GLOBAL-SCALE TEMPERATURE PATTERNS AND CLIMATE FORCING OVER THE PAST SIX CENTURIES

Michael E. Mann & Raymond S. Bradley,
Department of Geosciences,
University of Massachusetts

Malcolm K. Hughes,
Laboratory of Tree Ring Research,
University of Arizona,
to appear in *Nature*

Reconstructions of annual global surface temperature patterns over several centuries are now possible, based on the multivariate calibration of widely distributed high-resolution proxy climate indicators. These reconstructions provide insight into both the spatial and temporal nature of climatic variations during the past six centuries. Time-dependent correlations of these temperature reconstructions with time series representing greenhouse gases, solar irradiance, and volcanic aerosols, suggest that each of these forcings has played a role in the climatic variability of the past several centuries, with greenhouse gases appearing to emerge as the dominant forcing during the 20th century. Northern hemisphere mean annual temperatures for three of the past eight years are warmer than any other year since (at least) 1400 AD, at a greater than 99.5% level of confidence.

Knowing both the spatial and temporal patterns of climatic change over the past several centuries remains a key to assessing a possible anthropogenic impact on post-industrial climate[1]. In addition to the possibility of warming due to enhanced greenhouse gases during the past century, there is evidence that both solar irradiance and explosive volcanism have played an important part in forcing climate variations over the past several centuries[2,3]. The unforced "natural variability" of the climate system may also be quite important on multidecadal and century timescales[4,5]. If a faithful empirical description of climate variability can be obtained for the past several centuries, a more confident estimation can be made of the roles of different external forcings and internal sources of variability on past and recent climate. Because widespread instrumental climate data are available for only about one century, we must use proxy climate indicators combined with any very long instrumental records that are available to obtain such an empirical description of large-scale climate variability during past centuries. A variety of studies have sought to employ a "multiproxy" approach to study long-term climatic variations, in which a widely-distributed set of proxy and instrumental climate indicators are analyzed[1,5–8] to yield insights into long-term global climatic variations. Building on such past studies, we take a new statistical approach to reconstructing global patterns of annual temperature back through the 15th century, based on the calibration of multiproxy data networks by the dominant patterns of temperature variability in the instrumental record.

Using these statistically-verifiable yearly global temperature reconstructions, we analyze the spatiotemporal patterns of climatic change over the past half-millenium, and then take an empirical approach to estimating the relationship between global temperature changes, variations in volcanic aerosols, solar irradiance, and greenhouse gas concentrations during the same period.

Data

We employ a multiproxy network consisting of widely distributed high-quality annual resolution proxy climate indicators individually collected and formerly analyzed by many paleoclimate researchers (details and references are available; see Supplementary information). The network includes (see Figure 1a) the collection of annual resolution dendroclimatic, ice core, ice melt, and long historical records used by Bradley and Jones[6] combined with other coral, ice core, dendroclimatic, and long instrumental records. The long instrumental records have been formed into annual mean anomalies relative to the 1902-1980 reference period, and gridded onto a 5 degree by 5 degree grid (yielding 11 temperature gridpoint series and 12 precipitation gridpoint series dating back to 1820 or earlier) similar to that shown in Figure 1 (bottom). Certain densely sampled regional dendroclimatic datasets have been represented in the network by a smaller number of leading principal components (typically 3-11 depending on the spatial extent and size of the dataset). This form of representation insures a reasonably homogeneous spatial sampling in the multiproxy network (112 indicators back to 1820).

Potential limitations specific to each type of proxy data series must be carefully taken into account in building an appropriate network. Dating errors in a given record (e.g., incorrectly assigned annual layers or rings) are particularly detrimental if mutual information is sought to describe climate patterns on a year-by-year basis. Standardization of certain biological proxy records relative to estimated growth trends, and the limits of constituent chronology segment lengths, (e.g. in dendroclimatic reconstructions) can restrict the maximum timescale of climatic variability that is recorded[9]. However, the dendroclimatic data used were carefully screened for conservative standardization and sizeable segment lengths. Moreover, the mutual information contained in a diverse and widely distributed set of independent climatic indicators can more faithfully capture the consistent climate signal that is present, reducing the compromising effects of biases and weaknesses in the individual indicators.

Monthly instrumental land air and sea surface temperature[10] gridpoint data (Figure 1–bottom) from the period 1902-1995 are used to calibrate the proxy dataset. Although there are notable spatial gaps, this network covers significant enough portions of the globe to form reliable estimates of Northern Hemisphere ("NH") mean temperature, and certain regional indices of particular importance such as the "NINO3" eastern tropical Pacific surface temperature index often used to describe the El Niño phenomenon. The NINO3 eastern equatorial Pacific sea surface temperature index is constructed from the 8 gridpoints available within the conventional NINO3 box shown between 5S to 5N, 90W to 150W.

Multiproxy Calibration

While studies have demonstrated that well-chosen regional paleoclimatic reconstructions can act as surprisingly representative surrogates for large-scale climate[11–13], multiproxy networks nonetheless, appear to provide the greatest opportunity for large-scale paleoclimate reconstruction[6] and climate signal detection[5,1]. There is a rich tradition of multivariate statistical calibration approaches to paleoclimate reconstruction, particularly in the field of dendroclimatology where the relative strengths and weaknesses of various approaches to multivariate calibration have been well-studied[14,15]. Such approaches have been applied to regional dendroclimatic networks to reconstruct regional patterns of temperature[16,17] and atmospheric circulation[18−20] or specific climate phenomena such as the Southern Oscillation[21]. Largely because of the inhomogeneity of the information represented by different types of indicators in a true "multiproxy" network, we found conventional approaches (e.g. Canonical Correlation Analysis or "CCA" of the proxy and instrumental datasets) to be relatively ineffective. Our approach to climate pattern reconstruction relates closely to statistical approaches which have recently been applied to the problem of filling in sparse early instrumental climate fields, based on calibration of the sparse sub-networks against the more widespread patterns of variability that can be resolved in shorter datasets[22,23]. We first decompose the 20th century instrumental data into its dominant patterns of variability, and subsequently calibrate the individual climate proxy indicators against the time histories of these distinct patterns during their mutual interval of overlap. One can think of the instrumental patterns as "training" templates against which we calibrate or "train" the much longer proxy data (i.e., the "trainee" data) during the shorter calibration period which they overlap. This calibration allows us to subsequently solve an "inverse problem" whereby best estimates of surface temperature patterns are deduced back in time before the calibration period, from the multiproxy network alone.

