd modes exhibit a more heterogeneous spatial structure and carry much of the ENSO surface temperature pattern. For PC2, the modern trend remains within the envelope of past variability, while PC3 shows a steady long-term negative trend. Positive values of these 2 PCs during the Medieval era multiply negative loadings in the eastern and central tropical Pacific, to produce the La Nina-like MCA pattern discussed in the main article. PC 4 is reconstructed considerably less far back (AD 1400—see Table S3), and displays an even more heterogeneous spatial structure including a dipolar SST pattern in the North

Atlantic. The PC pattern displays pronounced multidecadal variability which projects onto the multidecadal AMO variability discussed in the main article, though this variability is less pronounced prior to vs. during instrumental era.

Sensitivity Tests. Additional basic tests were performed to evaluate the robustness of key features of the reconstruction with respect to data used. This includes a similar test to that shown in ref. *S1* investigating the robustness of the RegEM EIV reconstruction of northern hemisphere mean temperature to the exclusion of particular types of data. In addition to the tests described by ref. *S1* which removed alternatively (a) all tree-ring data or (b) 7 additional long-term proxy records associated with greater uncertainties or potential documented biases (showing the temperature reconstruction was robust to removal of either of these datasets), we here removed *both* data sets simultaneously from the predictor network (Fig. S8). This additional test reveals that with the resulting extremely sparse proxy network in earlier centuries, a skillful reconstruction is no longer possible prior to AD 1500. Nonetheless, even in this case, the resulting (unskillful) early reconstruction remains almost entirely within the estimated error bounds of the original reconstruction.

We also investigated the effect of alternative definitions of MCA and LIA intervals. Based on Fig. 1 of the main article, one could arguably choose slightly different multi-century intervals that characterize the main transition from relatively moderate to cool NH mean temperatures, specifically AD 1600-1850 for the LIA (instead of AD 1400-1700) and AD 900-1100 for the MCA (rather than AD 950-1250). This alternative analysis yields essentially the same basic LIA and MCA patterns as shown in the main article (Fig. S9). To investigate the robustness of the key features of the reconstructed pattern to the richness of the available proxy network, we performed a test in which the composite LIA reconstruction of AD 1400-1700 was performed not with the proxy data available during those interval, but instead with the proxy data available during entirety of the early MCA period (i.e. the network available back to AD 900—see Table S3). This test (Fig. S10) reveals the LIA pattern to be robust to whether the richer LIA or sparser MCA proxy network is used, suggesting that our proxy network is indeed adequate to capture the key spatial features of the reconstructions during the earlier MCA interval.

Details of GISS-ER model analysis. We performed 6 climate simulations with the coupled atmosphere-ocean climate model GISS-ER (*S12*) Following a control run to establish stable initial conditions, six transient runs extending from 850 to 1900 C.E. were performed. Solar forcing was applied across the ultraviolet, visible and infrared spectrum based on scaling by wavelength versus total irradiance as seen in modern satellite data, while the total irradiance through time was based on the time series of Bard et al (*S13*) derived from South Pole ice core ^{10}Be data and taking into account a small long-term geomagnetic modulation (*S14*) and a polar enhancement factor. The amplitude is scaled to give a top-of-the-atmosphere forcing of 1.1 W/m2 between the Maunder Minimum and the late 20th century (*S15*-16), with a second ensemble using twice that amplitude to test sensitivity. The model also includes the ozone response to solar irradiance variations, which is parameterized from the results of prior GISS modeling using a full atmospheric chemistry simulation (*S15*) for computational efficiency.

Individual runs and the ensemble mean results were analyzed using time periods within (or very near) the MCA and LIA periods analyzed in the reconstruction. Times were

chosen to maximize the solar forcing (1125-1275 for the MCA, 1650-1750 for the LIA), and results from the 2.2 W/m2 top-of-the-atmosphere forcing ensemble were used to maximize the response. As the magnitude of historical forcing is quite uncertain, results were scaled to be equivalent to that used in the NCAR CSM 1.4 simulation also analyzed (*S17*) in the paper for ease of comparison.

Variance across the six ensemble members is large, so that mid-latitude results are typically only marginally statistically significant for individual grid boxes, though tropical and high latitude results are significant. Even at mid-latitudes, however, the NAO/AO-like anomalies discussed in the text occur in all of the runs (though with shifted locations), so that the large-scale circulation changes are significant (see Figure S11).

Supporting figures

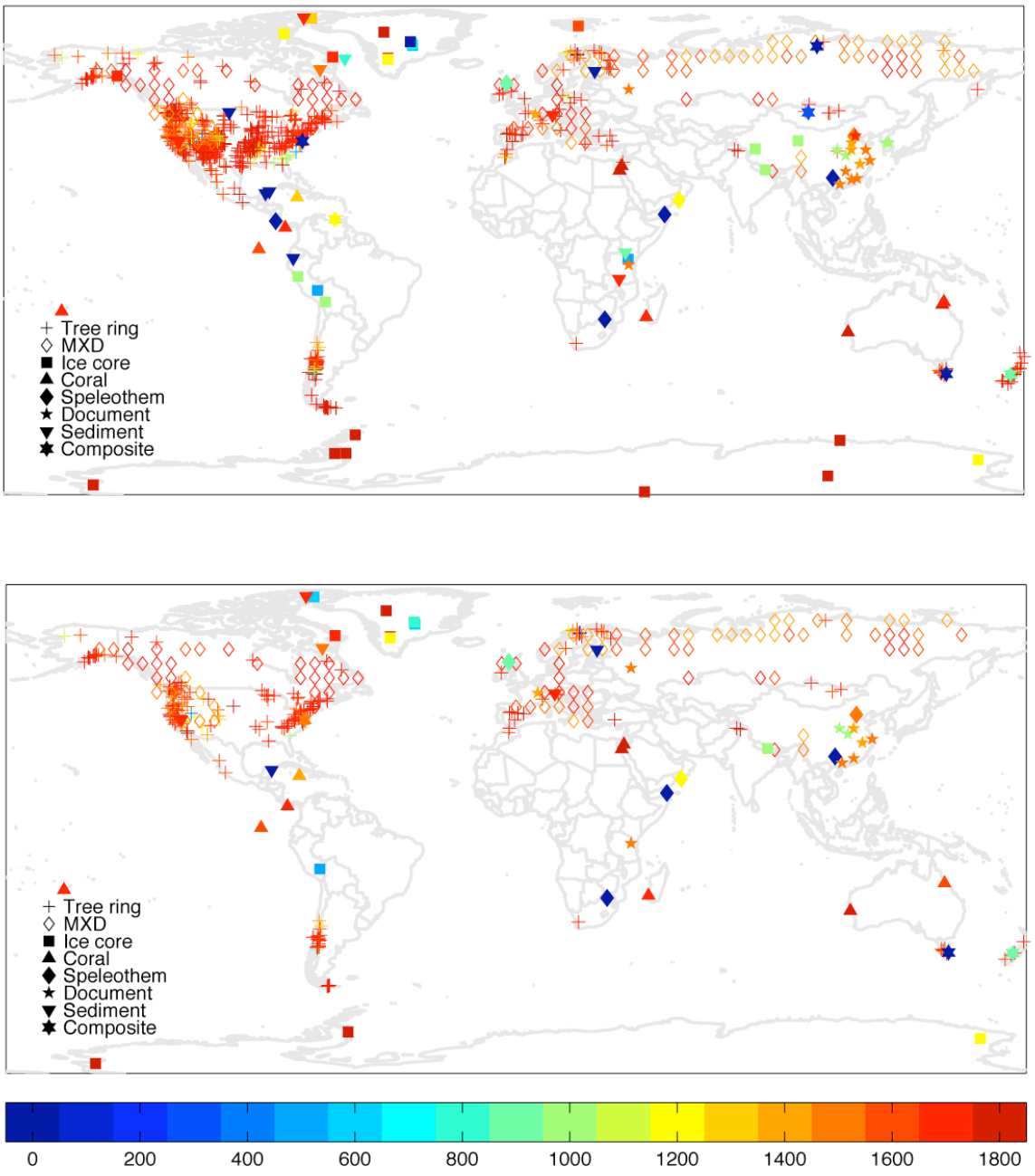


Figure S1: Spatial distribution of (a) full and (b) screened (based on 1850-1995 interval) proxy database (see also Table S1**).** Nine different proxy types are denoted with different symbols as shown in the map. Beginning dates of proxy records are represented by color scale shown.

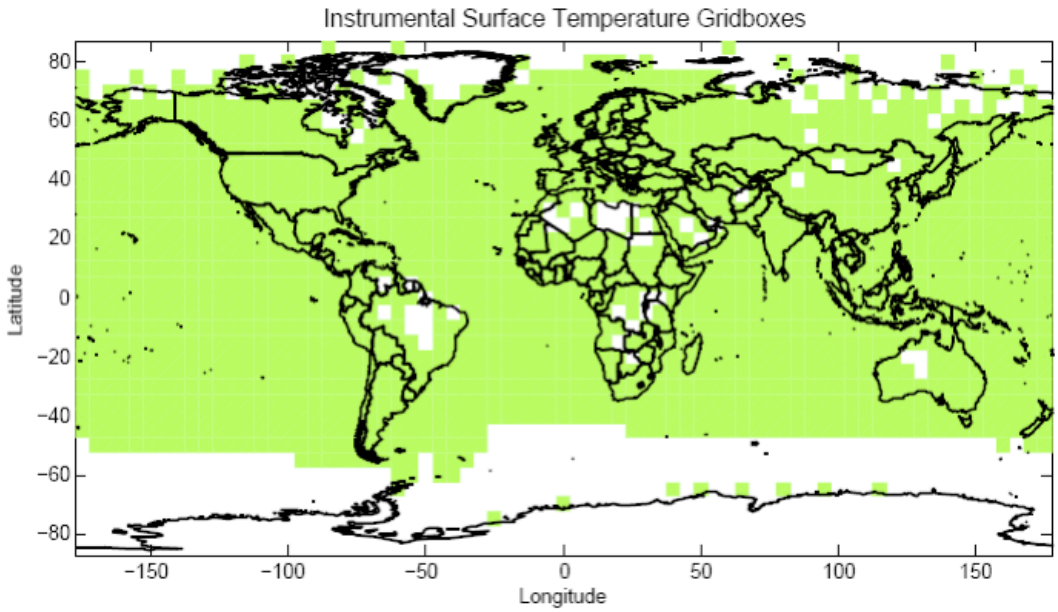


Figure S2: Distribution of infilled instrumental surface temperature gridboxes used in study.

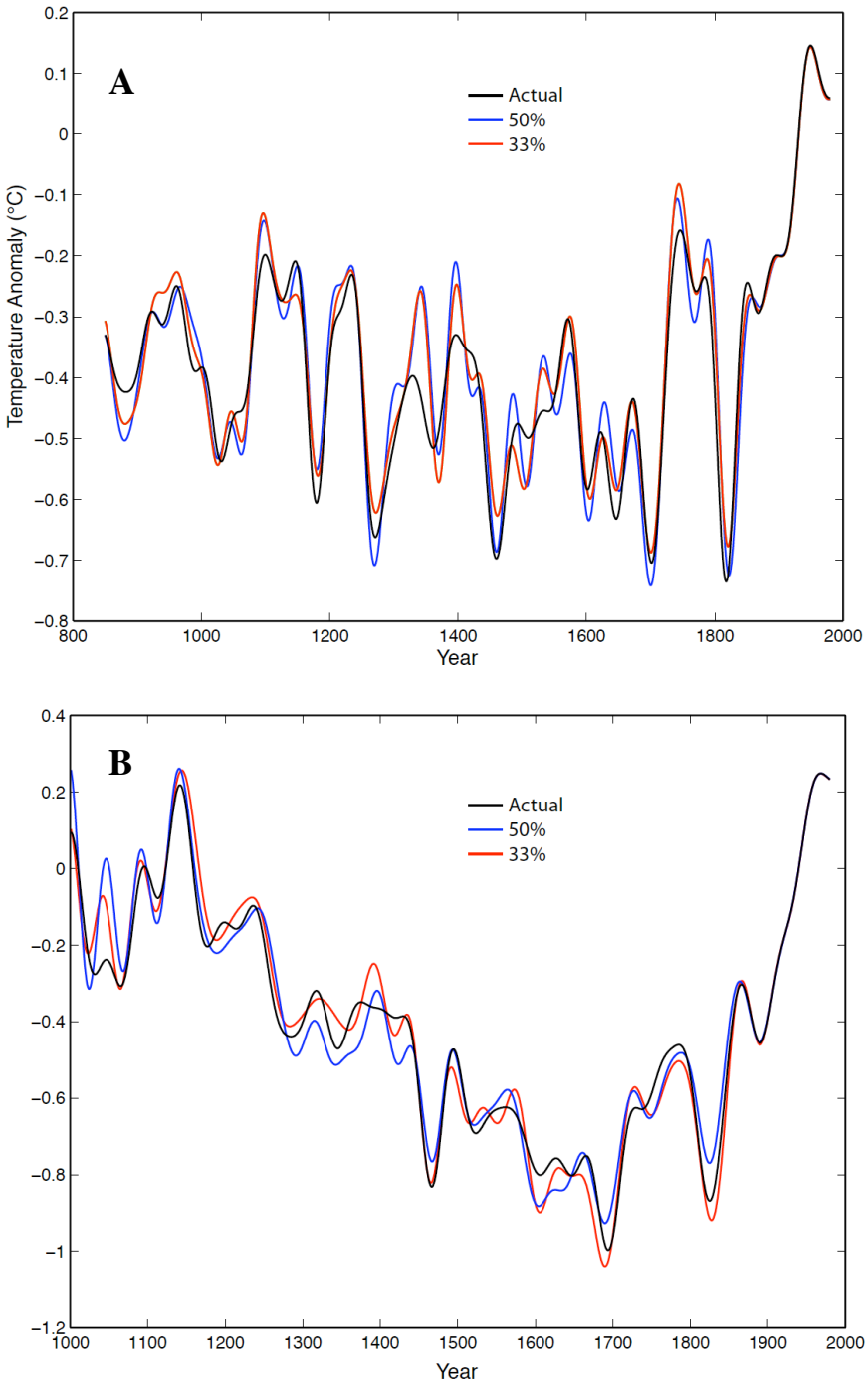
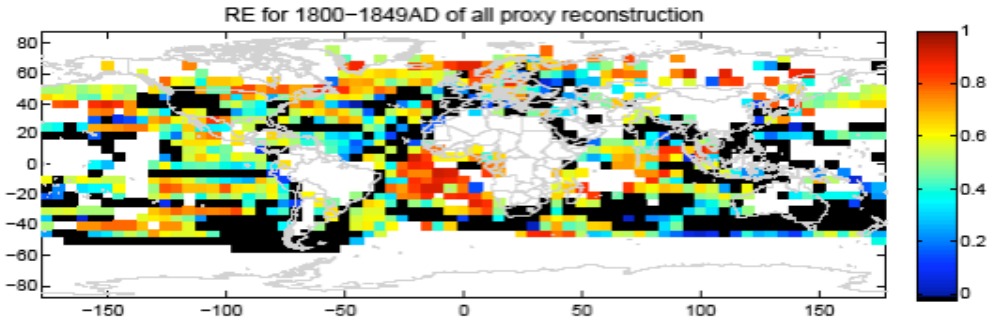
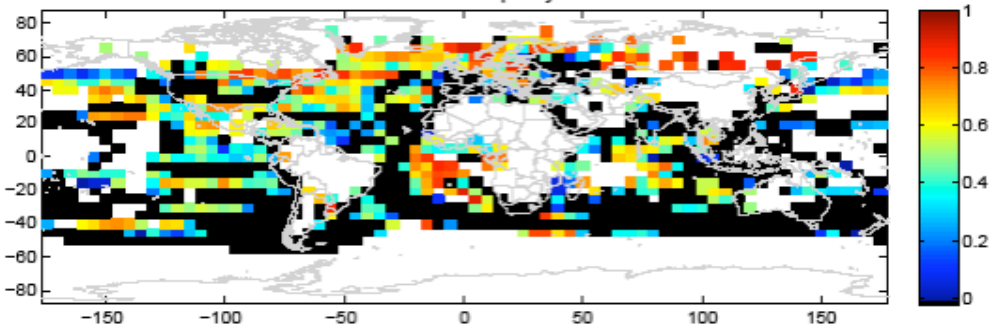