Implicit in our approach are at least three fundamental assumptions. We assume (1) that the indicators in our multiproxy trainee network are linearly related to one or more of the instrumental training patterns. In the relatively unlikely event that a proxy indicator represents a truly local climatic phenomenon which is uncorrelated with larger-scale climatic variations, or represents a highly non-linear response to climatic variations, this assumption will not be satisfied. We further assume (2) that a relatively sparse but widely distributed sampling of long proxy and instrumental records may nonetheless sample most of the relatively small number of degrees of freedom in climatic patterns at interannual and longer timescales. Regions not directly represented in the trainee network may nonetheless be indirectly represented through teleconnections with regions that are. The El Niño-Southern Oscillation (ENSO), for example, exhibits global-scale patterns of climatic influence[24], and is an example of a prominent pattern of variability which, if captured, can potentially describe variability in regions not directly sampled by the trainee data. Finally, we assume (3) that that the patterns of variability captured by the multiproxy network have analogues in the patterns we resolve in the shorter instrumental data. This latter assumption represents a fairly weak "stationarity" requirement–we don't require that the climate itself be stationary. In fact, we expect that some sizeable trends in the climate may be resolved by our reconstructions. We do, however, assume that the fundamental *spatial patterns* of variation which the climate has exhibited during the past century are similar to those by which it has varied during past recent centuries. Studies of instrumental surface temperature patterns suggest that such a form of stationarity holds up at least on multidecadal timescales, during the past century[23]. The statistical cross-validation exercises we describe later provide the best evidence that these key underlying assumptions hold.

We isolate the dominant patterns of the instrumental surface temperature data through principal component analysis[25] (PCA). PCA provides a natural smoothing of the temperature field in terms of a small number of dominant patterns of variability or "empirical eigenvectors". Each of these eigenvectors is associated with a characteristic spatial pattern or "Empirical Orthogonal Function" (EOF) and its characteristic evolution in time or "Principal Component" (PC). The ranking of the eigenvectors orders the fraction of variance they describe in the (standardized) multivariate data during the calibration period. The first 5 of these eigenvectors describe a fraction $\beta = 0.93$ (ie, 93%) of the of the global-mean ("GLB"), temperature variations, 85% of the northern hemisphere-mean ("NH") variations, 67% of the NINO3 index, and 76% of the non trend-related ("DETR") NH variance (see "Methods–statistics" section for a description of the β statistic used here as a measure of resolved variance). A sizeable fraction of the total multivariate spatiotemporal variance ("MULT") in the raw (instrumental) data (27%) is described by these 5 eigenvectors, or about 30% of the standardized variance (1=12%, 2=6.5%, 3=5%, 4=4%, 5=3.5%). Figure 2 shows the EOFS of the first 5 eigenvectors. The associated time histories of these patterns (PCs) and their reconstructed counterparts ("RPC"s) are discussed in the next section. The first eigenvector, associated with the significant global warming trend of the past century, describes much of the variability in the global (GLB=88%) and hemispheric (NH=73%) means. Subsequent eigenvectors, in contrast, describe much of the spatial variability relative to the large-scale means (ie, much of the remaining multivariate variance "MULT"). The second eigenvector is the dominant ENSO-related component, describing 41% of the variance in the NINO3 index. This eigenvector exhibits a modest negative trend which, in the eastern tropical Pacific, describes a "La Niña"-like cooling trend[26], which opposes warming in the same region associated with the global warming pattern of the first eigenvector. The third eigenvector is associated largely with interannual-to-decadal scale variability in the Atlantic basin and carries the well known temperature signature of the North Atlantic Oscillation (NAO)[27] and decadal tropical Atlantic dipole[28]. The fourth eigenvector describes a primarily multidecadal timescale variation with ENSO-scale and tropical/subtropical Atlantic features, while the fifth eigenvector is dominated by multidecadal variability in the entire Atlantic basin and neighboring regions that has been widely noted elsewhere[29–34].

We calibrate each of the indicators in the multiproxy data network against these empirical eigenvectors at annual mean resolution during the 1902-1980 training interval. While the seasonality of variability is potentially important– many extratropical proxy indicators for example reflect primarily warm-season variability[6,7]–we seek in the present study to resolve only annual mean conditions, exploiting the seasonal climatic persistence, and the fact that the mutual information from data reflecting various seasonal windows should provide complementary information regarding annual mean climatic conditions[10]. Following this calibration, we apply an overdetermined optimization procedure to determine the best combination of eigenvectors represented by the multiproxy network back in time on a year-by- year basis, with a spatial coverage dictated only by the spatial extent of the instrumental training data. From the reconstructed PCs or "RPC"s, spatial patterns and all relevant averages or indices can be readily determined. The details of the entire statistical approach are described in the "Methods–calibration" section.

The skill of the temperature reconstructions (ie, their statistical validity) back in time is established through a variety of complimentary independent cross-validation or "verification" exercises (see "Methods–verification" below). We summarize the key results of these experiments below (a Table detailing the quantitative results of the calibration and verification procedures is provided in *Natures's* on-line supplementary information):

1) In the reconstructions from 1820 onward based on the full multiproxy network of 112 indicators, 11 eigenvectors are skillfully resolved (# 1-5,7,9,11,14-16) describing ($\sim$70-80% of the variance in NH and GLB mean series in both calibration and verification (verification is based here on the independent 1854-1901 dataset which was witheld, as described in the "Methods–verification" section). Figure 3 shows the spatial patterns of calibration β, and verification β and the squared correlation statistic r^2, demonstrating highly significant reconstructive skill over widespread regions of the reconstructed spatial domain. 30% of the full spatiotemporal variance in the gridded dataset is captured in calibration, and 22% of the variance is verified in cross-validation.

Some of the degradation in the verification score relative to the calibration score may reflect the decrease in instrumental data quality in many regions before the 20th century rather than a true decrease in resolved variance. These scores thus compare favorably to the 40% total spatiotemporal variance that is described by simply filtering the raw 1902-1980 instrumental data with 11 eigenvectors used in calibration, suggesting that the multiproxy calibrations are describing a level of variance in the data reasonably close to the optimal "target" value. While a verification NINO3 index is not available from 1854-1901, correlation of the reconstructed NINO3 index with the available Southern Oscillation index (SOI) data from 1865-1901 of $r = -0.38$ ($r^2 = 0.14$) compares reasonably with its target value given by the correlation between the actual instrumental NINO3 and SOI index from 1902-1980 ($r = -0.72$). Furthermore, the correspondence between the reconstructed NINO3 index warm events and Quinn and Neal's[35] historical El Niño chronology back to 1820 (see "Methods–verification") is significant at the 98% level.

2)The calibrations back to 1760 based on 93 indicators, continue to resolve at least 9 eigenvectors (1-5,7,9,11,15) with no degradation of calibration or verification resolved variance in NH, and only slight degradation in MULT (calibration ~27%, verification ~17%). Our reconstructions are thus largely indistinguishable in skill back to 1760.