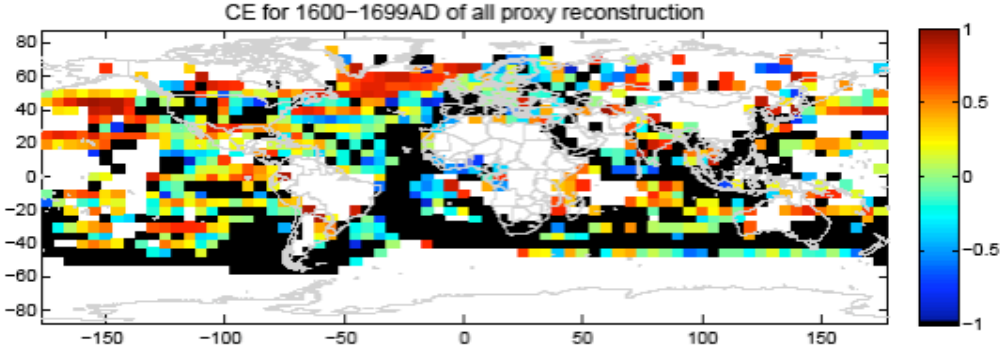
Figure S3: RegEM "pseudoproxy" tests. Shown are reconstructions of NH mean temperature (smoothed on timescales > 40 years for ease of comparison) using network "A" and SNR=0.4 as in ref. *S7*, comparing results using original criterion for selection of low-frequency TTLS cutoff (*K*) corresponding to 50% retained data variance, and revised criterion used in this study corresponding to 33% retained data variance. Results are shown for both **(a)** "NCAR CSM 1.4" simulation and **(b)** "GKSS ECHO-G" simulations (see ref *S7* for further details).


RE for 1800−1849AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.8
0.6
0.4
0.2
0
CE for 1800−1849AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.5
0
−0.5
−1
RE for 1700−1799AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.8
0.6
0.4
0.2
0
CE for 1700−1799AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.5
0
−0.5
−1

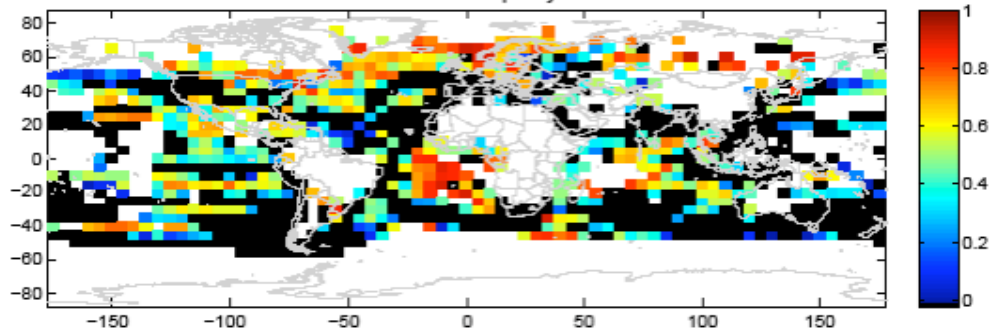
RE for 1600−1699AD of all proxy reconstruction



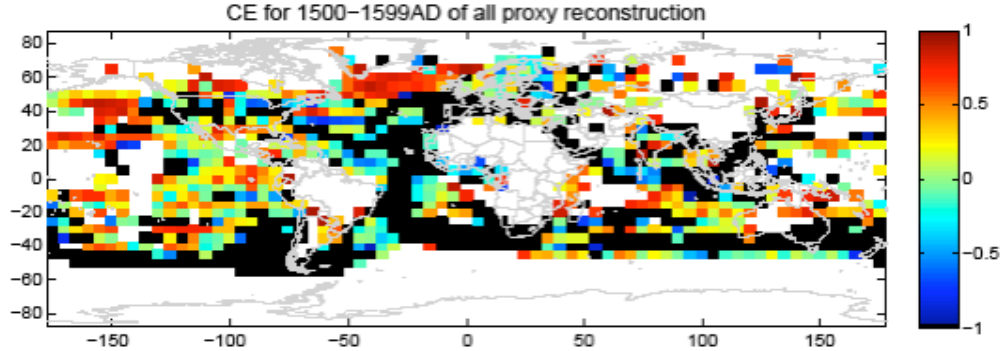
CE for 1600−1699AD of all proxy reconstruction

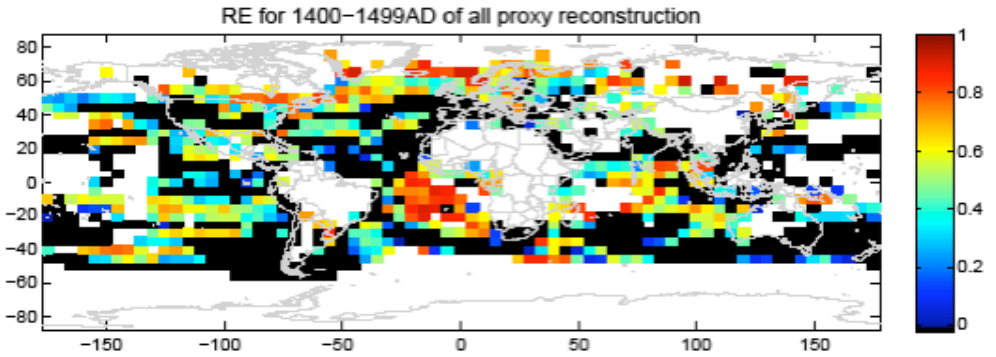


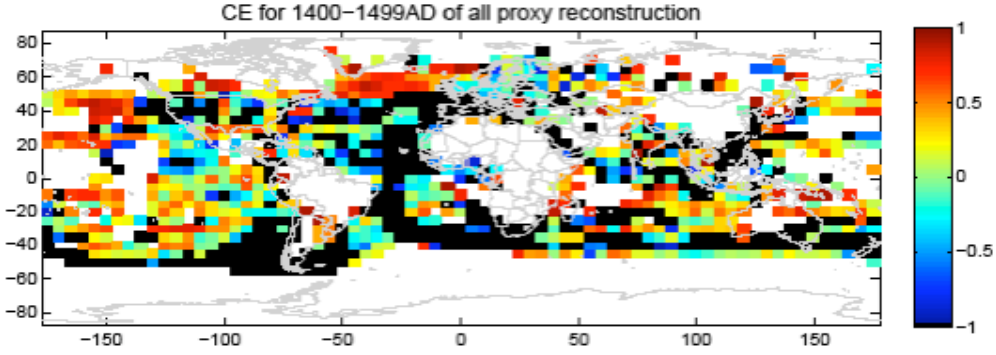
RE for 1500−1599AD of all proxy reconstruction

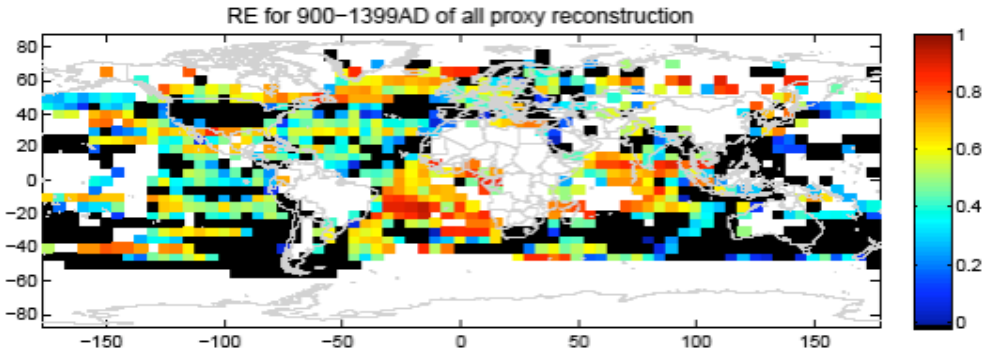


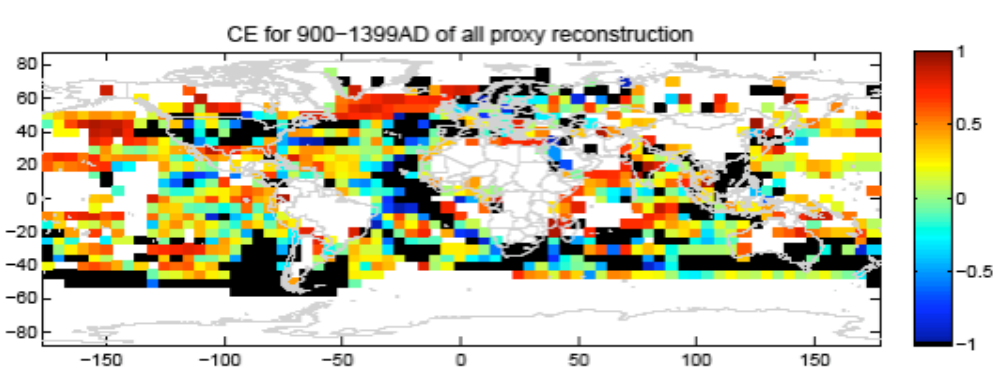
CE for 1500−1599AD of all proxy reconstruction



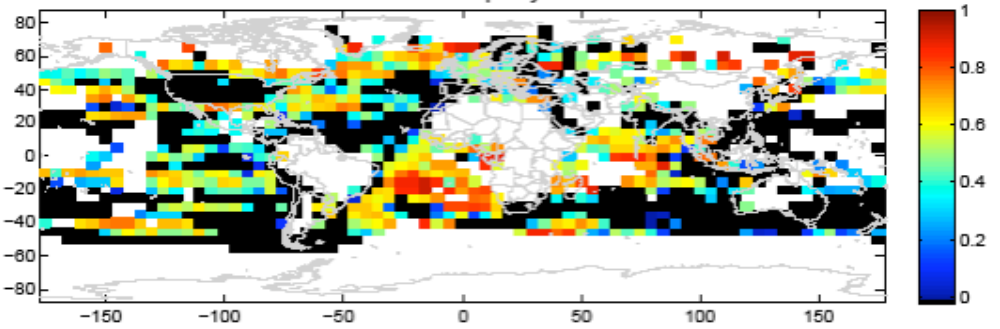

RE for 1400−1499AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.8
0.6
0.4
0.2
0


CE for 1400−1499AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.5
0
−0.5
−1

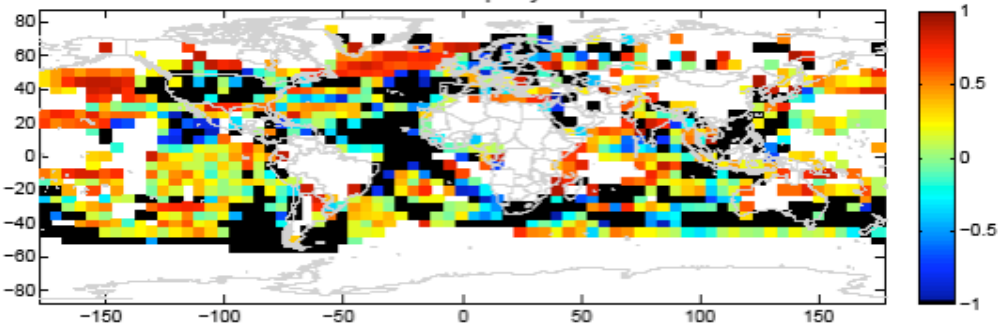

RE for 900−1399AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.8
0.6
0.4
0.2
0


CE for 900−1399AD of all proxy reconstruction
80
60
40
20
0
−20
−40
−60
−80
−150
−100
−50
0
50
100
150
1
0.5
0
−0.5
−1

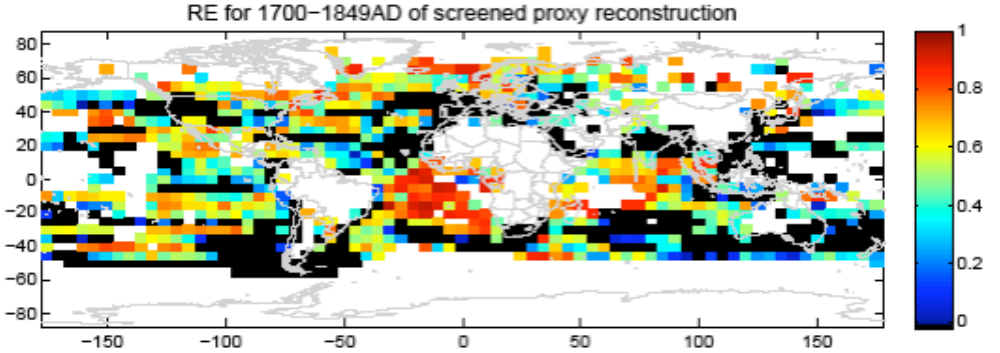
RE for 500−899AD of all proxy reconstruction



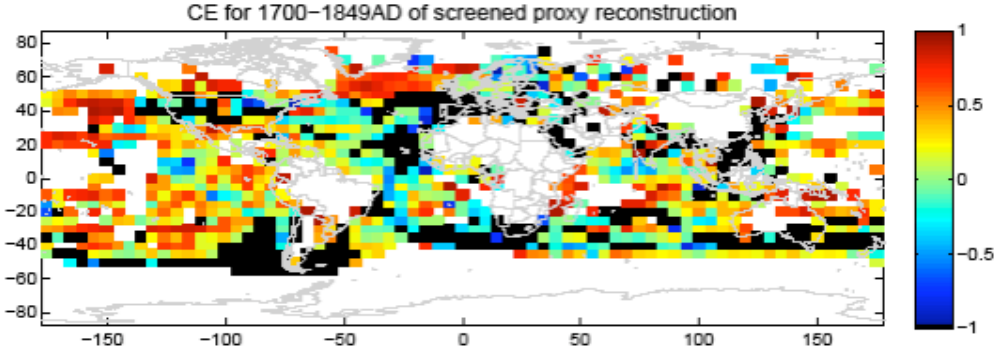
CE for 500−899AD of all proxy reconstruction



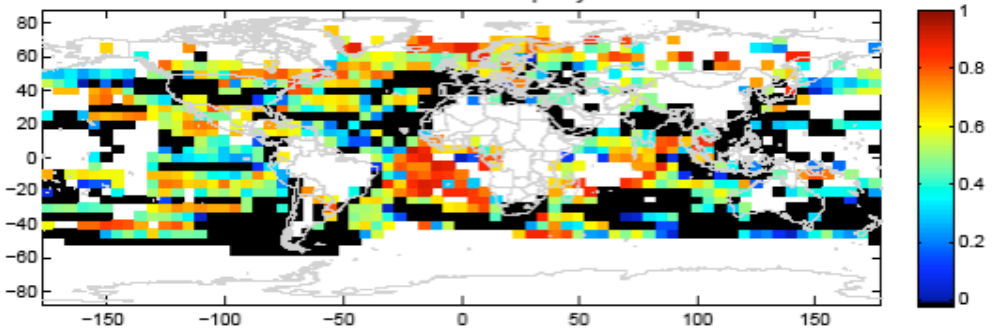
RE for 1700−1849AD of screened proxy reconstruction



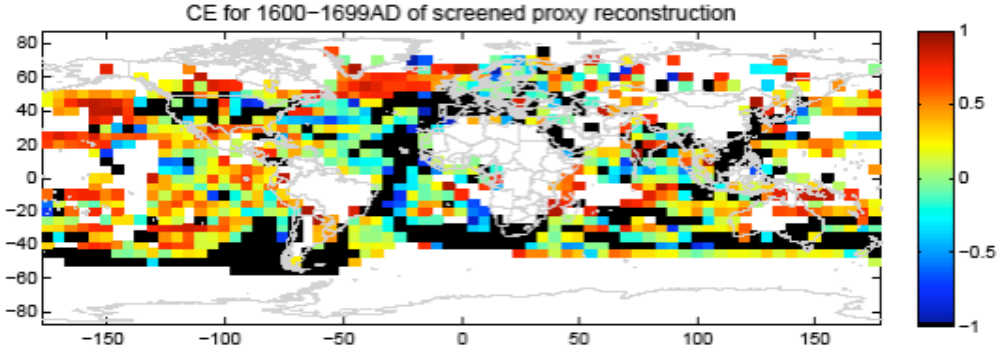
CE for 1700−1849AD of screened proxy reconstruction



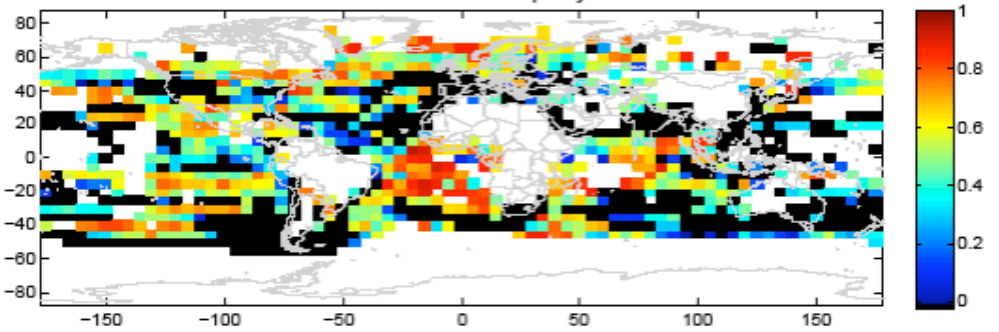
RE for 1600−1699AD of screened proxy reconstruction



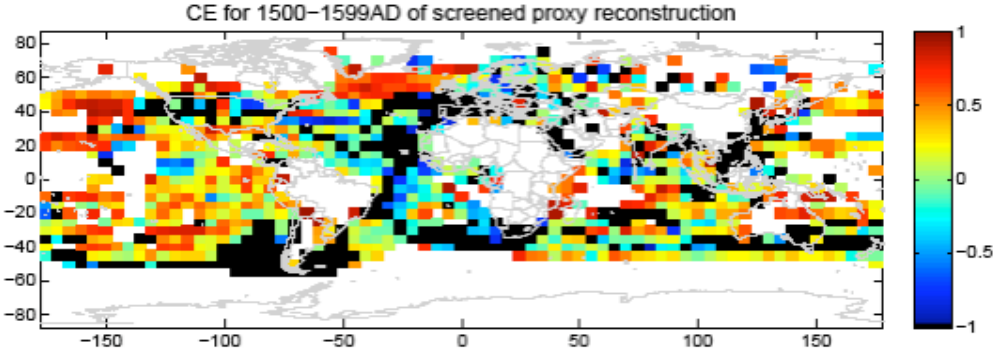
CE for 1600−1699AD of screened proxy reconstruction



RE for 1500−1599AD of screened proxy reconstruction



CE for 1500−1599AD of screened proxy reconstruction



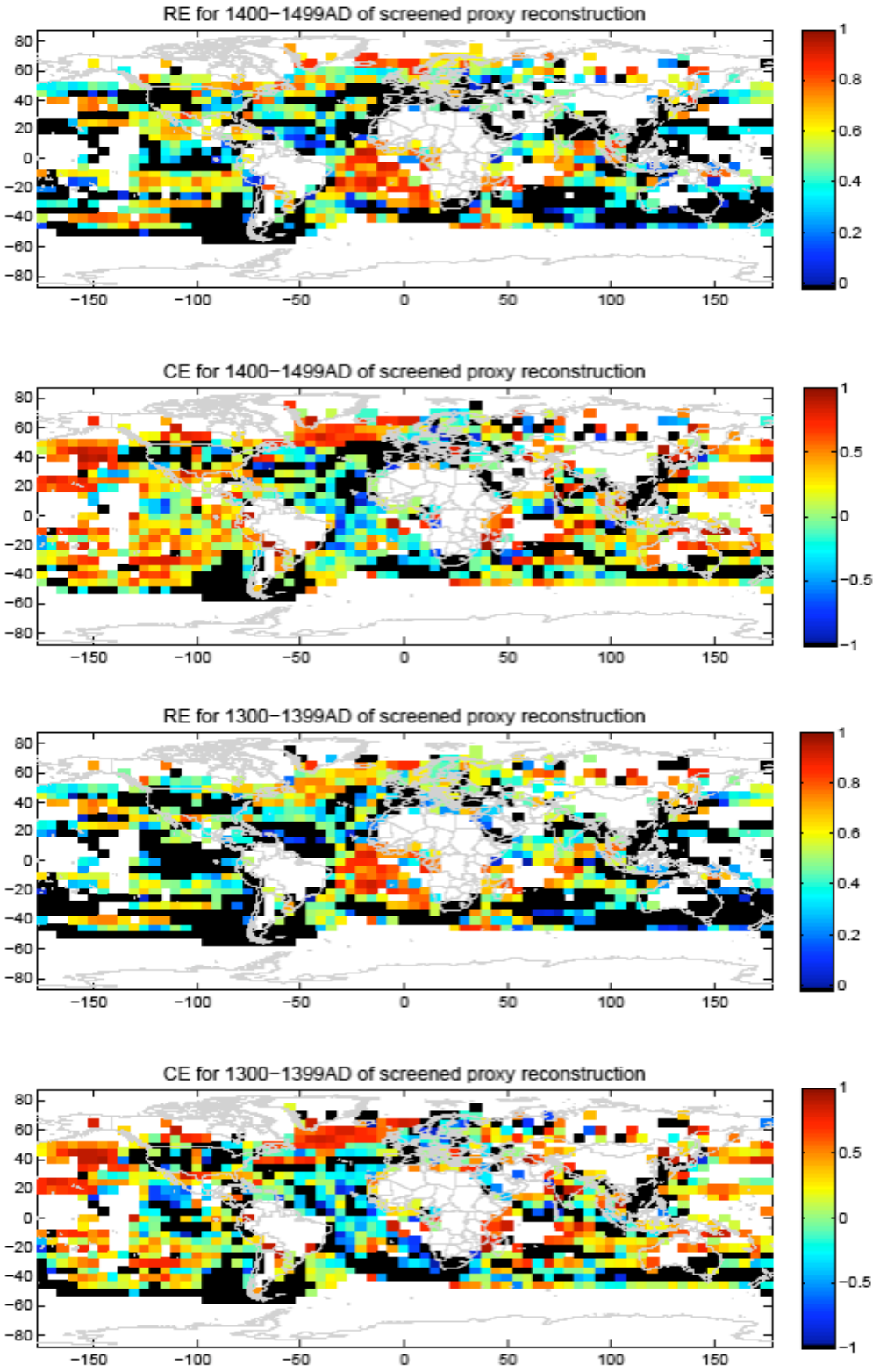
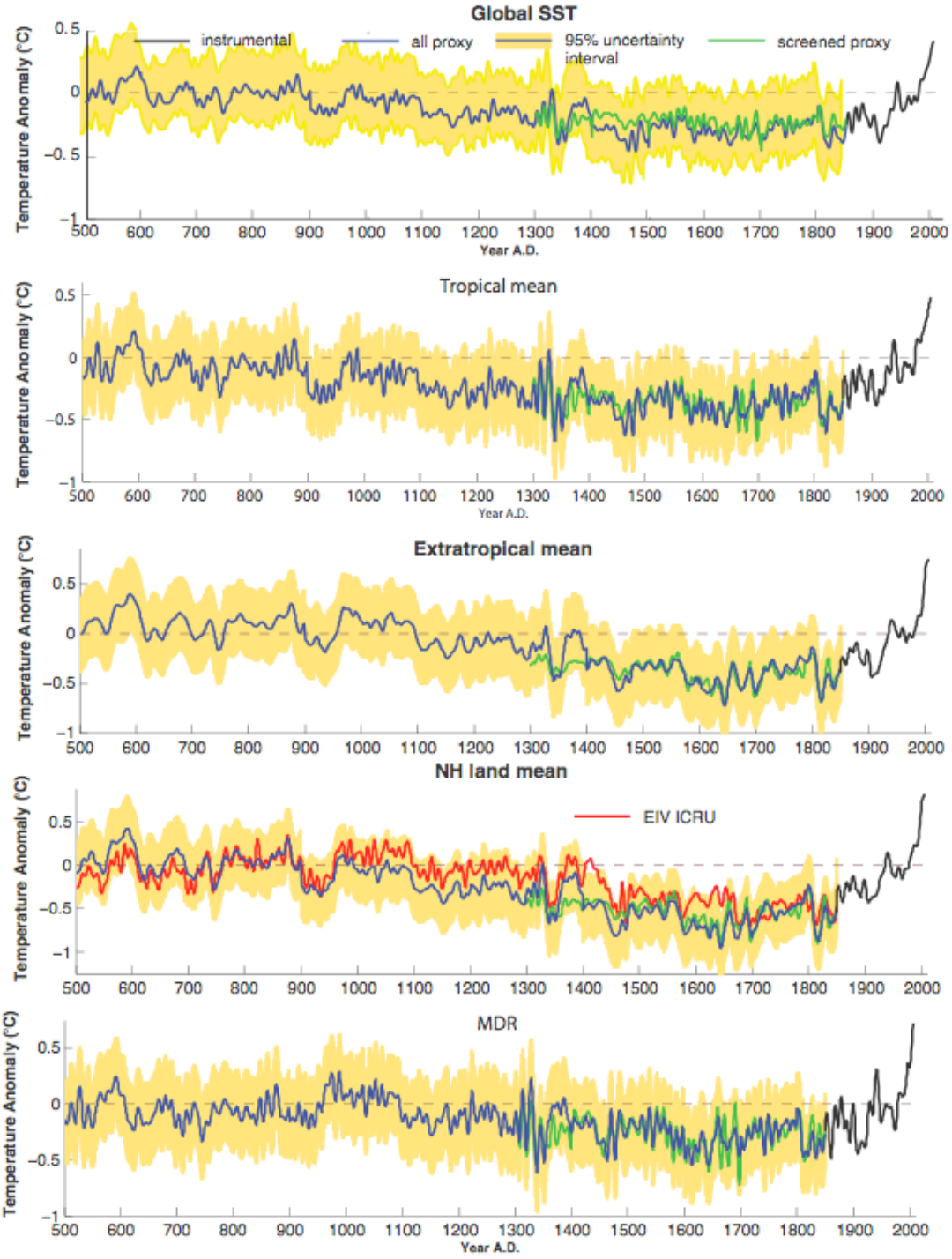


Figure S4: Spatial pattern of *RE* and *CE* validation statistics for each time interval of reconstruction. Both "all proxy" and "screened proxy" results are shown. Gridboxes that do not pass validation are blacked out. White regions contain too many missing values over the validation interval to calculate meaningful validation statistics.


Global SST
instrumental
all proxy
95% uncertainty interval
screened proxy
Temperature Anomaly (°C)
0.5
0
−0.5
−1
500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
Year A.D.
Tropical mean
Extratropical mean
NH land mean
EIV ICRU
MDR

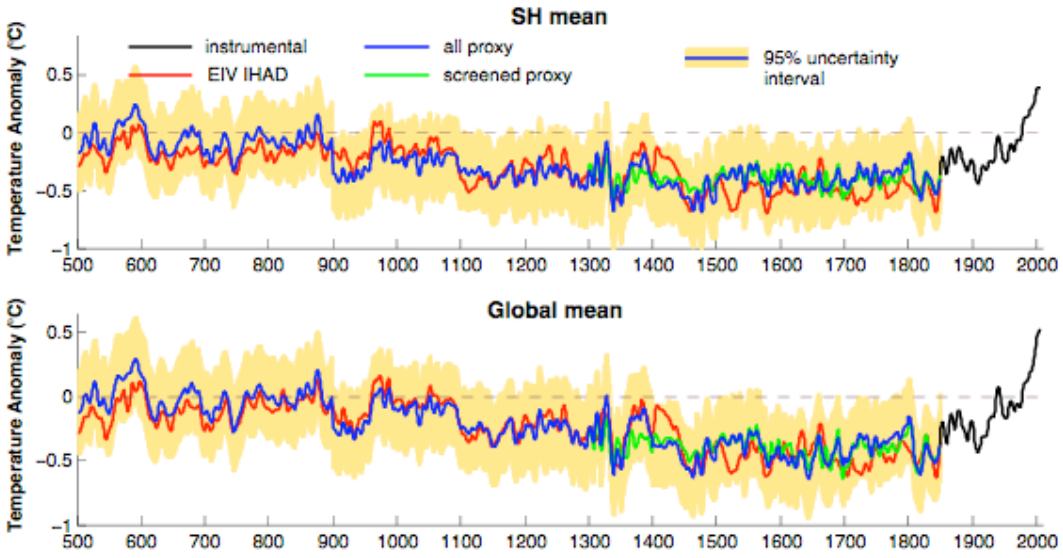
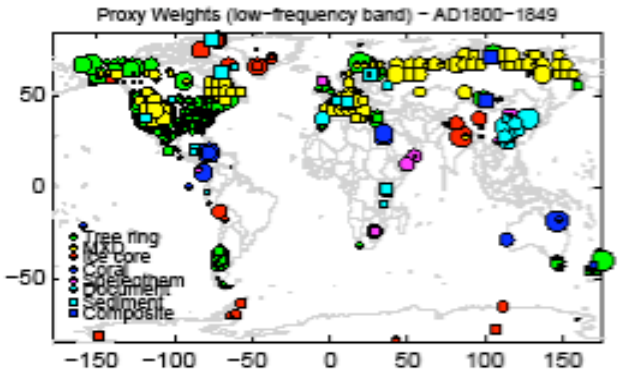
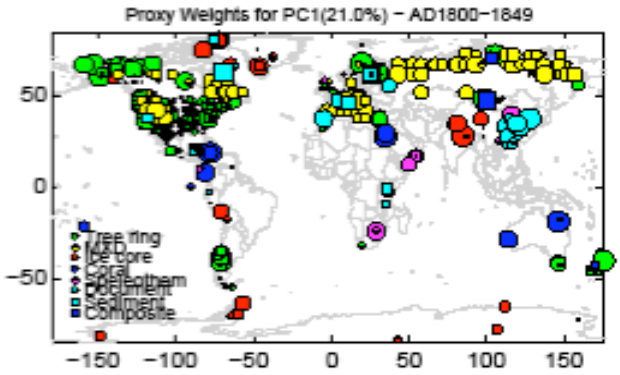
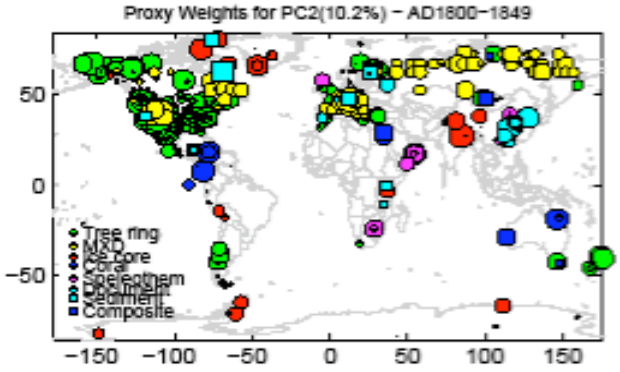
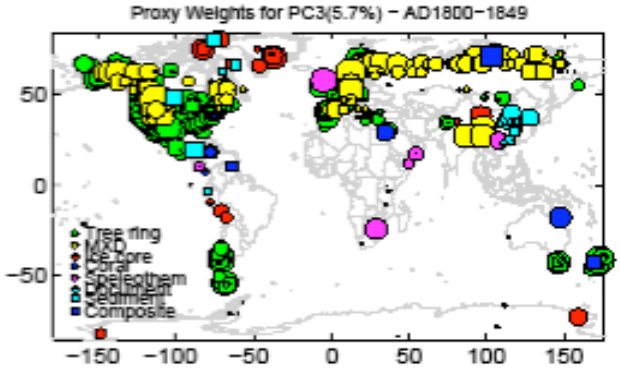