3)The network available back to 1700 of 74 indicators (including only 2 instrumental or historical indicators) skillfully resolves 5 eigenvectors (1,2,5,11,15) and shows some significant signs of decrease in reconstructive skill. In this case, about 60-70% of NH variance is resolved in calibration and verification, and about 14-18% of MULT in calibration, and 10-12% of MULT in verification. The verification r of NINO3 with the SOI is in the range of r ~-0.25 to -0.35, which is statistically significant (as is the correspondence with the Quinn historical chronology back to 1700) but notably inferior to the later calibrations. In short, both spatial patterns and large-scale means are skillfully resolved, but with significantly less resolved variance than in later calibrations.

4)The network of 57 indicators back to 1600 (including 1 historical record) skillfully resolves 4 eigenvectors (1,2,11,15). 67% of NH is resolved in calibration, and 53% in verification. 14% of MULT is resolved in calibration, and 12% of MULT in verification. A significant, but modest level of NINO3-scale variability is resolved in the calibrations.

5) The network of 24 proxy indicators back to 1450 resolves 2 eigenvectors (1,2) and ~40-50% of NH in calibration and verification. Only ~10% of MULT is resolved in calibration and ~5% in verification. There is no skillful reconstruction of ENSO-scale variability. Thus spatial reconstructions are of marginal usefulness this far back, though the largest-scale quantities are still skillfully resolved.

6) The multiproxy network of 22 indicators available back to 1400 resolves only the first eigenvector, associated with 40-50% of resolved variance in NH in calibration and verification. There is no useful resolution of spatial patterns of variability this far back. The sparser networks available before 1400 show little evidence of skill in reconstructing even the first eigenvector, terminating useful reconstruction at the initial year 1400 AD.

7) Experiments using trainee networks containing only proxy (i.e., no instrumental or historical) indicators establish the most truly independent cross-validation of the reconstruction since there is in this case neither spatial nor temporal dependence between the calibration and verification datasets. Such statistically significant verification is demonstrated at the gridpoint level (calibration and verification resolved variance~15% for the "MULT" statistic), at the largest scales (calibration and verification resolved variance ~60-65% for NH) and the ENSO-scale (90-95% statistical significance for all verification diagnostics). In contrast, networks containing only the 24 long historical or instrumental records available back to 1820 resolve only ~30% of NH in calibration or verification, and the modest multivariate calibration and verification resolved variance scores of MULT (~10%) are artificially inflated by the high degree of spatial correlation between the instrumental "multiproxy" predictor and instrumental predictand data. No evidence of skillful ENSO-scale reconstruction is evident in these latter reconstructions.

In short, the inclusion of the proxy data in the "multiproxy" network is essential for the most skillful reconstructions. Certain sub-components of the proxy dataset however (e.g., the dendroclimatic indicators) appear to be especially important in resolving the large-scale temperature patterns, with notable decreases in the NH scores obtained if all dendroclimatic indicators are witheld from the multiproxy network. On the other hand, the long-term trend in NH is relatively robust to the inclusion of dendroclimatic indicators in the network, suggesting that potential tree growth trend biases are not influential in multiproxy pattern reconstructions. The network of all combined proxy and long instrumental/historical indicators provide the greatest cross-validated estimates of skillful reconstruction, and are used in obtaining the reconstructions described below.

Temperature Reconstructions

The reconstructions henceforth discussed are derived using all indicators available, and using the optimal eigenvector subsets determined in the calibration experiments described above (11 from 1780-1980, 9 from 1760-1779, 8 from 1750-1759, 5 from 1700-1749, 4 from 1600-1699, 2 from 1450-1599, 1 from 1400-1449). To better illustrate the workings and effectiveness of the proxy pattern reconstruction procedure, we show as an example (Figure 4) the actual, EOF-filtered, and reconstructed temperature patterns for a year (1941) during the calibration interval. This year is a known ENSO year, associated with a warm eastern tropical Pacific, and a cold central North Pacific. Pronounced cold anomalies are also found over large-parts of Eurasia. The proxy-reconstructed pattern captures these features, although in a relatively smoothed sense (describing about 30% of the full variance in that pattern), and is remarkably similar to the raw data once it has been filtered by retaining only the 11 eigenvectors (#s 1-5,7,9,11,14-16) used in pattern reconstruction. It is thus visually apparent that the multiproxy network is quite capable of resolving much of the structure resolved by the eigenvectors retained in the calibration process.

It is interesting to consider the temporal variations in the first 5 reconstructed PCs (Figure 5a). The positive trend in RPC #1 during the 20th century is clearly exceptional in the context of the long-term variability in the associated eigenvector, and indeed describes much of the unprecedented warming trend evident in the NH reconstruction. Equally interesting is the negative trend in RPC # 2 during the past century which is also anomalous in the context of the longer-term evolution of the associated eigenvector. The recent negative trend is associated with a pattern of cooling in the eastern tropical Pacific (superimposed on warming associated with the pattern of eigenvector #1) which may be a modulating negative feedback on global warming[26]. RPC #5 exhibits notable multidecadal variability throughout both the modern and pre-calibration interval, associated with the wavelike trend of warming and subsequent cooling of the North Atlantic this century discussed earlier[29–33] and the longer-term multidecadal oscillations in that region detected in a previous analysis of proxy climate networks[5]. This variability may be associated with ocean-atmosphere processes related to the North Atlantic thermohaline circulation[34,4].

The long-term trends in the reconstructed annual mean NH series (Figure 5b) are quite similar to those of the decadal Northern Hemisphere summer temperature reconstructions of Bradley and Jones[6], exhibiting pronounced cold periods during the mid 17th and 19th centuries, and somewhat warmer intervals during the mid 16th and late 18th centuries, with almost all years prior to the 20th century well below the 20th century climatological mean. Taking into account the uncertainties in our NH reconstruction (see "Methods–calibration"), it appears that the years 1990, 1995 and now 1997 (this value recently calculated and not shown) each exhibit anomalies that are greater than any other year back to 1400 at 3 standard errors, or, roughly a 99.7% level of certainty. It should be kept in mind that hemispheric mean values are not associated with globally or hemispherically uniform trends. An example of the global pattern for an historically documented[35] "very strong" El Niño year (1791) is shown in Figure 6a demonstrating the classic warm eastern tropical Pacific and cold central North Pacific SST patterns. Analysis of ENSO variability in these reconstructions is discussed in more detail elsewhere[36]. The reconstructed pattern for 1816 is also shown (Figure 6b) exhibiting quite anomalous cold throughout much of the northern hemisphere (even relative to this generally cold decade) but with a quadrapole pattern of warmth near Newfoundland and the Near East, and enhanced cold in the eastern United States and Europe consistent with the anomalous atmospheric circulation associated with the North Atlantic Oscillation (NAO) pattern. Such a pattern is indeed observed in empirical[37] and model-based studies[38] of the atmospheric response to volcanic forcing. We infer in the 1816 temperature pattern a climatic response to the explosive Tambora eruption of April 1815 based on both the anomalous hemispheric coolness and the superimposed NAO-like pattern. Reconstructed time series RPCs #1-5, the NH series, NINO3 index and reconstructions for specific gridpoints can be obtained through the NOAA paleoclimatology web site *http://www.ngdc.noaa.gov/paleo/paleo.html*.