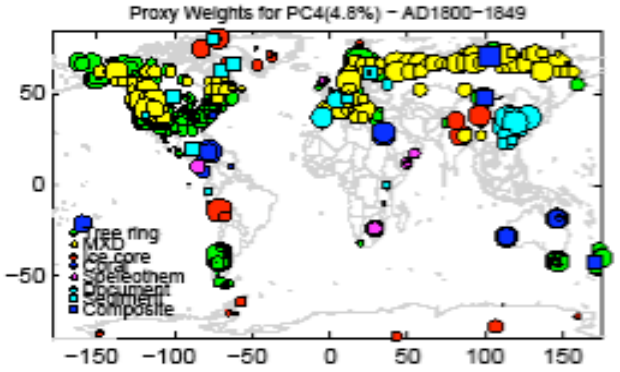
Figure S5: Additional temperature time series reconstructions. Shown are hemispheric and global mean temperatures averages and regional averages/indices not shown in the main article (see "Regional averages and indices" subsection above). Reconstructed series are shown with estimated 95% uncertainty intervals and modern instrumental series.

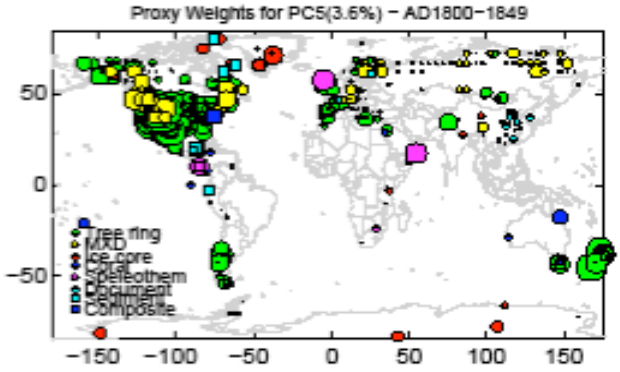

Proxy Weights (low−frequency band) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

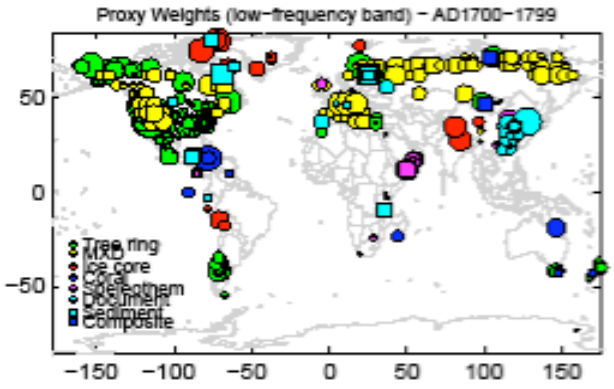

Proxy Weights for PC1(21.0%) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

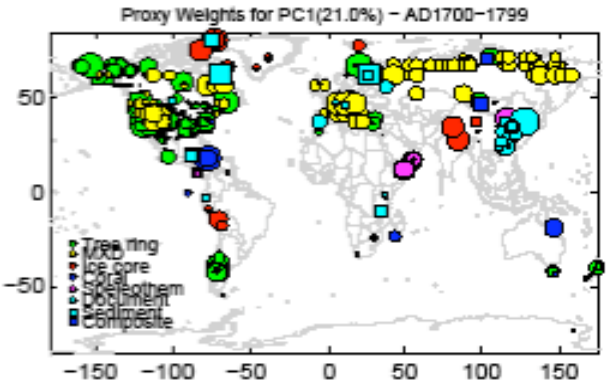

Proxy Weights for PC2(10.2%) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

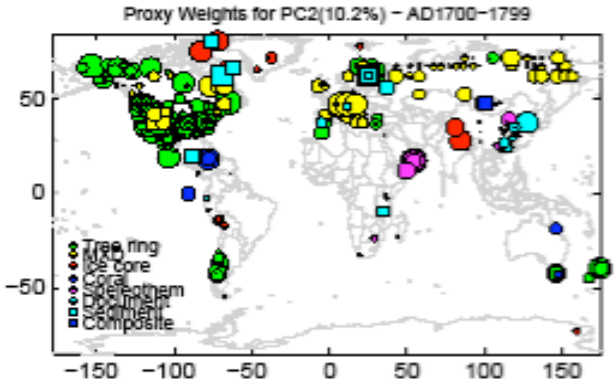

Proxy Weights for PC3(5.7%) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

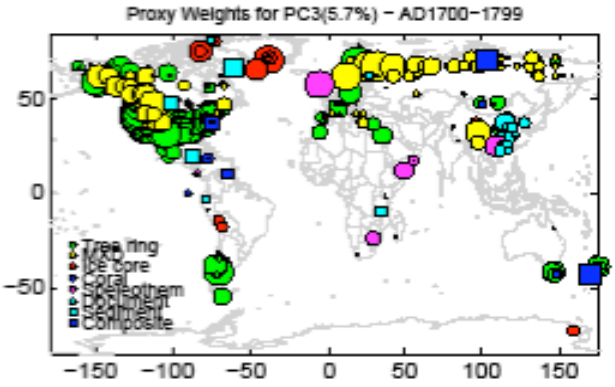

Proxy Weights for PC4(4.8%) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

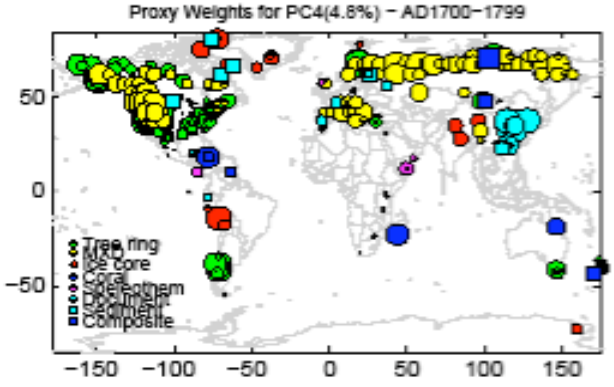

Proxy Weights for PC5(3.6%) − AD1800−1849
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

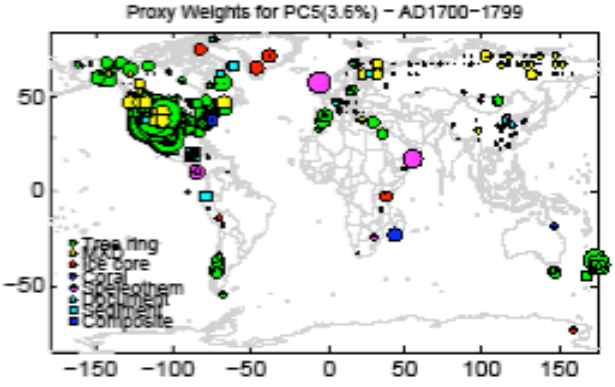

Proxy Weights (low−frequency band) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

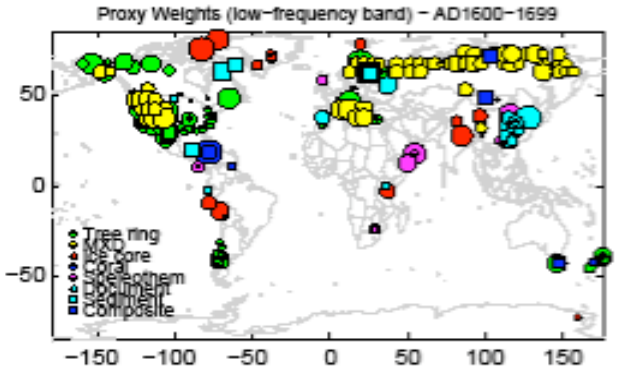

Proxy Weights for PC1(21.0%) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

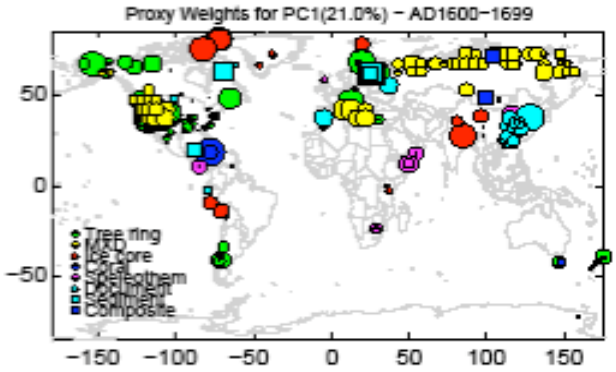

Proxy Weights for PC2(10.2%) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

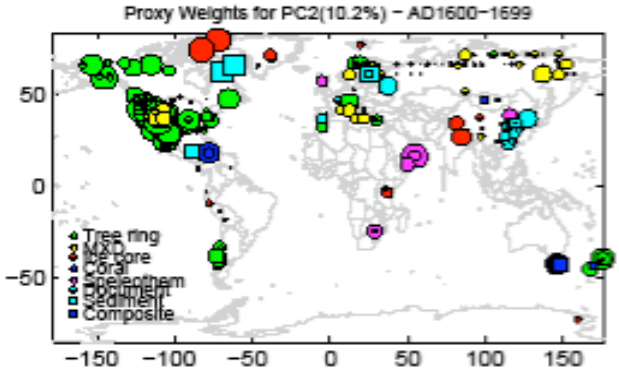

Proxy Weights for PC3(5.7%) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

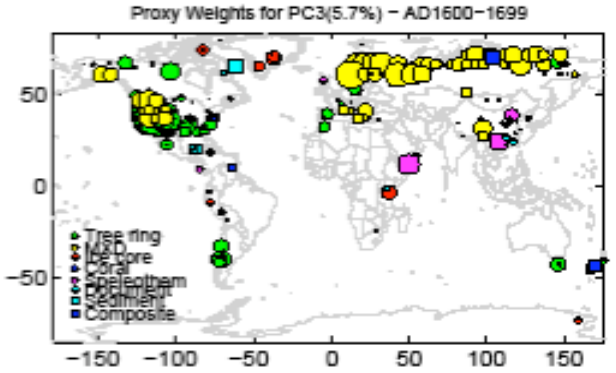

Proxy Weights for PC4(4.8%) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

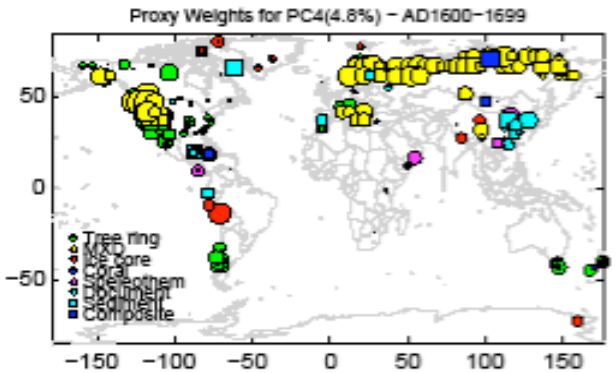

Proxy Weights for PC5(3.6%) − AD1700−1799
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Sheet/otherm
Documnt
Sediment
Composite

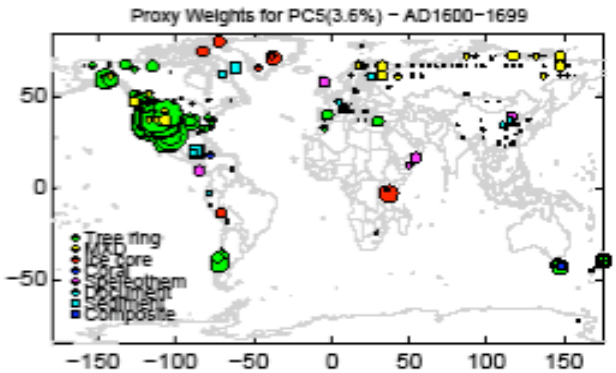

Proxy Weights (low−frequency band) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

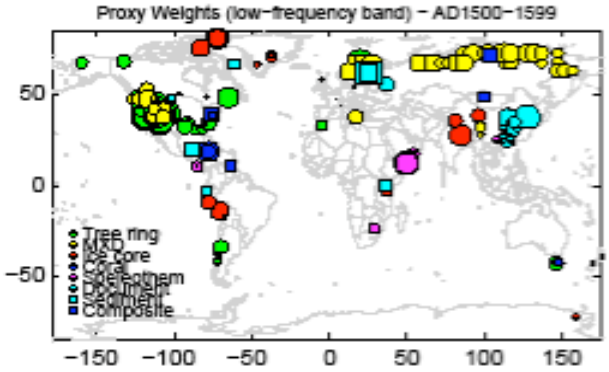

Proxy Weights for PC1(21.0%) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

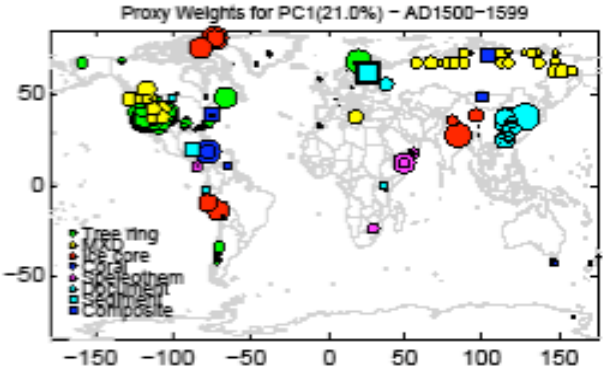

Proxy Weights for PC2(10.2%) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

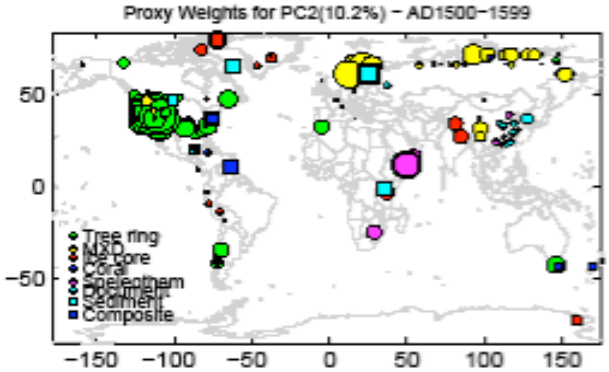

Proxy Weights for PC3(5.7%) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

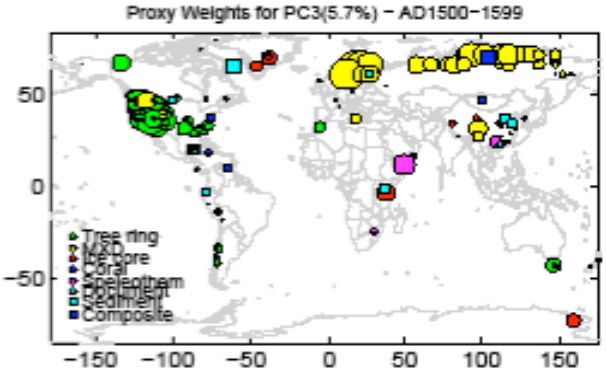

Proxy Weights for PC4(4.8%) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

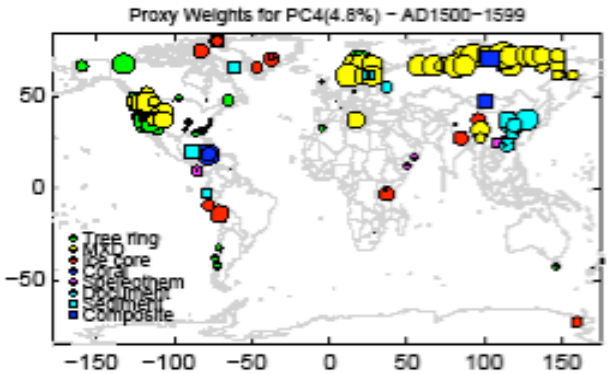

Proxy Weights for PC5(3.6%) − AD1600−1699
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite
50
0
−50
−150
−100
−50
0
50
100
150

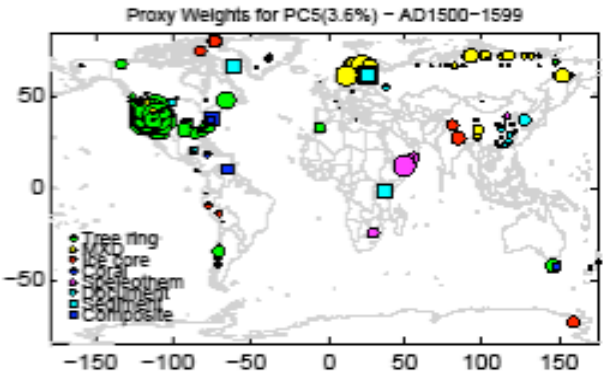

Proxy Weights (low−frequency band) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

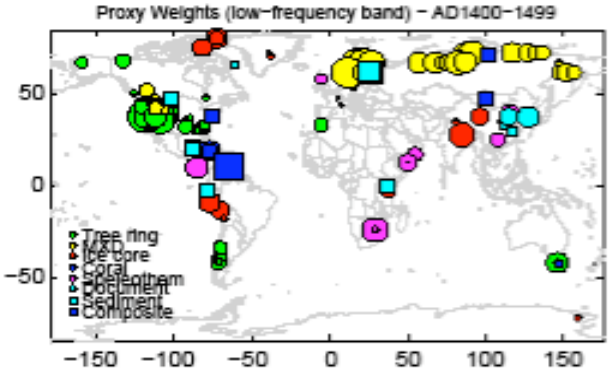

Proxy Weights for PC1(21.0%) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

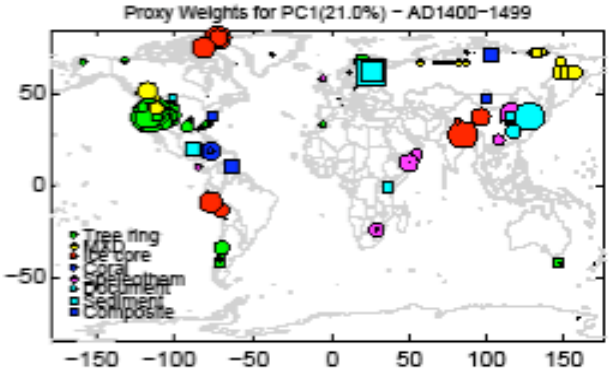

Proxy Weights for PC2(10.2%) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

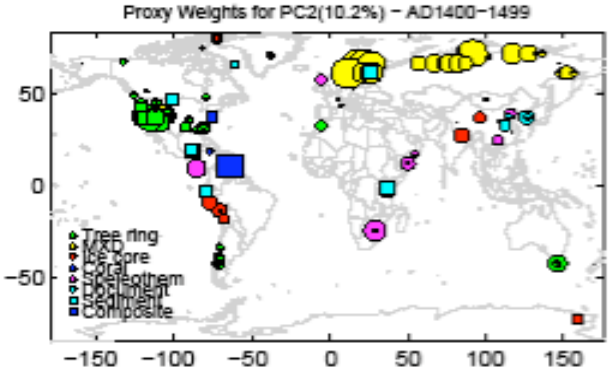

Proxy Weights for PC3(5.7%) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

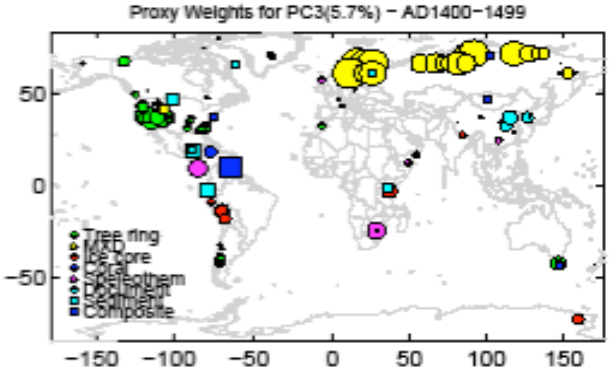

Proxy Weights for PC4(4.8%) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

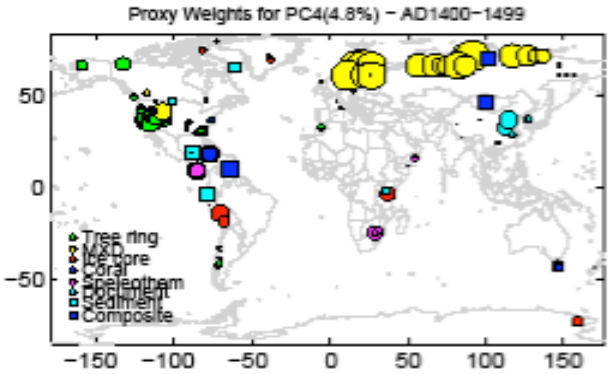

Proxy Weights for PC5(3.6%) − AD1500−1599
Tree ring
MXD
Ice core
Coral
Speleotherm
Documents
Sediment
Composite
50
0
−50
−150 −100 −50 0 50 100 150

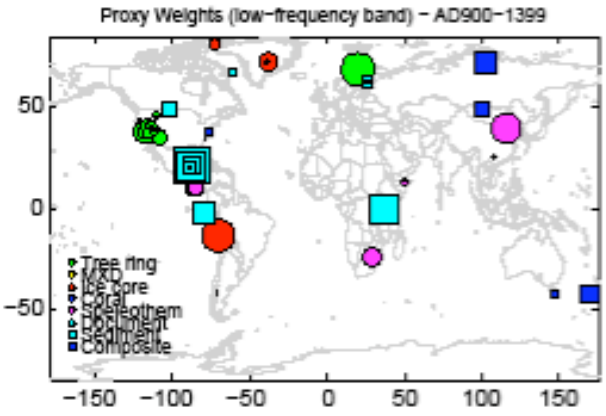

Proxy Weights (low−frequency band) − AD1400−1499
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
Ice core
Coral
Speleothem
Sediment
Composite

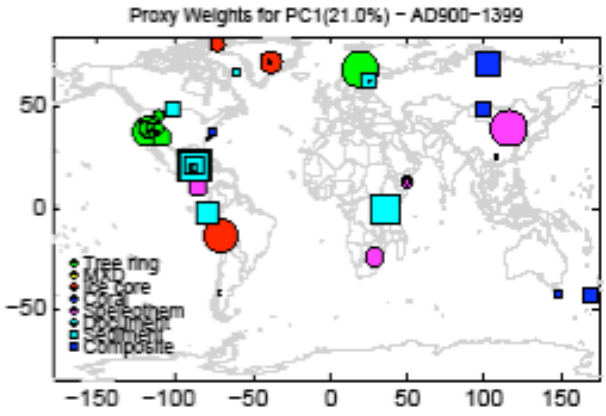

Proxy Weights for PC1(21.0%) − AD1400−1499
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
Ice core
Coral
Speleothem
Sediment
Composite

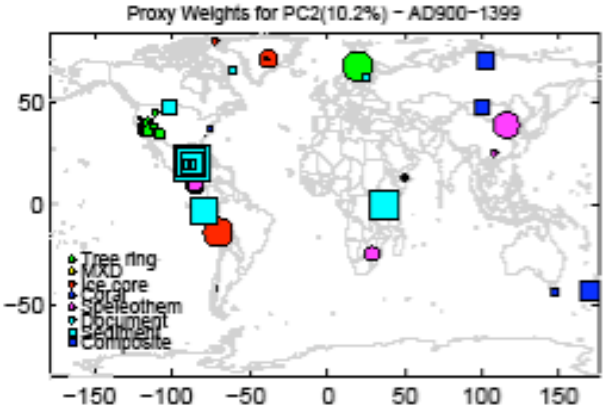

Proxy Weights for PC2(10.2%) − AD1400−1499
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
Ice core
Coral
Speleothem
Sediment
Composite

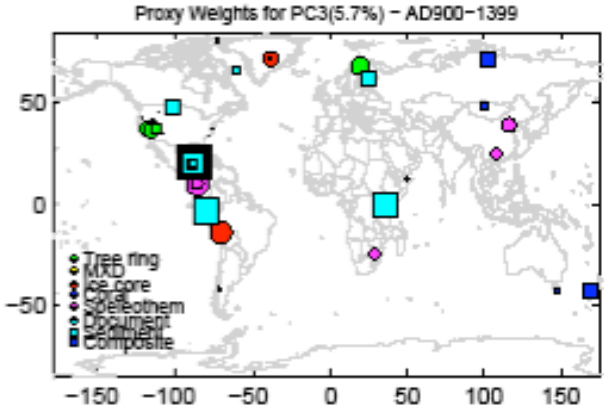

Proxy Weights for PC3(5.7%) − AD1400−1499
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
Ice core
Coral
Speleothem
Sediment
Composite


Proxy Weights for PC4(4.8%) − AD1400−1499
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
Ice core
Coral
Speleothem
Sediment
Composite


Proxy Weights (low−frequency band) − AD900−1399
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite


Proxy Weights for PC1(21.0%) − AD900−1399
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite


Proxy Weights for PC2(10.2%) − AD900−1399
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite


Proxy Weights for PC3(5.7%) − AD900−1399
50
0
−50
−150 −100 −50 0 50 100 150
Tree ring
MXD
Ice core
Coral
Speleothem
Document
Sediment
Composite