Attribution of Forcings

Finally, we take an empirical approach to detecting the possible effects of external forcings on the climate. We take the reconstructed NH series as a diagnostic of the global climate, and examine its relationship with three candidate external forcings during the period 1610-1995 including (i) CO_2 measurements[39] as a proxy for total greenhouse gas changes, (ii) reconstructed solar irradiance variations[2] and (iii) the weighted historical dust veil index (DVI) of explosive volcanism used by Jones and Bradley[40] (see Figure 31.1 therein) updated with recent data[41]. While we caution the reader that historical series for these forcing agents are imperfectly known or measured, they do nonetheless represent our best estimates of the time-histories of the corresponding forcings. The reader is referred to more detailed discussions of the estimation, and potential sources of uncertainty or bias in these series[2,39,40]. Industrial aerosol forcing of the climate has also been suggested as an important forcing of recent climate[42−43], but its physical basis is still controversial[44], and difficult to estimate observationally. Noting that in any case, this forcing is not believed to be important before about 1940, its omission should be inconsequential in our long-term detection approach. Our empirical signal detection is complementary to that of model-based "fingerprint" signal detection studies[42,43,45]; while our empirical approach relies on the faithfulness of the reconstructed forcing series and on the assumption of a linear and contemporaneous response to forcings, it does not suffer the potential weaknesses of incomplete representations of internal feedback processes[26], poorly constrained parameterizations of climatic responses[44], and underestimated natural variability[1] in model-based studies. To the extent that the response to forcing is not contemporaneous, but rather is delayed owing to the inertia of the slow-response components of the climate system (e.g., the ocean and cryosphere), our detection approach will tend to underestimate the response to forcings, making the approach a relatively conservative one.

We estimate the climatic response to the three forcings based on an evolving multivariate regression method (Figure 7). This time-dependent correlation approach generalizes on previous studies of (fixed) correlations between long-term northern hemisphere temperature records and possible forcing agents[2,3]. Normalized regression (that is, correlation) coefficients r are simultaneously estimated between each of the three forcing series and the NH series from 1610 to 1995 in a 200 year moving window. The first calculated value centered at 1710 is based on data from 1610 to 1809, and the last value, centered at 1895, is based on data from 1796 to 1995–i.e., the most recent 200 years. A 200 year window width was chosen to insure that any given window contains enough samples to provide good signal-to-noise ratios in correlation estimates. Nonetheless, all of the important conclusions drawn below are robust to choosing other reasonable (e.g., 100 year) window widths.

We test the significance of the correlation coefficients (r) relative to a null hypothesis of random correlation arising from natural climate variability, taking into account the reduced degrees of freedom in the correlations owing to substantial trends and low-frequency variability in the NH series. The reduced degrees of freedom are modeled in terms of first-order Markovian "red noise" correlation structure of the data series, described by the lag-one autocorrelation coefficient ρ during a 200 year window. This parameter ranges from $\rho = 0.48$ in the first window (1610-1809) to $\rho = 0.77$ in the final window (1796-1995) of the moving correlation, the considerably larger recent value associated with the substantial global warming trend of the past century. This latter trend has been shown to be inconsistent with red noise[46] and could thus itself be argued as indicative of externally forced variability. An argument could, in this sense, be made for using the smaller pre-industrial value $\rho = 0.48$ of the NH series in estimating the statistical degrees of freedom appropriate for the null hypothesis of natural variability. Nonetheless, we make the conservative choice of adopting the largest value $\rho = 0.77$ as representative of the natural serial correlation in the series. We employ Monte Carlo simulations to estimate the likelihood of chance spurious correlations of such serially correlated noise with each of the 3 actual forcing series. The associated confidence limits are approximately constant between sliding 200 year windows. For (positive) correlations with both CO_2 and solar irradiance, the confidence levels are both approximately 0.24 (90%), 0.31 (95%), 0.41 (99%), while for the "whiter", relatively trendless DVI index, the confidence levels for (negative) correlations are somewhat lower (-0.16, -0.20, -0.27 respectively). A one-sided significance test is employed in each case because the physical nature of the forcing dictates a unique expected sign to the correlations (positive for CO_2 and solar irradiance variations, negative for the DVI fluctuations).

The correlation statistics indicate highly significant detection of solar irradiance forcing in the NH series during the "Maunder Minimum" of solar activity from the mid 17th to early 18th century which corresponds to an especially cold period. In turn, the steady increase in solar irradiance from the early 19th century through the mid 20th century coincides with the general warming over the period, exhibiting peak correlation during the mid 19th century. The regression against solar irradiance indicates a sensitivity to changes in the "solar constant" of about $0.1\mathrm{K}/(\mathrm{W/m}^2)$, which is consistent with recent model-based studies[42]. Greenhouse (CO_2) forcing, on the other hand, shows no sign of significance until a large positive correlation sharply emerges as the moving window slides into the 20th century. The partial correlation with CO_2 indeed dominates over that of solar irradiance for the most recent 200 year interval, as increases in temperature and CO_2 simultaneously accelerate through 1995, while solar irradiance levels off after the mid 20th century. It is reasonable to infer that greenhouse gas forcing is now the dominant external forcing of the climate system. Explosive volcanism exhibits the expected marginally significant negative correlation with temperature during much of 1610-1995 period, most pronounced in the 200 year window centered near 1830 which includes the most explosive volcanic events.

A variety of general circulation[42,47] and energy balance model experiments[43,45,48] as well as statistical comparisons of 20th century global temperatures with forcing series[49] suggest that, although both solar and greenhouse gas forcings play some role in explaining 20th century climate trends, greenhouse gases appear to play an increasingly dominant role during the 20th century. Such a proposition is consistent with the results of this study.