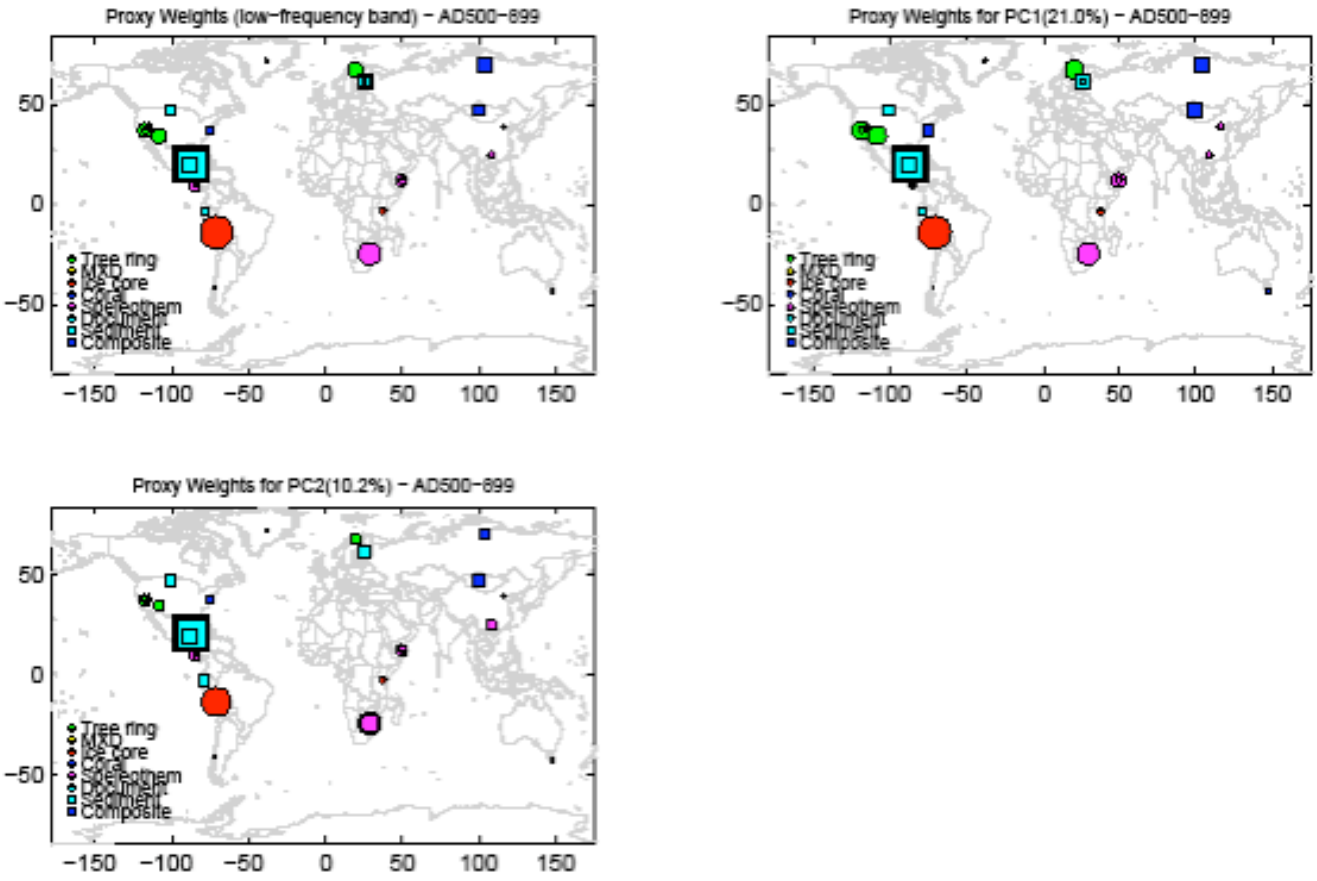
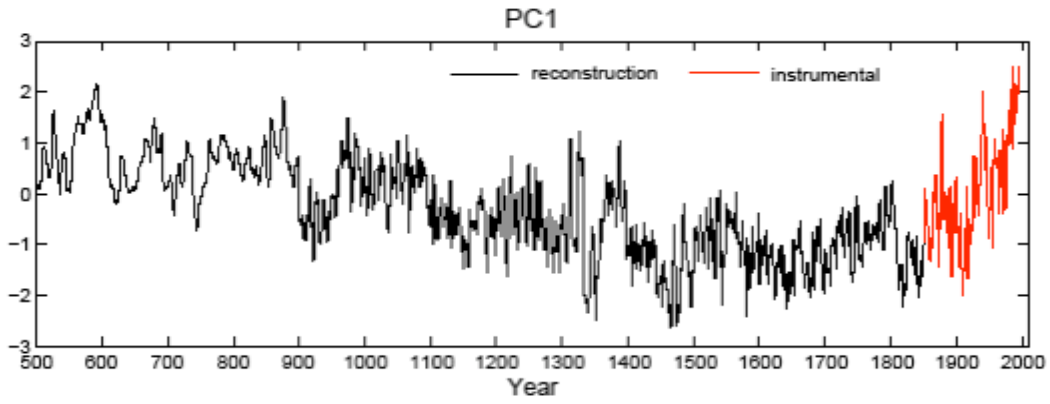
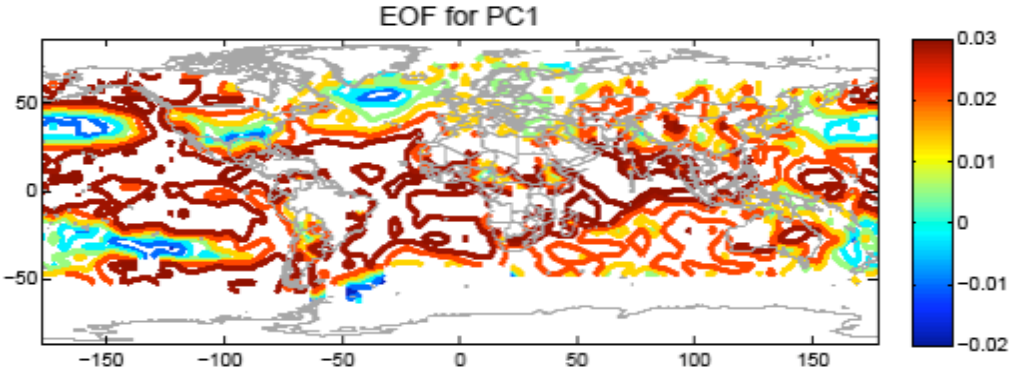
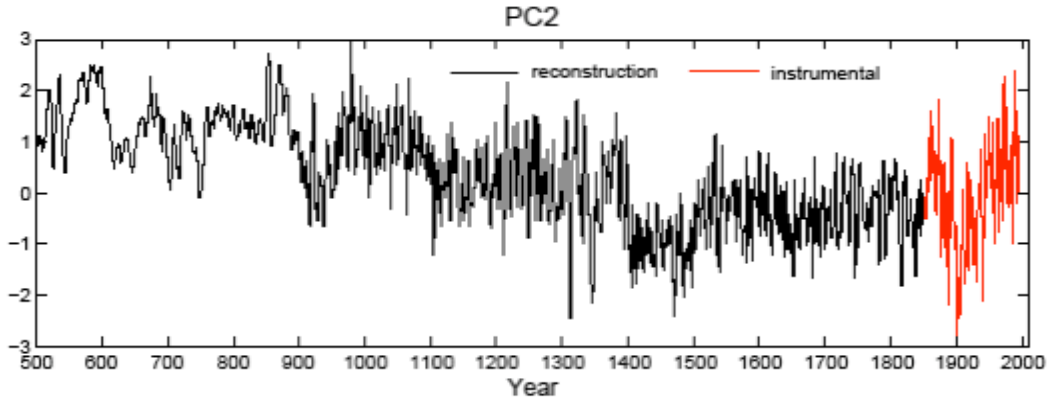
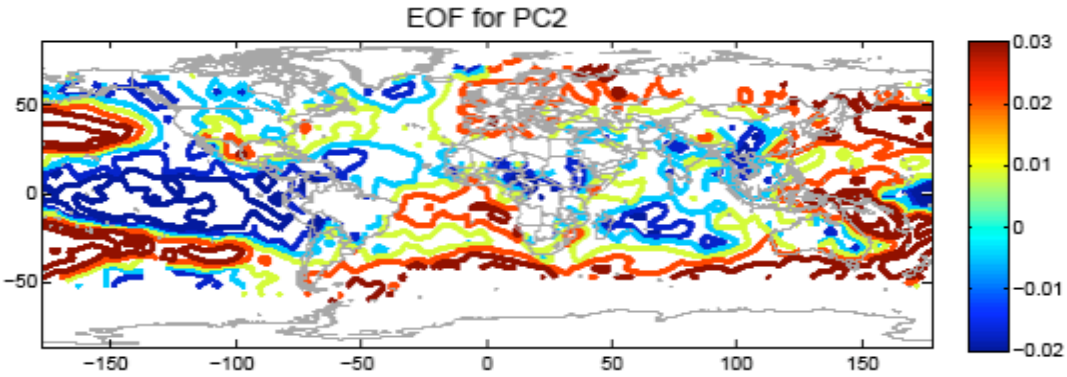


Figure S6: Low-Frequency weights on proxies for each interval of reconstruction. Shown are weights for individual reconstructed low-frequency PCs and the overall low-frequency reconstruction itself. Relative weights are shown by size of symbol.


PC1
reconstruction
instrumental
Year


EOF for PC1


PC2
reconstruction
instrumental
Year


EOF for PC2

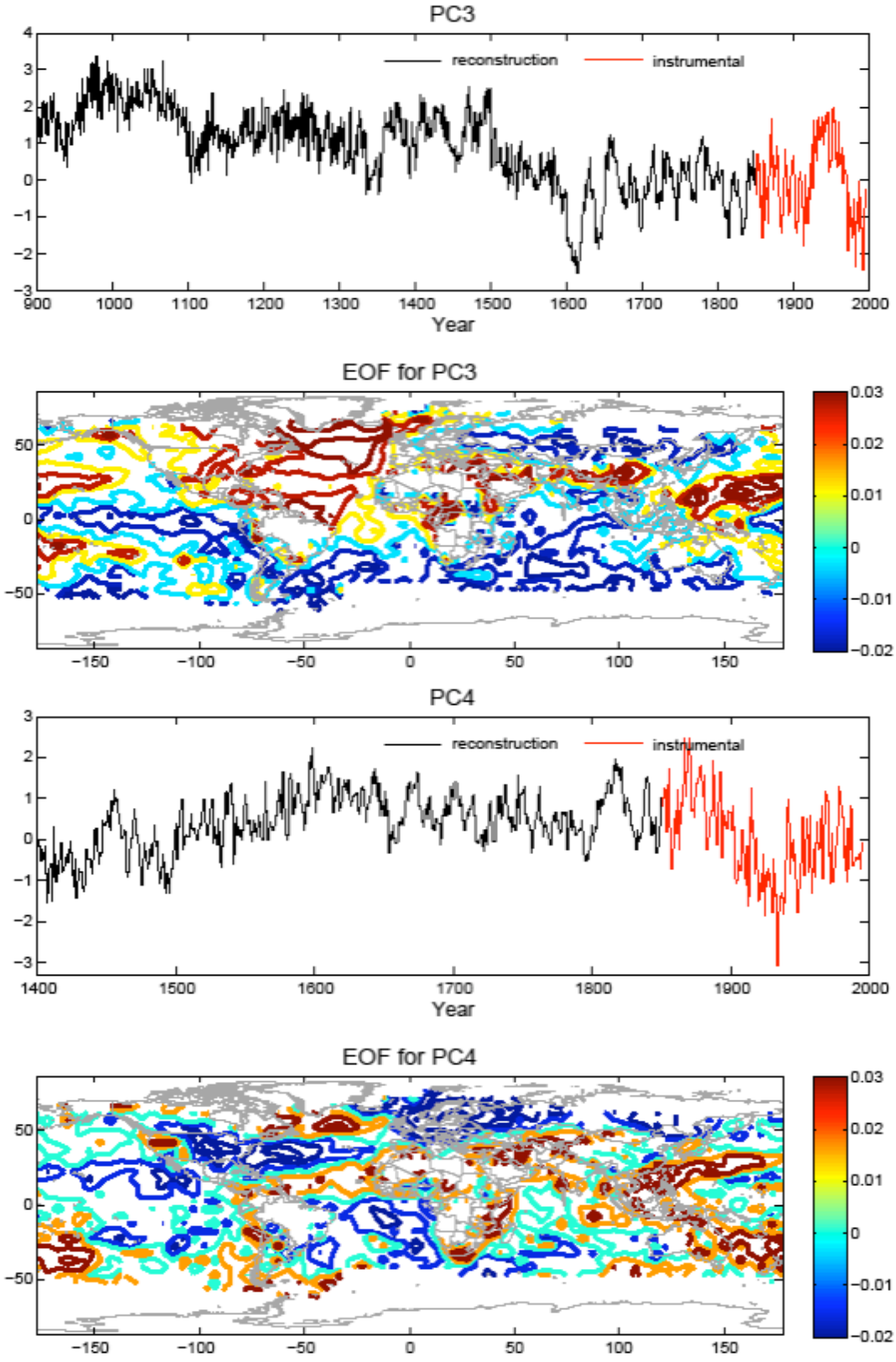


Figure S7: EOF patterns and associated PC series. Shown are the EOF patterns and PC series (instrumental PC series shown in red, reconstructed PC series shown in black) for each of the 4 leading surface temperature modes. All series are decadally smoothed.

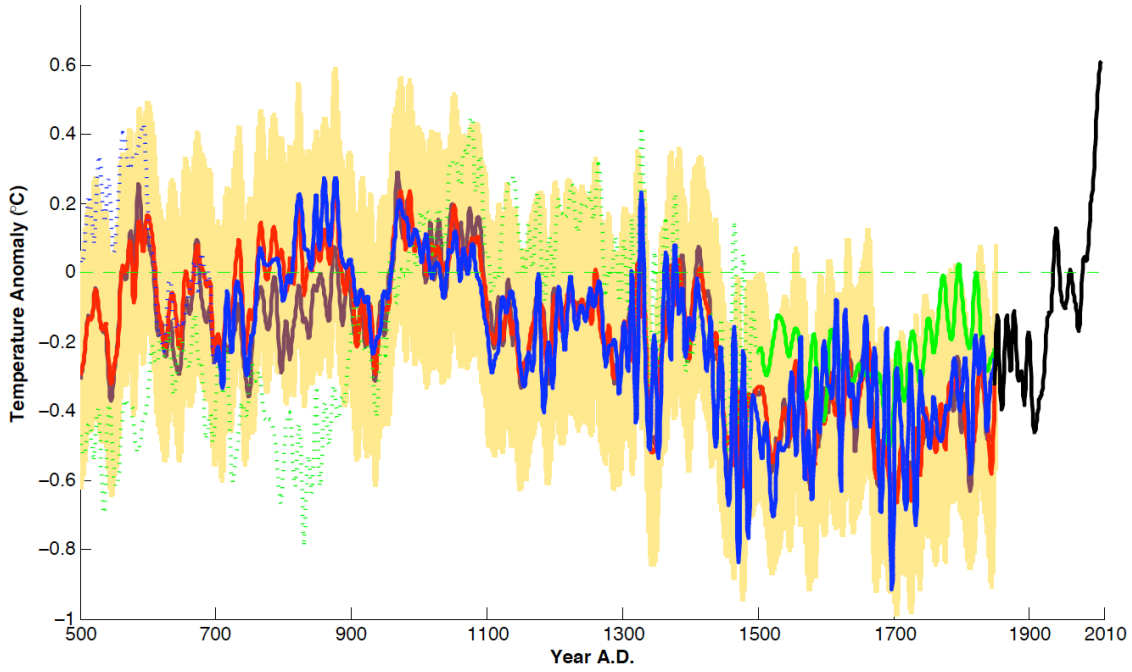


Figure S8: Sensitivity of NH mean reconstruction to exclusion of selected proxy record. Reconstructions are shown based on "all proxy" network (red, with two standard error region shown in yellow) proxy network with all tree-ring records removed (blue), proxy network with a group of 7 long-term proxy with greater uncertainties and/or potential biases as discussed in ref. *S1* (brown) and both tree-ring data and the group of 7 records removed (green; dashed before AD 1500 indicates reconstruction no longer passes validation).

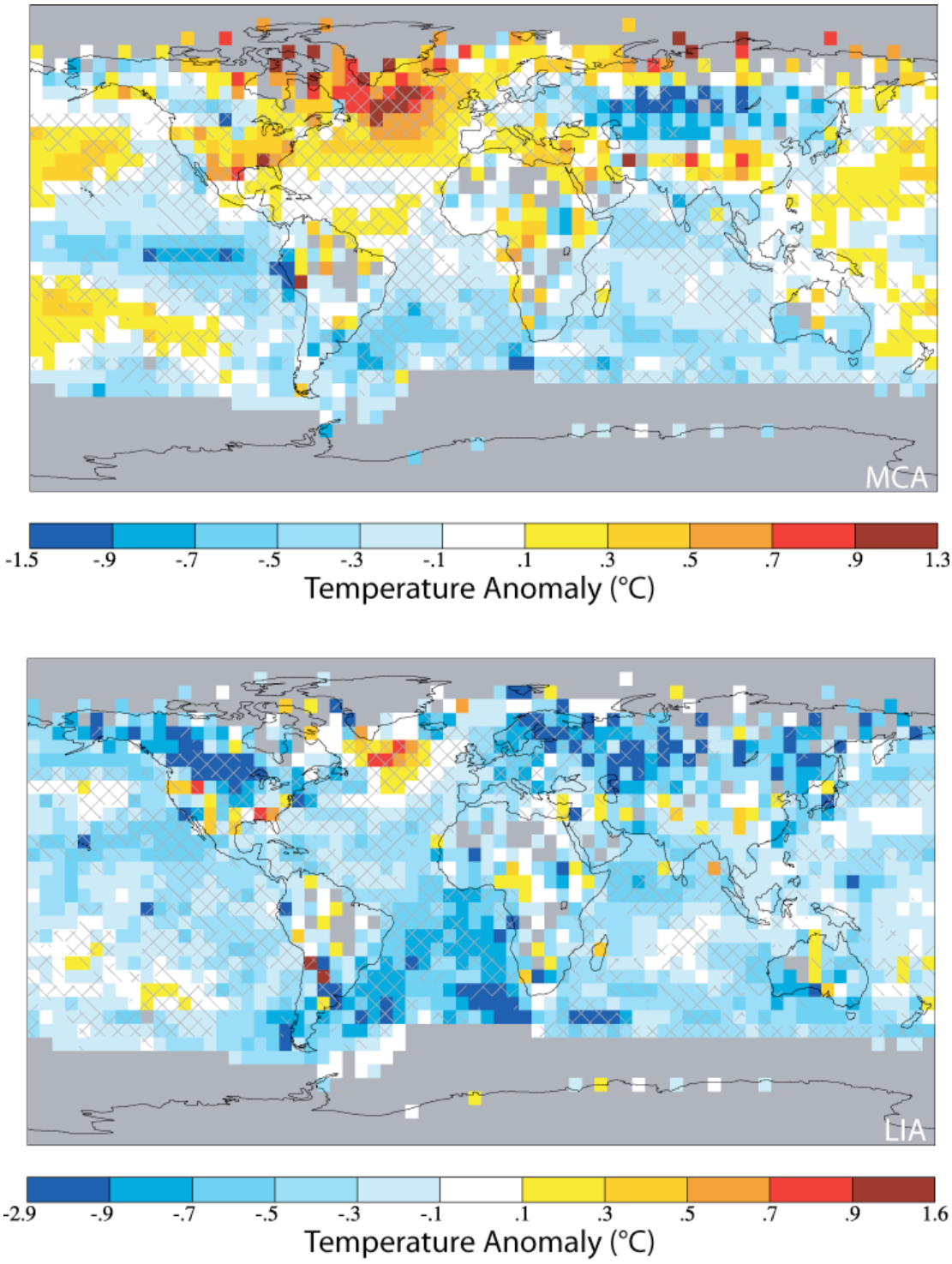


Figure S9: Patterns of alternatively defined LIA and MCA intervals. As in Figure 2 of main article, but using alternative intervals (AD 1600-1850 and AD 900-1100, respectively) to define the MCA (top) and LIA intervals (bottom).