As larger numbers of high quality proxy reconstructions become available in diverse regions of the globe it may be possible to assimilate a more globally representative multiproxy data network. Given the high level of skill possible in large-scale reconstruction back to 1400 with the present network, it is reasonable to hope that it may soon be possible to faithfully reconstruct mean global temperatures back over the entire millenium, resolving for example the enigmatic medieval period[7]. Geothermal measurements from boreholes[50] recover long-term temperatures trends without many of the complications of traditional proxy indicators and, in combination with traditional multiproxy networks, may prove helpful in better resolving trends over many centuries. With a better knowledge of how the climate has varied prior to the 20th century, we will be able to place even better constraints on the importance of natural and anthropogenic factors governing the climate of the past few centuries, factors which will no doubt continue to affect climate variability in the future, in addition to any anthropogenic effects.

Methods

Statistics We use as our primary diagnostic of calibration and verification reconstructive skill the conventional "resolved variance" statistic,

$$\beta = 1 - \sum (y_{ref} - \hat{y})^2 \,/\, \sum y_{ref}^2$$

where y_{ref} is the reference series (e.g., the raw data in the case of calibration or the verification dataset in the case of verification) and $\hat{y}$ is the series being compared to it (e.g., the proxy-reconstructed data for either calibration or verification). We compute β for each gridpoint, and for the NH, GLB, and MULT quantities. The sum extends over the time interval of comparison, and for the multivariate case (MULT), over all gridpoints as well. We also computed a calibration β statistic for the detrended NH series (DETR) to distinguish between explanatory variance associated with the noteable trend of the 20th century, and that related to departures from the trend.

β is a quite rigorous measure of the similarity between two variables, measuring their correspondence not only in terms of the relative departures from mean values (as does the correlation coefficient r) but also in terms of the means and absolute variance of the two series. For comparison correlation (r) and squared-correlation (r^2) statistics are also determined. The expectation value for two random series is $\beta = -1$. Negative values of β may in fact be statistically significant for sufficient temporal degrees of freedom. Nonetheless, the threshold $\beta = 0$ defines the simple "climatological" model in which a series is assigned its long-term mean. In this sense, statistically significant negative values of β might still be considered questionable in their predictive or reconstructive skill. Due to the more rigorous "match" between two series sought by β, highly significant values of β are possible even when r^2 is only marginally significant.

Significance levels were determined for r^2 from standard one-sided tables, accounting for decreased degrees of freedom due to serial correlation. Significance levels for β were estimated by Monte Carlo simulations, also taking serial correlation into account. Serial correlation is assumed to follow from the null model of AR(1) red noise, and degrees of freedom are estimated based on the lag-one autocorrelation coefficients (ρ) for the two series being compared. While the values of ρ differ from gridpoint-to-gridpoint, this variation is relatively small, making it simplest to use the ensemble average values of ρ over the domain ($\rho \approx 0.2$).

Calibration With the spatial sampling of $M = 1082$ continuous monthly gridpoint surface temperature anomaly (ie, de-seasonalized) temperature data used (Figure 1b), the $N = 1128$ months of data available from 1902-1995 were sufficient for a unique, overdetermined eigenvector decomposition (note that $N' = 94$ years of the annual mean data would, in contrast, not be sufficient).

For each gridpoint, the mean was removed, and the series was normalized by its standard deviation. A standardized data matrix $\underline{\mathbf{T}}$ of the data is formed by weighting each gridpoint by the cosine of its central latitude to ensure areally-proportional contributed variance, and a conventional Principal Component Analysis (PCA) is performed,

$$\underline{\mathbf{T}} = \Sigma_{k=1}^{K} \lambda_k \mathbf{u_k}^{\dagger} \mathbf{v_k}$$

decomposing the dataset into its dominant spatiotemporal eigenvectors. The M-vector or empirical orthogonal function (EOF) $\mathbf{v}_k$ describes the relative spatial pattern of the kth eigenvector, the N-vector $\mathbf{u}_k$ or principal component (PC) describes its variation over time, and the scalar λ_k describes the associated fraction of resolved (standardized and weighted) data variance.

In a given calibration exercise, we retain a specified subset of the annually-averaged eigenvectors, the annually-averaged PCs denoted by $\overline{u}^k{}_n$, where $n = 1, \ldots, \overline{N}$, $\overline{N} = 79$ is the number of annual averages used of the N month length dataset. In practice, only a small subset N_{eofs} of the highest rank eigenvectors turn out to be usefull in these exercises from the standpoint of verifiable reconstructive skill. An objective criterion was used to determine the particular set of eigenvectors which should be used in the calibration as follows. Preisendorfer's[25] selection rule "rule N" was applied to the multiproxy network to determine the approximate number N_{eofs} of significant independent climatic patterns that are resolved by the network, taking into account the spatial correlation within the multiproxy dataset. Because the ordering of various eigenvectors in terms of their prominence in the instrumental data, and their prominence as represented by the multiproxy network, need not be the same, we allowed for the selection of non-contiguous sequences of the instrumental eigenvectors. We chose the optimal group of N_{eofs} eigenvectors, from among a larger set (e.g., the first 16) of the highest-rank eigenvectors, as the group of eigenvectors which maximized the calibration explained variance. It was encouraging from a consistency standpoint that this subset typically corresponded quite closely to the subset which maximized the *verification* explained variance statistics (see below), but the objective criterion was, as it should be, independent of the verification process. We emphasize, furthermore, that statistical significance was robustly established, as neither the measures of statistical skill nor the reconstructions themselves were highly sensitive to the precise criterion for selection. In addition to the above means of cross-validation, we also tested the network for sensitivity to the inclusion or elimination of particular trainee data (e.g., instrumental/historical records, non-instrumental/historical records, or dendroclimatic proxy indicators).

These N_{eofs} eigenvectors were trained against the N_{proxy} indicators, by finding the least-squares optimal combination of the N_{eofs} PCs represented by each individual proxy indicator during the $\overline{N} = 79$ year training interval from 1902-1980 (the training interval is terminated at 1980 because many of the proxy series terminate at or shortly after 1980). The proxy series and PCs were formed into anomalies relative to the same 1902-1980 reference period mean, and the proxy series were also normalized by their standard deviations during that period. This proxy-by-proxy calibration is well posed (i.e., a unique optimal solution exists) as long as $\overline{N} > N_{eofs}$ (a limit never approached in this study) and can be expressed as the least-squares solution to the overdetermined matrix equation,

$$\underline{\mathbf{U}}\mathbf{x} = \mathbf{y^{(p)}}$$

$$\underline{\mathbf{U}} = \begin{bmatrix} \overline{u}^{(1)}{}_1 & \overline{u}_1^{(2)} & \ldots & u_1^{(N_{eofs})} \\ \overline{u}^{(1)}{}_2 & \overline{u}_2^{(2)} & \ldots & u_2^{(N_{eofs})} \\ \vdots & & & \\ \overline{u}^{(1)}{}_{\overline{N}} & \overline{u}_{\overline{N}}^{(2)} & \ldots & u_{\overline{N}}^{(N_{eofs})} \end{bmatrix}$$

is the matrix of annual PCs, and

$$\mathbf{y^{(p)}} = \begin{bmatrix} y^{(p)}{}_1 \\ y^{(p)}{}_2 \\ \vdots \\ y^{(p)}{}_{\overline{N}} \end{bmatrix}$$

is the time series $\overline{N}-$vector for proxy record p.

The N_{eofs}-length solution vector $\mathbf{x} = \mathbf{G^{(p)}}$ is obtained by solving the above overdetermined optimization problem by singular value decomposition (SVD) for each proxy record $p = 1, \ldots, P$. This yields a matrix of coefficients relating the different proxies to their closest linear combination of the N_{eofs} PCs,

$$\underline{\mathbf{G}} = \begin{bmatrix} G^{(1)}{}_1 & G^{(1)}{}_2 & \ldots & G^{(1)}_{N_{eofs}} \\ G^{(2)}{}_1 & G^{(2)}{}_2 & \ldots & G^{(2)}_{N_{eofs}} \\ \vdots & & & \\ G^{(P)}{}_1 & G^{(P)}{}_2 & \ldots & G^{(P)}_{N_{eofs}} \end{bmatrix}$$

This set of coefficients will not provide a single consistent solution, but rather represents an overdetermined relationship between the optimal weights on each of the N_{eofs} PCs and the multiproxy network.

Proxy-reconstructed patterns are thus obtained during the pre-calibration interval by the year-by-year solution of the overdetermined matrix equation,

$$\underline{\mathbf{G}}\mathbf{z} = \mathbf{y_{(j)}}$$

where $\mathbf{y_{(j)}}$ is the predictor vector of values of each of the P proxy indicators during year j. The predictand solution vector $\mathbf{z} = \widehat{\mathbf{U}}$ contains the least-squares optimal values of each of the N_{eofs} PCs for a given year. This optimization is overdetermined (and thus well constrained) as long as $P > N_{eofs}$ which is always realized in this study. It is noteworthy that, unlike conventional paleoclimate transfer function approaches, there is no specific relationship between a given proxy indicator and a given predictand (ie, reconstructed PC). Instead, the best common choice of values for the small number of N_{eofs} predictands is determined from the mutual information present in the multiproxy network during any given year. The reconstruction approach is thus relatively resistant to errors or biases specific to any small number of indicators during a given year.

This yearly reconstruction process leads to annual sequences of the optimal reconstructions of the retained PCs, which we term the reconstructed principal components or "RPCs" and denote by $\hat{\mathbf{u}}^{\mathbf{k}}$. Once the RPCS are determined, the associated temperature patterns are readily obtained through the appropriate eigenvector expansion,

$$\widehat{\underline{\mathbf{T}}} = \Sigma_{k=1}^{N_{eofs}} \lambda_k \hat{\mathbf{u}}_{\mathbf{k}}^{\dagger} \mathbf{v_k}$$

while quantities of interest (e.g., NH) are calculated from the appropriate spatial averages, and appropriate calibration and verification resolved variance statistics are calculated from the raw and reconstructed data.

Several checks were performed to insure a reasonably unbiased calibration procedure. The histograms of calibration residuals were examined for possible heteroscedasticity, but were found to pass a χ^2 test for Gaussian distributedness at reasonably high levels of significance (NH: 95% level, NINO3: 99% level). The spectrum of the calibration residuals for these quantities, were, furthermore, found to be approximately "white", showing little evidence for preferred or deficiently resolved timescales in the calibration process. Having established reasonably unbiased calibration residuals, we were able to calculate uncertainties in the reconstructions by assuming that the unresolved variance is Gaussian distributed over time. This variance increases back in time (the increasingly sparse multiproxy network calibrates smaller fractions of variance), yielding error bars which expand back in time.

Verification Verification resolved variance statistics (β) were determined based on two distinct verification datasets including (a) the sparse subset of the gridded data ($M' = 219$ gridpoints) for which independent values are available from 1854-1901 (see Figure 1b) and (b) the small subset of 11 very long instrumental estimated temperature gridpoint averages (10 in Eurasia, 1 in North America–see Figure 1a) constructed from the longest available station measurements. Each of these "gridpoint" series shared at least 70% of their variance with the corresponding temperature gridpoint available from 1854-1980, providing verification back to at least 1820 in all cases (and back through the mid and early 18th century in many cases). Note that this latter verification dataset is only temporally but not spatially independent of the multiproxy network itself, which contains these long instrumental gridpoint series as a small subset of the network. In case (a), NH and GLB verification statistics are computed as well as the multivariate ("MULT") gridpoint level verification statistic, although these quantities represent different spatial samplings than in the full calibration dataset owing to the sparser sampling of the verification period. Case (b) provides a longer-term albeit an even less spatially-representative multivariate verification statistic ("MULTb"). In this case, the spatial sampling does not permit meaningful estimates of NH or GLB mean quantities. In any of these diagnostics, a positive value of β is statistically significant at greater than 99% confidence as established from Monte Carlo simulations. Verification skills for the NINO3 reconstructions are estimated by other means, since the actual NINO3 index is not available far beyond the beginning of the calibration period. The (negative) correlation r of NINO3 with the Southern Oscillation Index ("SOI") annual-mean from 1865-1901 (obtained from P.D. Jones, personal communication) and a squared congruence statistic g^2 measuring the categorical match between the distribution of warm NINO3 events and the distribution of warm episodes according to the Quinn and Neal[35] historical chronology (available back through 1525) were used for statistical cross-validation of our reconstruction NINO3 index, with significance levels estimated from standard one-sided tables and Monte Carlo simulations respectively. The results of all calibration and verification experiments are provided in Supplementary Information.

1 Barnett, T.P., Santer, B., Jones,P.D, Bradley, R.S. & K.R. Briffa, Estimates of low frequency natural variability in near-surface air temperature, *Holocene*, **6**, 255–263 (1996).

2 Lean, J., Beer, J. & Bradley, R.S., Reconstruction of solar irradiance since 1610: Implications for climate change, *Geophys. Res. Lett.*, **22**, 3195–3198 (1995).

3 Crowley, T.J. & Kim, K.Y., Comparison of proxy records of climate change and solar forcing, *Geophys. Res. Lett.*, **23**, 359–362 (1996).

4 Delworth, T.D., Manabe, S. & Stouffer, R.J., Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model, *J. Climate*, **6**, 1993–2011 (1993).

5 Mann, M.E., Park, J., & Bradley, R.S., Global interdecadal and century-scale oscillations during the past five centuries, *Nature*, **378**, 266–270 (1995).