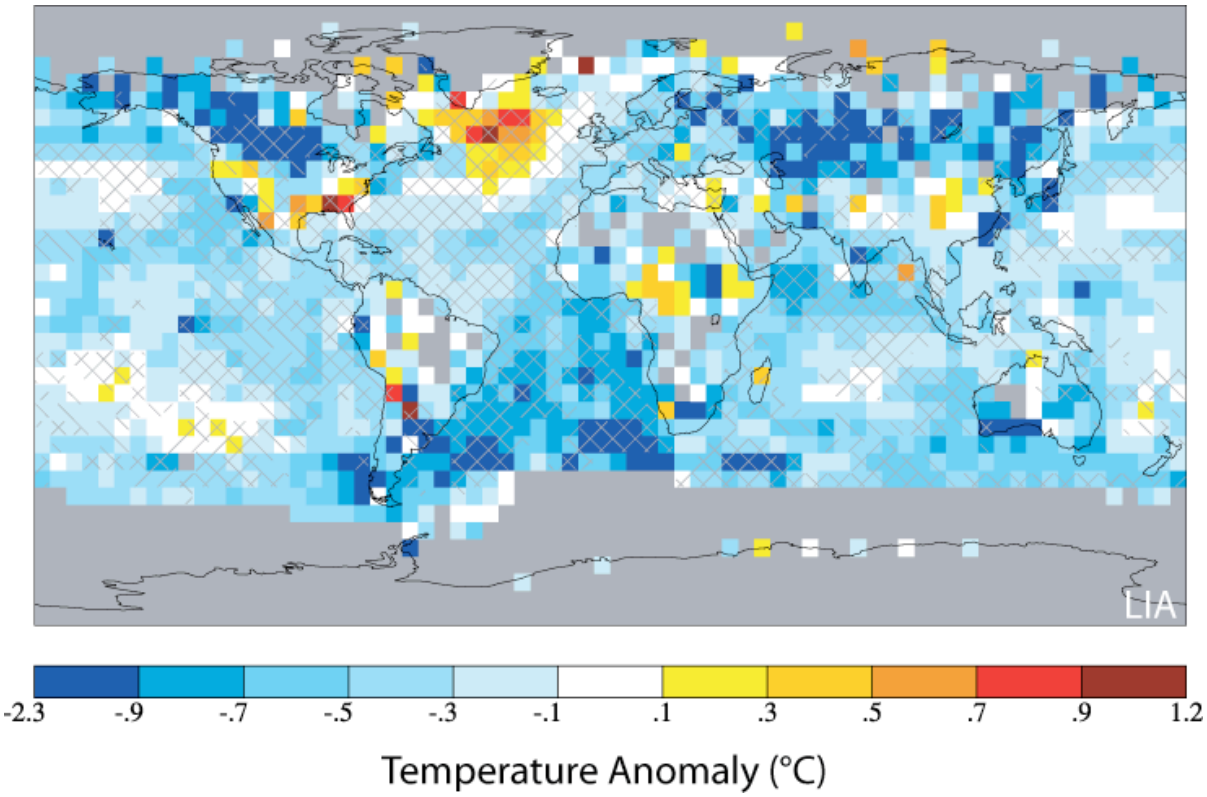


Figure S10: Sensitivity of reconstructed spatial pattern to richness of available proxy network. As in Figure 2 of main article, but using MCA network (network of proxy data available to beginning of AD 900-999 step of reconstruction—see Table S3) to reconstruct LIA interval.

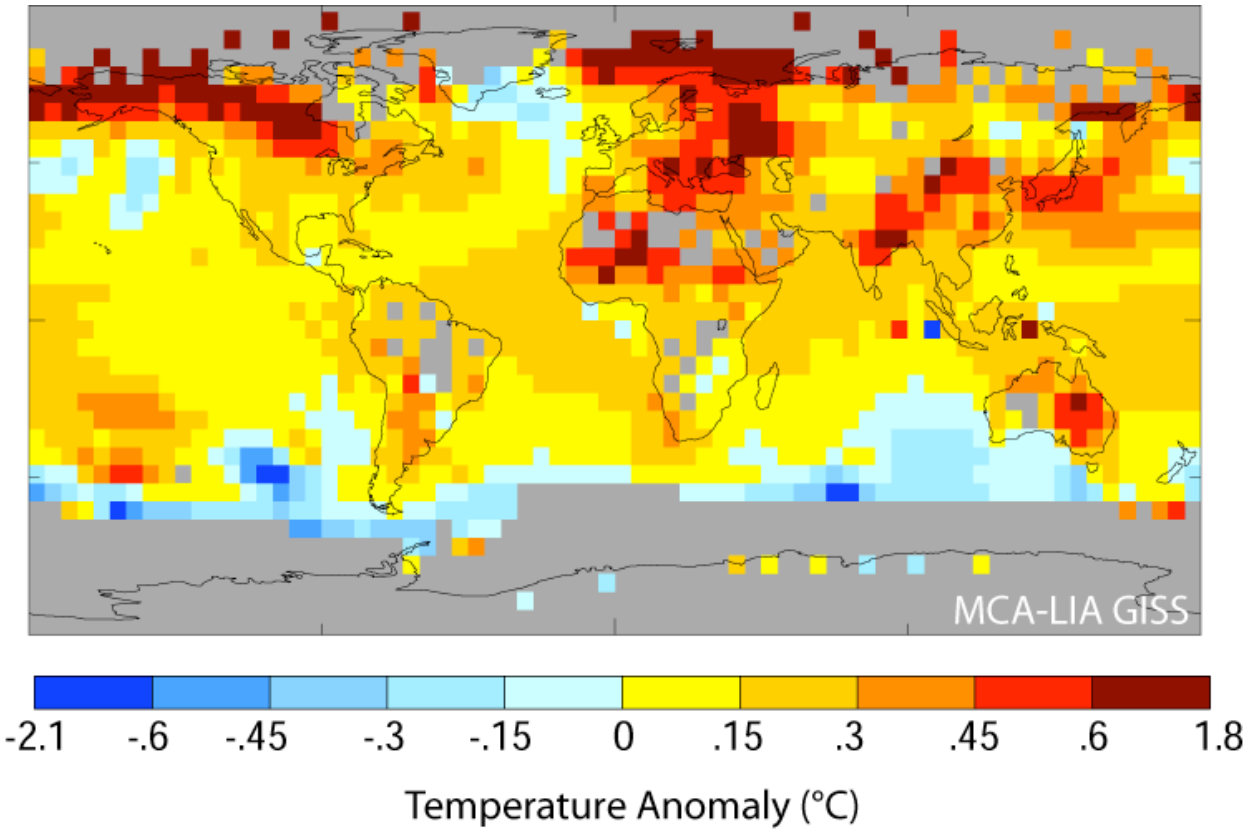


Figure S11: Alternative GISS-ER MCA-LIA Surface Temperature pattern. As in Figure 3c of main article, but showing a single realization of the forced GISS-ER simulation in place of the ensemble mean, which shows greater warming over Europe.

Supporting tables

Table S1: Number of annually, decadally, and total proxy data available from different sources for different starting years.

(a) Full proxy data set

	NH proxies						SH proxies						
	Annually resolved proxies			Annually+decadally resolved proxies				Annually resolved proxies			Annually+decadally resolved proxies		
interval	Dendro	Others	Total	Dendro	Others	Total	interval	Dendro	Others	Total	Dendro	Others	Total
1800-1855	889	104	993	889	147	1036	1800-1855	143	22	165	143	30	173
1700-1799	563	95	658	563	137	700	1700-1799	78	12	90	78	17	95
1600-1699	276	91	367	276	132	408	1600-1699	39	9	48	39	13	52
1500-1599	150	89	239	150	127	277	1500-1599	22	8	30	22	12	34
1400-1499	105	18	123	105	46	151	1400-1499	14	8	22	14	12	26
1300-1399	62	16	78	62	40	102	1300-1399	8	8	16	8	12	20
1200-1299	41	12	53	41	36	77	1200-1299	6	7	13	6	11	17
1100-1199	29	12	41	29	34	63	1100-1199	5	7	12	5	11	16
1000-1099	18	12	30	18	28	46	1000-1099	4	7	11	4	9	13
900-999	15	10	25	15	25	40	900-999	2	7	9	2	9	11
800-899	13	9	22	13	24	37	800-899	2	6	8	2	7	9
700-799	11	8	19	11	23	34	700-799	2	6	8	2	7	9
600-699	11	8	19	11	22	33	600-699	1	6	7	1	7	8
500-599	7	7	14	7	21	28	500-599	1	6	7	1	7	8
400-499	7	7	14	7	21	28	400-499	1	4	5	1	4	5
300-399	6	7	13	6	21	27	300-399	1	4	5	1	4	5
200-299	4	6	10	4	20	24	200-299	0	4	4	0	4	4
100-199	3	6	9	3	20	23	100-199	0	4	4	0	4	4
0-99	2	6	8	2	19	21	0-99	0	4	4	0	4	4

	Global proxies					
	Annually resolved proxies			Annually+decadally resolved pro		
interval	Dendro	Others	Total	Dendro	Others	Total
1800-1855	1032	126	1158	1032	177	1209
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

b) Screened proxy data set

	NH proxies						SH proxies						
	Annually resolved proxies			Annually+decadally resolved proxies				Annually resolved proxies			Annually+decadally resolved proxies		
Interval	Dendro	Others	Total	Dendro	Others	Total	Interval	Dendro	Others	Total	Dendro	Others	Total
1800-1855	305	98	403	305	115	420	1800-1855	50	14	64	50	14	64
1700-1799	225	90	315	225	107	332	1700-1799	31	9	40	31	9	40
1600-1699	117	86	203	117	102	219	1600-1699	13	7	20	13	7	20
1500-1599	62	85	147	62	100	162	1500-1599	8	6	14	8	6	14
1400-1499	41	14	55	41	23	64	1400-1499	5	6	11	5	6	11
1300-1399	18	12	30	18	20	38	1300-1399	2	6	8	2	6	8
1200-1299	12	9	21	12	17	29	1200-1299	1	5	6	1	5	6
1100-1199	7	9	16	7	17	24	1100-1199	1	5	6	1	5	6
1000-1099	3	9	12	3	16	19	1000-1099	1	5	6	1	5	6
900-999	2	7	9	2	13	15	900-999	0	5	5	0	5	5
800-899	2	6	8	2	12	14	800-899	0	4	4	0	4	4
700-799	2	6	8	2	12	14	700-799	0	4	4	0	4	4
600-699	2	6	8	2	11	13	600-699	0	4	4	0	4	4
500-599	2	5	7	2	10	12	500-599	0	4	4	0	4	4
400-499	2	5	7	2	10	12	400-499	0	3	3	0	3	3
300-399	1	5	6	1	10	11	300-399	0	3	3	0	3	3
200-299	1	5	6	1	10	11	200-299	0	3	3	0	3	3
100-199	1	5	6	1	10	11	100-199	0	3	3	0	3	3
0-99	0	5	5	0	10	10	0-99	0	3	3	0	3	3

	Global proxies					
	Annually resolved proxies			Annually+decadally resolved prox		
Interval	Dendro	Others	Total	Dendro	Others	Total
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: Skill scores for pseudoproxy experiments. The experiments compare results from the selection rules used by ref. *S9* with the alternative selection rules discussed above. In all cases, 'white noise' pseudoproxy networks were used, with signal-to-noise variance ratio SNR=0.4 as defined in ref. *S9*. Calibration was performed using the "long" calibration interval (1850-1999) and validation was performed over the entire pre-calibration interval (note that, as shown in ref. *S9* this leads to higher *CE* scores than use of the short (roughly 50 year) validation period typically available for actual proxy reconstructions). For each experiment, the best average validation scores achieved are highlighted in yellow. Scores are for decadally-smoothed data as in refs. *S1* and *S9*.

CSM: SNR=0.4, long calibration, white noise network **A**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	Mult RE (NH)	Mult CE (NH)	Mult r^2 (NH)
4	50% (4)	0.93	0.67	0.71	0.24	0.00	0.12	0.29	0.08	0.16
4	40% (3)	0.94	0.72	0.74	0.25	0.01	0.13	0.28	0.07	0.16
4	33% (2)	0.94	0.71	0.73	0.30	0.08	0.13	0.35	0.16	0.17
4	25% (2)	0.94	0.71	0.73	0.30	0.08	0.13	0.35	0.16	0.17
4	15% (1)	0.93	0.68	0.73	0.29	0.07	0.13	0.33	0.13	0.16
3	50% (4)	0.94	0.69	0.72	0.26	0.02	0.13	0.30	0.09	0.15
3	40% (3)	0.93	0.67	0.72	0.27	0.04	0.13	0.31	0.11	0.16
3	33% (2)	0.94	0.72	0.73	0.30	0.08	0.13	0.34	0.14	0.16
3	25% (1)	0.94	0.69	0.73	0.28	0.05	0.12	0.31	0.10	0.15
3	15% (1)	0.94	0.69	0.73	0.28	0.05	0.12	0.31	0.10	0.15

GKSS: SNR=0.4, long calibration, white noise network **A**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	Mult RE (NH)	Mult CE (NH)	Mult r^2 (NH)
3	50% (4)	0.96	0.86	0.87	0.33	-0.01	0.18	0.36	0.07	0.19
3	40% (3)	0.96	0.86	0.87	0.34	0.01	0.20	0.38	0.09	0.22
3	33% (2)	0.96	0.88	0.88	0.37	0.05	0.22	0.41	0.14	0.25
3	25% (2)	0.96	0.88	0.88	0.37	0.05	0.22	0.41	0.14	0.25
3	15% (1)	0.96	0.86	0.88	0.37	0.05	0.22	0.40	0.12	0.25
2	50% (4)	0.96	0.87	0.88	0.33	-0.01	0.19	0.34	0.04	0.21
2	40% (3)	0.96	0.86	0.87	0.33	-0.01	0.20	0.34	0.04	0.22
2	33% (2)	0.96	0.87	0.88	0.33	-0.01	0.21	0.34	0.04	0.22
2	25% (1)	0.96	0.87	0.89	0.35	0.02	0.22	0.37	0.07	0.24
2	15% (1)	0.96	0.87	0.89	0.35	0.02	0.22	0.37	0.07	0.24

CSM: SNR=0.4, long calibration, white noise network **B**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	Mult RE (NH)	Mult CE (NH)	Mult r^2 (NH)
2	50% (3)	0.81	0.07	0.29	0.15	-0.12	0.04	0.17	-0.08	0.04
2	40% (2)	0.83	0.20	0.36	0.18	-0.08	0.05	0.19	-0.05	0.05
2	33% (2)	0.83	0.20	0.36	0.18	-0.08	0.05	0.19	-0.05	0.05
2	25% (2)	0.83	0.20	0.36	0.18	-0.08	0.05	0.19	-0.05	0.05
2	15% (1)	0.84	0.21	0.35	0.17	-0.09	0.05	0.19	-0.06	0.05
1	50% (3)	0.82	0.15	0.32	0.16	-0.11	0.04	0.15	-0.10	0.04
1	40% (2)	0.83	0.20	0.36	0.17	-0.09	0.05	0.17	-0.08	0.05
1	33% (2)	0.83	0.20	0.36	0.17	-0.09	0.05	0.17	-0.08	0.05
1	25% (2)	0.83	0.20	0.36	0.17	-0.09	0.05	0.17	-0.08	0.05
1	15% (1)	0.83	0.20	0.35	0.17	-0.09	0.05	0.16	-0.09	0.05