6 Bradley, R.S. & Jones, P.D., 'Little Ice Age' summer temperature variations: their nature and relevance to recent global warming trends, *Holocene*, **3**, 367–376 (1993).

7 Hughes, M.K. & Diaz, H.F., Was there a 'Medieval Warm Period" and if so, where and when?, *Clim. Change*, **26**, 109–142 (1994).

8 Diaz, H.F., & Pulwarty, R.S., An analysis of the time scales of variability in centuries-long ENSO-sensitive records in the last 1000 years, *Clim. Change*, **26**, 317–342 (1994).

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

10 Jones, P.D. & Briffa, K.R., Global surface air temperature variations during the 20th century: Part 1–Spatial, temporal and seasonal details, *Holocene*, **1**, 165–179 (1992).

11 Bradley, R.S., Are there optimum sites for global paleotemperature reconstruction?, in *Climatic Variations and Forcing Mechanisms of the last 2000 years*, ed. P.D. Jones, R.S. Bradley & Jouzel, J., Springer-Verlag, Berlin, 603–624 (1996).

12 Jacoby, G.C. & D'Arrigo, R. Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America, *Clim. Change* **14**, 39–59 (1989).

13 Jacoby, G.C., D'Arrigo, R.D. & Tsevegyn, D., Mongolian tree rings and 20th-century warming, *Science*, **9**, 771–773 (1996).

14 Fritts, H.C., Blasing, T.J., Hayden, B.P. & Kutzbach, J.E., Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate, *J. Appl. Meteor.*, **10**, 845–864 (1971).

15 Cook, E.R., Briffa, K.R., Jones, P.D., Spatial regression methods in dendroclimatology: a review and comparison of two techniques, *Int. J. Climatol.*, **14**, 379–402 (1994)

16 Briffa, K.R., Jones, P.D., Schweingruber, F.H., Tree-ring density reconstructions of summer temperature patterns across western North American since 1600, *J. Climate*, **5**, 735–753 (1992).

17 Schweingruber, F.H., Briffa, K.R., Jones, P.D., Yearly maps of summer temperatures in western Europe from A.D. 1750 to 1975 and Western North American from 1600 to 1982, *Vegetatio*, **92**, 5–71 (1991)

18 Fritts, H., *Reconstructing large-scale climatic patterns from tree ring data*, (The University of Arizona press, Tucson, 1991).

19 Guiot, J., The combination of historical documents and biological data in the reconstruction of climate variations in space and time. *Palaeoklimaforschung*, **7**, 93–104 (1988).

20 D'Arrigo, R.D., Cook, E.R., Jacoby, G.C. & Briffa, K.R., NAO and sea surface temperature signatures in tree-ring records from the North Atlantic sector, *Quat. Sci. Rev.*, **12**, 431–440 (1993).

21 Stahle, D.W. & Cleaveland, M.K., Southern Oscillation extremes reconstructed from tree rings of the Sierra Madre Occidental and southern Great Plains, *J. Climate*, **6**, 129–140 (1993).

22 Smith, T.M., Reynolds, R.W., Livezey, R.E., Stokes, D.C., Reconstruction of historical sea surface temperatures using empirical orthogonal functions, *J. Climate*, **9**, 1403–1420 (1996).

23 Kaplan, A. et al, Analyses of global sea surface temperature 1856-1991, *J. Geophys. Res.* (in the press).

24 Halpert, M.S., Ropelewski, C.F., Surface temperature patterns associated with the Southern Oscillation, *J. Climate*, **5**, 577–593 (1992)

25 Preisendorfer, R.W., *Principal component analysis in meteorology and oceanography*, Elsevier, Amsterdam (1988).

26 Cane, M. et al, Twentieth-century sea surface temperature trends, *Science*, **275**, 957–960 (1997).

27 Hurrell, J.W., Decadal trends in the North Atlantic Oscillation, regional temperatures and precipitation, *Science*, **269**, 676–679 (1995).

28 Chang, P., Ji, L. & Li, H., A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions, *Nature*, **385**, 516–518 (1997).

29 Folland, C.K., Parker, D.E. & F.E. Kates, Worldwide marine temperature fluctuations 1856-1981, *Nature*, **310**, 670–673 (1984).

30 Kushnir, Y., Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions, *J. Climate*, **7**, 141–157 (1994).

31 Schlesinger, M.E. & Ramankutty, N., An oscillation in the global climate system of period 65–70 years, *Nature*, **367**, 723–726 (1994).

32 Mann, M.E. & Park, J., Global-scale modes of surface temperature variability on interannual to century timescales, *J. Geophys. Res.*, **99**, 25819-25833 (1994).

33 Mann, M.E. & Park, J., Joint spatiotemporal modes of surface temperature and sea level pressure variability in the northern hemisphere during the last century, *J. Climate*, **9**, 2137–2162 (1996).

34 Delworth, T.D., Manabe, S. & Stouffer, R.J., Multidecadal climate variability in the Greenland Sea and surrounding regions: a coupled model simulation, *Geophys. Res. Lett*, **24**, 257–260 (1997).

35 Quinn, W.H. & Neal, V.T., The historical record of El Niño events, in *Climate Since A.D. 1500* (eds Bradley, R.S. & Jones, P.D., 623–648, Routledge, London, 1992).

36 Mann, M.E., Bradley, R.S., and M.K. Hughes, Long-term variability in the El Niño-Southern Oscillation and associated teleconnections. in Diaz, H.F. & Markgraf, V. (eds), *El Niño and the Southern Oscillation: Multiscale Variability and its Impacts on Natural Ecosystems and Society*, edited by H.F. Diaz and V. Markgraf, Cambridge University Press, 1998 (in the press).

37 Kelly, P.M., Jones, P.D. & Pengqun, J., The spatial response of the climate system to explosive volcanic eruptions, *Int. J. Climatol.*, **16**, 537–550 (1996).

38 Kirchner, I. & Graf, H.F., Volcanos and El Niño: signal separation in Northern Hemisphere winter, *Clim. Dyn.*, **11**, 341–358, (1995).

39 Raynaud, D., Barnola, J.M., Chappellaz, J. & Martinerie, P., Changes in trace gas concentrations during the last 2000 years, and more generally, the Holocene, in *Climatic Variations and Forcing Mechanisms of the last 2000 years*, ed. P.D. Jones, R.S. Bradley & Jouzel, J., Springer-Verlag, Berlin, 547–562 (1996).

40 Bradley, R.S. & Jones, P.D., Records of explosive volcanic eruptions over the last 500 years, in *Climate Since A.D. 1500* (eds Bradley, R.S. & Jones, P.D., 606–622, Routledge, London, 1992).

41 Robock A. & Free M.P., 1995: Ice cores as an index of global volcanism from 1850 to the present. *J. Geophys. Research*, **100**, 11549–11567 (1995).

42 Hegerl, G.C., Hasselmann, K., Cubasch, U., Mitchell, J.F.B., Roeckner, E., Voss., R. & Waszkewitz, J., Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change, *Clim. Dyn.*, **13**, 613634 (1997).

43 North, G.R. & M.J. Stevens, Detecting climate signals in the surface temperature record, *J. Climate* (in the press).

44 Hansen, J, Sato, M. & Ruedy, R., The missing climate forcing, *Phil. Trans. R. Soc. Lond.*, **B 352**, 231–240 (1997).

45 Stevens, M.J. & North, G.R., Detection of the Climate Response to the Solar Cycle, *J. Climate* (in the press).

46 Mann, M.E. & Lees, J., Robust estimation of background noise and signal detection in climatic time series, *Clim. Change*, **33**, 409–445 (1996).

47 Rind, D. & Overpeck, J., Hypothesized causes of decadal-to-century climate variability: climate model results, *Quat. Sci. Rev.*, **12**, 357–374 (1993).

48 Wigley, T.M.L. & Raper, S.C.B., Natural variability of the climate system and detection of the greenhouse effect, *Nature*, **344**, 324–327 (1990).

49 Thomson, D.J., Dependence of global temperatures on atmospheric CO_2 and solar irradiance, *Proc. Nat. Acad. Sci.*, **94**, 8370–8377 (1997).

50 Pollack, H.N. & D.S. Chapman, Underground records of changing climte, *Sci. Am*, **268**, 44–50 (1993).

Acknowledgements

This work benefited greatly from helpful discussions with Mark Cane, Ed Cook, Mike Evans, Alexey Kaplan, and collaborators at the Lamont Doherty Earth Observatory. We acknowledge discussions with K. Briffa, T. Crowley, P. Jones, S. Manabe, R. Saravanan, and K. Trenberth, as well as the comments of G. Hegerl and two anonymous reviewers. We are indebted to Rosanne D'Arrigo, David Fisher, Gordon Jacoby, Judith Lean, Alan Robock, David Stahle, Charles Stockton, Eugene Vaganov, Ricardo Villalba, and the numerous contributors to the International Tree-Ring Data Bank. We are also grateful to Frank Keimig, Martin Munro, Richard Holmes, and Caspar Ammann for their technical assistance. This work was supported by the National Science Foundation and the Department of Energy. M.E.M. acknowledges support through the Alexander Hollaender Distinguished Postdoctoral Research Fellowship program of the Department of Energy. This work is a contribution to the NSF- and NOAA-sponsored Analysis of Rapid and Recent Climatic Change (ARRCC) project.

Figure Captions

Figure 1. Data used in this study. (top) Distribution of annual resolution proxy indicators used in this study. Dendroclimatic reconstructions are indicated by "tree" symbols, ice core/ice melt proxies by "star" symbols, coral records by "C" symbols. Long historical records and instrumental "gridpoints" series are shown by squares (temperature), or diamonds (precipitation). Groups of "+" symbols indicate principal components of dense tree ring sub-networks, with the number of such symbols indicating the number of retained principal components. Sites are shown dating back to at least 1820 (red), 1800 (blue-green), 1750 (green), 1600 (blue), and 1400 (black). Certain sites (e.g., the Quelccaya ice core) consist of multiple proxy indicators (e.g., multiple cores, and both $\delta^{18}o$ isotope and accumulation measurements). (bottom) Distribution of the (1082) nearly continuous available land air/sea surface temperature gridpoint data available from 1902 onward indicated by shading. The squares indicate the subset of 219 gridpoints with nearly continuous records extending back to 1854 that are used for verification. Northern Hemisphere (NH) and global (GLB) mean temperature are estimated as areally-weighted (ie, cosine latitude) averages over the Northern hemisphere and global domains respectively.

Figure 2. Emirical orthogonal functions (EOFs) for the five leading eigenvectors of the global temperature data from 1902-1980. The gridpoint areal weighting factor used in the PCA procedure has been removed from the EOFs so that relative temperature anomalies can be inferred from the patterns.

Figure 3. Spatial patterns of reconstruction statistics. Top, calibration β (top–based on 1902-1980 data); middle, verification (based on 1854-1901 data) β; bottom, verification r^2 For the β statistic, values that are insignificant at the 99% level are shown in gray, while negative, but 99% significant values are shown in yellow, and significant positive values are shown in two shades of red. For the r^2 statistic, statistically insignificant values (or any gridpoints with unphysical values of correlation $r < 0$) are indicated in gray. The color scale indicates values significant at the 90% (yellow), 99% (light red), and 99.9% (dark red) levels (these significance levels are slightly higher for the calibration statistics which are based on a longer period of time). A description of significance level estimation is provided in the Methods section.

Figure 4. Comparison of the proxy-based spatial reconstructions of the anomaly pattern for 1941 vs the raw data. based on (top) actual, (middle) EOF-filtered, and (bottom) proxy-reconstructed data. Anomalies (relative to 1902-1980 climatology) are indicated by the color scale shown.

Figure 5. Time reconstructions (solid lines) along with raw data. (a) Principal Components (RPCs) 1-5 (b) Northern hemisphere mean temperature (NH) in ^{o}C. In both cases, the zero line corresponds to the 1902-1980 calibration mean of the quantity. For (b), raw data is shown through 1995 and positive and negative 2σ uncertainty limits are shown by the light dotted lines surrounding the solid reconstruction, calculated as described in the Methods section.

Figure 6. Reconstructed annual temperature patterns for two example years (i) 1791 (ii) 1816. The colors indicate regions which exceed (either positively or negatively) the threshold indicated in ^{o}C. The zero baseline is defined by the 1902-1980 climatological mean for each gridpoint.

Figure 7. Relationship of Northern hemisphere mean (NH) temperature with 3 candidate forcings between 1610 and 1995. (i) reconstructed NH temperature series from 1610-1980 update with raw data from 1981-1995. (i) greenhouse gases represented by atmospheric CO_2 measurements. (ii) reconstructed solar irradiance. (iii) weighted volcanic dust veil index (DVI). (iv) evolving multivariate correlation of NH series with the 3 forcings (i) (ii) and (iii). The time axis denotes the *center* of a 200 year moving window. One-sided (positive) 90%,95%,99% significance levels (see text) for correlations with CO_2 and solar irradiance are shown by horizontal dashed lines, while the one-sided (negative) 90% significance threshold for correlations w/ the DVI series is shown by the horizontal dotted line. The gray bars indicates two different 200 year window of data, with the long-dashed vertical lines indicating the center of the corresponding windows.