GKSS: SNR=0.4, long calibration, white noise network **B**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	MultRE (NH)	Mult CE (NH)	Mult r^2 (NH)
2	50% (3)	0.86	0.53	0.59	0.25	-0.13	0.13	0.27	-0.06	0.14
2	40% (2)	0.91	0.69	0.70	0.27	-0.09	0.14	0.30	-0.02	0.16
2	33% (2)	0.91	0.69	0.70	0.27	-0.09	0.14	0.30	-0.02	0.16
2	25% (1)	0.91	0.70	0.70	0.29	-0.06	0.14	0.31	-0.01	0.16
2	15% (1)	0.91	0.70	0.70	0.29	-0.06	0.14	0.31	-0.01	0.16
1	50% (3)	0.86	0.56	0.59	0.28	-0.08	0.14	0.29	-0.04	0.14
1	40% (2)	0.89	0.63	0.64	0.29	-0.07	0.14	0.30	-0.02	0.15
1	33% (2)	0.89	0.63	0.64	0.29	-0.07	0.14	0.30	-0.02	0.15
1	25% (1)	0.90	0.69	0.69	0.30	-0.05	0.15	0.31	-0.01	0.16
1	15% (1)	0.90	0.69	0.69	0.30	-0.05	0.15	0.31	-0.01	0.16

CSM: SNR=0.4, long calibration, white noise network **D (208 proxies)**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	Mult RE (NH)	Mult CE (NH)	Mult r^2 (NH)
4	50% (4)	0.92	0.63	0.66	0.30	0.08	0.19	0.33	0.13	0.21
4	40% (3)	0.93	0.65	0.67	0.31	0.09	0.19	0.33	0.13	0.21
4	33% (3)	0.93	0.65	0.67	0.31	0.09	0.19	0.33	0.13	0.21
4	25% (2)	0.93	0.68	0.70	0.28	0.05	0.18	0.30	0.09	0.20
4	15% (1)	0.96	0.78	0.79	0.31	0.09	0.19	0.34	0.14	0.22

3	50% (4)	0.92	0.61	0.66	0.30	0.08	0.19	0.32	0.12	0.20
3	40% (3)	0.93	0.64	0.67	0.31	0.09	0.19	0.33	0.13	0.21
3	33% (3)	0.93	0.64	0.67	0.31	0.09	0.19	0.33	0.13	0.21
3	25% (2)	0.94	0.69	0.71	0.26	0.03	0.17	0.26	0.05	0.18
3	15% (1)	0.96	0.78	0.79	0.29	0.06	0.18	0.30	0.10	0.20

GKSS: SNR=0.4, long calibration, white noise network **D (208 proxies)**

# inst PCs	Low-f cutoff Pct (*K*)	NH *RE*	NH *CE*	NH r^2	Mult *RE* (Glb)	Mult CE (Glb)	Mult r^2 (Glb)	Mult RE (NH)	Mult CE (NH)	Mult r^2 (NH)
2	50% (4)	0.91	0.69	0.84	0.38	0.07	0.27	0.40	0.12	0.28
2	40% (3)	0.96	0.88	0.90	0.41	0.11	0.27	0.46	0.19	0.30
2	33% (2)	0.97	0.89	0.91	0.41	0.11	0.27	0.44	0.19	0.30
2	25% (1)	0.97	0.91	0.93	0.41	0.11	0.27	0.43	0.17	0.30
2	15% (1)	0.97	0.91	0.93	0.41	0.11	0.27	0.43	0.17	0.30
1	50% (4)	0.91	0.70	0.85	0.35	0.02	0.25	0.35	0.06	0.26
1	40% (3)	0.97	0.90	0.92	0.37	0.05	0.25	0.38	0.09	0.27s
1	33% (2)	97	0.89	0.91	0.35	0.03	0.24	0.36	0.07	0.26
1	25% (1)	0.97	0.91	0.93	0.37	0.06	0.26	0.39	0.11	0.27
1	15% (1)	0.97	0.91	0.93	0.37	0.06	0.26	0.39	0.11	0.27

Table S3: Parameter choices for RegEM proxy temperature reconstructions. Provided are the # of Instrumental PCs (*M*), and RegEM TTLS truncation parameters for low-frequency (f<0.05 cycle/year) ('*K*') and high-frequency (f>0.05 cycle/year) ('K_{high}') bands, as a function of the network used over each period. Note that the number of spatial degrees of freedom in the reconstruction for each of the two frequency bands is the smaller of *M* and the value of *K* associated with the frequency band in question.

	All proxy reconstruction				**Screened proxy reconstruction**		
interval	M	*K*	K_{high}	interval	M	*K*	K_{high}
1800-1855	7	3	6	1800-1855	6	3	6
1700-1799	6	3	5	1700-1799	6	2	6
1600-1699	6	3	6	1600-1699	7	2	7
1500-1599	5	2	6	1500-1599	4	2	7
1400-1499	4	2	7	1400-1499	3	2	8
1300-1399	3	2	6	1300-1399	3	1	6
1200-1299	3	2	6	1200-1299	2	1	4
1100-1199	3	2	6	1100-1199	2	1	4
1000-1099	3	2	6	1000-1099	2	1	3
900-999	3	2	6	900-999	3	1	3
800-899	2	2	3	800-899	2	1	2
700-799	2	2	3	700-799	2	1	2
600-699	2	2	3	600-699	2	1	2
500-599	2	2	3	500-599	2	1	2

Table S4: Decadal multivariate and domain-mean (hemispheric and global mean) validation skill scores. 95% significance levels from Monte Carlo simulations are indicated at bottom in red. Note that hemispheric and global mean series are defined as averages of all available gridboxes as shown in Fig. S2.

Skill scores for all proxy reconstruction

	NH mult		SH mult		Global mult		NH mean		SH mean		Global mean	
Interrval	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE
1800-1849	0.36	-0.85	-0.01	-1.24	0.25	-1.00	0.54	0.14	0.52	0.14	0.53	0.14
1700-1799	0.33	-0.96	-0.35	-1.99	0.12	-1.34	0.54	0.12	0.50	0.05	0.52	0.09
1600-1699	0.30	-1.03	-0.81	-3.01	-0.04	-1.76	0.52	0.08	0.47	0.02	0.50	0.06
1500-1599	0.27	-1.11	-0.54	-2.42	0.02	-1.59	0.54	0.07	0.49	0.02	0.51	0.05
1400-1499	0.25	-1.17	-0.33	-1.95	0.07	-1.46	0.48	0.04	0.50	0.02	0.52	0.03
1300-1399	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.42	-0.18	0.45	-0.20	0.44	-0.19
1200-1299	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.42	-0.18	0.45	-0.20	0.44	-0.19
1100-1199	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.42	-0.18	0.45	-0.20	0.44	-0.19
1000-1099	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.42	-0.18	0.45	-0.20	0.44	-0.19
900-999	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.42	-0.18	0.45	-0.20	0.44	-0.19
800-899	0.19	-1.34	-0.01	-1.24	0.13	-1.30	0.41	-0.20	0.43	-0.13	0.42	-0.16
700-799	0.19	-1.34	-0.01	-1.24	0.13	-1.30	0.41	-0.20	0.43	-0.13	0.42	-0.16
600-699	0.19	-1.34	-0.01	-1.24	0.13	-1.30	0.41	-0.20	0.43	-0.13	0.42	-0.16
500-599	0.19	-1.34	-0.01	-1.24	0.13	-1.30	0.41	-0.20	0.43	-0.13	0.42	-0.16
95% sig.	0.06	-1.86	-0.04	-1.79	0.02	-1.83	-0.02	-0.98	0.01	-0.91	-0.01	-0.95

Skill scores for screened reconstruction

	NH mult		SH mult		Global mult		NH mean		SH mean		Global mean	
Interval	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE
1800-1849	0.36	-0.85	-0.01	-1.24	0.25	-1.00	0.54	0.13	0.53	0.18	0.54	0.15
1700-1799	0.33	-0.96	-0.35	-1.99	0.12	-1.34	0.54	0.13	0.53	0.18	0.54	0.15
1600-1699	0.30	-1.03	-0.81	-3.01	-0.04	-1.76	0.54	0.13	0.52	0.13	0.53	0.13
1500-1599	0.27	-1.11	-0.54	-2.42	0.02	-1.59	0.53	0.08	0.52	0.15	0.52	0.11
1400-1499	0.25	-1.17	-0.33	-1.95	0.07	-1.46	0.52	0.07	0.51	0.17	0.52	0.12
1300-1399	0.20	-1.33	-0.02	-1.26	0.13	-1.30	0.51	0.06	0.51	0.09	0.51	0.07
95% sig.	0.06	-1.86	-0.04	-1.79	0.02	-1.83	-0.02	-0.98	0.01	-0.91	-0.01	-0.95

Multivariate PC Filtered Scores

		NH		SH		Global	
Interval	PCs	RE	CE	RE	CE	RE	CE
1800-1849	7	0.59	0.29	0.56	0.25	0.58	0.28
1700-1799	6	0.53	0.20	0.52	0.19	0.53	0.19
1600-1699	6	0.53	0.20	0.52	0.19	0.53	0.19
1500-1599	5	0.50	0.14	0.48	0.12	0.49	0.14
1400-1499	4	0.47	0.09	0.45	0.07	0.46	0.08
1300-1399	3	0.34	-0.13	0.38	-0.05	0.35	-0.11
1200-1299	3	0.34	-0.13	0.38	-0.05	0.35	-0.11
1100-1199	3	0.34	-0.13	0.38	-0.05	0.35	-0.11
1000-1099	3	0.34	-0.13	0.38	-0.05	0.35	-0.11
900-999	3	0.34	-0.13	0.38	-0.05	0.35	-0.11
800-899	2	0.27	-0.25	0.34	-0.12	0.29	-0.21
700-799	2	0.27	-0.25	0.34	-0.12	0.29	-0.21
600-699	2	0.27	-0.25	0.34	-0.12	0.29	-0.21
500-599	2	0.27	-0.25	0.34	-0.12	0.29	-0.21
95% sig.		0.06	-1.86	-0.04	-1.79	0.02	-1.83

Table S5: Decadal validation skill scores for regional averages and indices. 95% significance levels from Monte Carlo simulations are indicated at bottom in red. Note that regional averages and indices are defined as spatial averages over available gridboxes in region passing validation over a given interval, as shown in Figure S4.

Skill Scores for Indices

Interval	AMO		Extratrop		Tropical		Niño3		PDO		NH Land		TAMDR		Global SST	
	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE	RE	CE
1800-1849	0.47	0.05	0.52	0.16	0.47	0.12	0.47	0.37	0.54	0.55	0.52	0.05	0.35	-0.17	0.49	0.16
1700-1799	0.52	0.08	0.53	0.19	0.46	0.02	0.43	0.35	0.50	0.46	0.55	0.07	0.25	-0.38	0.46	0.08
1600-1699	0.48	0.07	0.51	0.15	0.42	0.01	0.38	0.34	0.45	0.42	0.52	0.04	0.24	-0.28	0.42	0.05
1500-1599	0.53	-0.02	0.52	0.13	0.48	0.01	0.41	0.37	0.38	0.38	0.56	0.04	0.26	-0.37	0.45	0.04
1400-1499	0.46	-0.08	0.51	0.12	0.46	-0.01	0.32	0.13	0.43	0.40	0.54	0.04	0.15	-0.48	0.45	0.03
1300-1399	0.48	-0.03	0.47	0.09	0.49	0.10	0.40	0.29	0.39	0.43	0.50	0.03	0.34	-0.17	0.46	0.08
1200-1299	0.48	-0.03	0.47	0.09	0.49	0.10	0.40	0.29	0.39	0.43	0.50	0.03	0.34	-0.17	0.46	0.08
1100-1199	0.48	-0.03	0.47	0.09	0.49	0.10	0.40	0.29	0.39	0.43	0.50	0.03	0.34	-0.17	0.46	0.08
1000-1099	0.48	-0.03	0.47	0.09	0.49	0.10	0.40	0.29	0.39	0.43	0.50	0.03	0.34	-0.17	0.46	0.08
900-999	0.48	-0.03	0.47	0.09	0.49	0.10	0.40	0.29	0.39	0.43	0.50	0.03	0.34	-0.17	0.46	0.08
800-899	0.45	-0.09	0.48	0.13	0.46	0.00	0.39	0.25	0.45	0.45	0.51	0.00	0.16	-0.59	0.45	0.04
700-799	0.45	-0.09	0.48	0.13	0.46	0.00	0.39	0.25	0.45	0.45	0.51	0.00	0.16	-0.59	0.45	0.04
600-699	0.45	-0.09	0.48	0.13	0.46	0.00	0.39	0.25	0.45	0.45	0.51	0.00	0.16	-0.59	0.45	0.04
500-599	0.45	-0.09	0.48	0.13	0.46	0.00	0.39	0.25	0.45	0.45	0.51	0.00	0.16	-0.59	0.45	0.04
95% sig.	-0.05	-1.26	-0.05	-1.00	0.02	-1.00	0.18	-0.47	-0.19	-0.84	0.02	-0.98	-0.02	-0.85	-0.02	-0.91

Note: TAMDR= Tropical Atlantic Main Development Region.

Supporting references

S1. M.E. Mann *et al., Proc. Natl. Acad. Sci.* **105**, 13252 (2008).

S2. K.R. Briffa, P.D. Jones, F.H. Schweingruber , T.J. Osborn, *Nature* **393**, 450 (1998).

S3. K.R. Briffa *et al*., *J. Geophys Res. Atmos.* **106**, 2929 (2001).

S4. S. Rutherford *et al., J. Climate* **18**, 2308 (2005).

S5. J. Luterbacher, D. Dietrich, E. Xoplaki, M. Grosjean, H. Wanner, *Science* **303**, 1499 (2004).

S6. T. Schneider, *J. Climate* **14**, 853, 2001

S7. S. Rutherford, M.E. Mann, T.L. Delworth, R. Stouffer, *J. Climate* **16**, 462 (2003).

S8. P. Brohan, J.J. Kennedy, I. Harris, S.F.B. Tett, P.D. Jones, *J. Geophys. Res. Atmos.* **111**, D12106 (2006).

S9. M.E. Mann, S. Rutherford, E. Wahl, C. Ammann, *J. Geophys. Res. Atmos*. **112**, D12109 (2007).

S10. M.E. Mann, R.S. Bradley, M.K. Hughes, *Nature* **392**, 779 (1998).

S11. M.E. Mann, R.S. Bradley, M.K. Hughes, *Geophys. Res. Lett*. **26**, 759 (1999).

S12. G.A. Schmidt *et al.*, *J. Clim*. ***19***, 153 (2006).

S13. E. Bard, G. Raisbeck, F. Yiou, J. Jouzel, *Quat. Sci. Rev.* **26**, 2301 (2007).

S14. M. Korte, C.G. Constable, *Earth Plan. Sci. Lett.* **236**, 348 (2005).

S15. D.T. Shindell *et al., Geophys. Res. Lett.*, *33, L24706*, doi:10.1029/2006GL027468 (2006).

S16. Y.M. Wang, J. Lean, N. R. Sheeley, *Astrophysical Journal* **625**, 622 (2005).

S17. C.M. Ammann, F. Joos, D. Schimel, B. L. Otto-Bliesner, R. Tomas, *Proc. Natl. Acad. Sci.* **104**, 3713 (2007).